On the Statistical and Scaling Properties of Observed and Simulated Soil Moisture

by

Ara Ko

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2018 by the
Graduate Supervisory Committee:

Giuseppe Mascaro, Chair
Enrique R. Vivoni
Soe Myint
Zhihua Wang
Rebecca Muenich

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

Soil moisture (θ) is a fundamental variable controlling the exchange of water and energy at the land surface. As a result, the characterization of the statistical properties of θ across multiple scales is essential for many applications including flood prediction, drought monitoring, and weather forecasting. Empirical evidences have demonstrated the existence of emergent relationships and scale invariance properties in θ fields collected from the ground and airborne sensors during intensive field campaigns, mostly in natural landscapes. This dissertation advances the characterization of these relations and statistical properties of θ by (1) analyzing the role of irrigation, and (2) investigating how these properties change in time and across different landscape conditions through θ outputs of a distributed hydrologic model. First, θ observations from two field campaigns in Australia are used to explore how the presence of irrigated fields modifies the spatial distribution of θ and the associated scale invariance properties. Results reveal that the impact of irrigation is larger in drier regions or conditions, where irrigation creates a drastic contrast with the surrounding areas. Second, a physically-based distributed hydrologic model is applied in a regional basin in northern Mexico to generate hyperresolution θ fields, which are useful to conduct analyses in regions and times where θ has not been monitored. For this aim, strategies are proposed to address data, model validation, and computational challenges associated with hyperresolution hydrologic simulations. Third, analyses are carried out to investigate whether the hyperresolution simulated θ fields reproduce the statistical and scaling properties observed from the ground or remote sensors. Results confirm that (i) the relations between spatial mean and

standard deviation of θ derived from the model outputs are very similar to those observed in other areas, and (ii) simulated θ fields exhibit the scale invariance properties that are consistent with those analyzed from aircraft-derived estimates. The simulated θ fields are then used to explore the influence of physical controls on the statistical properties, finding that soil properties significantly affect spatial variability and multifractality. The knowledge acquired through this dissertation provides insights on θ statistical properties in regions and landscape conditions that were never investigated before; supports the refinement of the calibration of multifractal downscaling models; and contributes to the improvement of hyperresolution hydrologic modeling.

This work is dedicated to my husband Joohan Lee and my boy Luke Taeyang Lee, as well as my family in Korea who have always believed and supported me.

ACKNOWLEDGMENTS

First, I would like to thank my primary advisor Professor Giuseppe Mascaro. It has been a great experience to be his PhD student. I appreciate all his contribution of time, excellent guidance, general advice to support to make mu PhD experience productive and exciting. I also appreciate my co-advisor Professor Enrique Vivoni for his great guidance and mentorship. His expertise and work ethic have been a source of great inspiration. I would also like to thank the rest of my dissertation committee Dr. Soe Myint, Zhihua Wang, and Rebecca Muenich for their great insight and guidance. A special thanks to the members of our hydrology research group, Jenita Gautam, Xin Guan, Adil Mounir, Jorge Cazares-Rodriguez, Tiantian Xiang, Nicole Templeton, Adam Schreiner-McGraw, Kristen Whitney, and Vivian Verduzco who have been sources of great friendship and support and have helped make my graduate experience memorable and productive. My PhD research was supported by the National Aeronautics and Space Administration (NASA) research award (grant NNX12AP79G) for the project (Ensemble generation of downscaled soil moisture from satellite observations) and by startup funding of Dr. Mascaro from ASU and I am appreciative of the financial support. I also want to thank the Research Computing facility and support team at ASU for their high-performance computing environment and large data storage to make my high-resolution simulations possible.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

TABLE Page

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER 1

INTRODUCTION

Background

Soil moisture (θ) is a key state variable linking water, energy, and carbon fluxes at land surface (Lin et al., 1994; Entekhabi, 1995). It controls several hydrological processes such as infiltration, runoff, storage and drainage and determines the partitioning of the incoming radiation between latent and sensible heat fluxes (Vivoni et al., 2008a; Heathman et al., 2012; Vereecken et al., 2014; Martínez García et al., 2014; Bertoldi et al., 2014). Soil moisture is highly variable in space and time, due to the combined effects of soil texture (Cosh and Brutsaert, 1999; Vereecken et al., 2007), topography (Grayson et al., 1997; Western et al., 1999), vegetation (Mohanty et al., 2000; Teuling and Troch, 2005), and climate (Grayson et al., 1997; Lawrence and Hornberger, 2007). As a result, characterizing the spatial variability of θ across multiple scales over a wide range of settings is essential for a several applications, including informing climate, hydrologic and atmospheric models (Entekhabi et al., 2010; Mascaro and Vivoni, 2012), monitoring drought (Svoboda et al., 2002), and supporting agricultural and irrigation management (Tao et al., 2003), among others.

Soil moisture is mainly measured through ground sensors at different depths. While useful to characterize the temporal variability of this hydrologic variable, these observations are representative of a very limited area or support (a few meters) so that their utility for hydrologic applications at basin scale is limited. During the last two decades, spatially-distributed θ fields have become available globally from satellite

remote sensing products, including the Advanced Microwave Scanning Radiometer (AMSR-E; Njoku et al., 2003), the Soil Moisture Ocean Salinity (SMOS; Kerr et al., 2001) mission, and the most recently-launched Soil Moisture Active and Passive (SMAP; Entekhabi et al., 2010) mission. While satellite products have been extremely useful for applications at global and continental scales (e.g., Rebel et al., 2012; Reichle et al., 2007), their resolution is still too coarse to provide information at watershed scales (Crow and Wood, 2002).

The investigation of the spatial variability of θ across different scales requires the availability of high-resolution θ estimates. These have been provided at ~1 km resolution in a limited number of cases by airborne sensors flown during intensive field campaigns (e.g. Washita 1999 Experiments; Southern Great Plains (SGP) Hydrology Experiments of 1997 and 1999; Soil Moisture Experiments (SMEX) of 2002, 2003, 2004 and 2005; and National Airborne Field Experiment 2005 and 2006). Data from these datasets have been crucial to advance the scientific knowledge on the statistical properties of soil moisture and their control by land surface properties. A large number of studies have analyzed the relation between the spatial mean soil moisture ($<\theta>$) and the variability (σ_θ) in the θ fields using the datasets from the field campaigns (Crow and Wood, 2002; Brocca et al., 2010, 2012; Lawrence and Hornberger, 2007; Vereecken et al., 2007; Vivoni et al., 2008a; Famiglietti et al., 1999, 2008). These have shown contradictory patterns on the shape of the relation such as increasing variability with decreasing $<\theta>$ (Famiglietti et al., 1999), decreasing variability with decreasing $<\theta>$ (Vivoni et al., 2008a) and an increase up to a certain value of $<\theta>$ followed by a decrease (Brocca et al., 2010, 2012), which

were ascribed to the effects of climate and landscape features (Mascaro et al., 2011). Further analyses are needed to better characterize the simple, but very useful σ_θ vs. $<\theta>$ relation across different climates and environmental settings.

Another group of studies have used the aircraft-derived soil moisture products collected from the field campaigns to demonstrate that soil moisture exhibits scale invariance and multifractal properties. When a variable exhibits scale invariance, the statistical properties controlling its spatial distribution (e.g., the statistical moments) have clear relationships across different aggregation scales. The evidence of these properties has been then used to develop downscaling models that allow reproducing the variability of θ at fine scales (~1 km) starting from coarse (~50 km) estimates from satellite sensors (Hu et al., 1998; Kumar, 1999; Kim and Barros, 2002b; Perry and Niemann, 2008; Kaheil et al., 2008; Mascaro et al., 2010). Unfortunately, the evidence of scale invariance and multifractality has been only observed in a limited number of cases during intensive field campaigns. As for the σ_θ vs. $<\theta>$ relation, additional work is needed to evaluate whether soil moisture exhibits scale invariance and multifractality (i) across all climate regimes; (ii) beyond the scaling regime from ~1 km to ~50 km; and (iii) in landscapes where anthropogenic effects are significant, such are irrigated areas.

An alternative approach to generate high-resolution θ estimates that could be used to support these analyses is the application of distributed hydrological models run at high spatial resolution. However, due to computational and data challenges, high-resolution (<100 m) hydrologic simulations have been conducted in a very limited number of cases and mostly in small basins (Mahmood and Vivoni, 2008; Xiang et al., 2014). As a result,

only a few studies have evaluated whether simulated θ fields are characterized by the same statistical properties of observed measurements (Peters-Lidard et al., 2001; Grayson et al., 2002; Vivoni et al., 2010; Mascaro et al., 2015). For example, Vivoni et al. (2010) applied a distributed hydrologic model in a small basin in northwest Mexico and showed that the σ_θ vs. $<\theta>$ relation is well reproduced. Based on this positive result, these authors used the model outputs to investigate the physical controls on the soil moisture spatiotemporal variability and its feedback with evapotranspiration fluxes. Studies like the one from Vivoni et al. (2010) demonstrate that simulated θ fields are useful to advance our knowledge on the statistical properties of soil moisture fields and the underlying physical controls, including land surface (terrain, soil, and vegetation) properties and climate.

As summarized next, Chapters 2-5 of this thesis address research questions that have the overarching goal of advancing the knowledge of the statistical properties of soil moisture, with particular focus on scale invariance and multifractality. For this aim, soil moisture data obtained from remote sensors and hyper-resolution hydrologic simulations will be utilized.

Outline of Chapters 2-5

Previous work of Mascaro et al. (2011) showed preliminary evidence that the presence of irrigated agricultural fields impacts significantly the scale invariance properties observed in natural landscapes. However, no study has quantified these impacts and included their effects on downscaling schemes. The analyses presented in Chapter 2 aim at answering two research questions:

1. *What is the effect of irrigated agricultural areas on the scale invariance and multifractal properties of soil moisture?*
2. *Can we improve soil moisture downscaling models in areas with presence of irrigation?*

For this aim, the scale invariance and multifractal properties of aircraft-based θ fields are compared in two areas characterized by different climate and combinations of natural landscapes and croplands. The data are part of the National Airborne Field Experiments 2005 (NAFE05) and 2006 (NAFE06) campaigns conducted in Australia. A methodological framework has been proposed to (1) identify the irrigated areas from remote sensing data, (2) quantify the impact of irrigation on the scale invariance properties, and (3) apply a multifractal downscaling model while accounting for the spatial heterogeneity due to irrigated croplands.

As previously mentioned, the evidence of scale invariance in soil moisture fields has been only observed in a limited number of cases during intensive field campaigns, when soil moisture estimates at high spatial resolution were available from airborne sensors. An alternative way of obtaining soil moisture fields for this type of studies is

through the use of distributed hydrologic models run at high spatial resolution (hereafter, hyperresolution). However, hyperresolution simulations have been conducted in a very limited number of cases and mostly in small basins (Xiang et al., 2014). The work presented in Chapter 3 has the goal of addressing the research question:

3. *Can we generate reliable high-resolution θ fields over large areas and for a long-time period, using a physically-based distributed hydrologic model?*

In a recent opinion paper, Wood et al. (2011) identified several challenges that are still preventing hyperresolution hydrologic simulations, including high computational cost and lack of meteorological forcings and terrain characterization data at high spatial resolution. In Chapter 3, some of these challenges have been addressed by applying the parallel version of the Triangulated Irregular Network (TIN)-based Real-Time Integrated Basin Simulator (tRIBS) hydrologic model (Ivanov et al., 2004a) to the Rio Sonora basin (RSB) in northwest Mexico. Simulations have been conducted at an average spatial resolution of ~88 m over 10 years. A downscaling scheme accounting for terrain effects has been applied to generate meteorological forcings at a resolution of ~1 km starting from reanalysis products at ~12 km. In addition, spatial maps of soil parameters have been generated at ~250 m by combining global soil texture maps and a local soil map (~6 km). The model was validated against independent observation of soil moisture, and with estimates of land surface temperature from the Moderate Imaging Spectroradiometer (MODIS) through a set of spatial analysis tools, including Taylor diagrams, connectivity analysis, and empirical orthogonal function analysis to quantify model performances and diagnose the potential causes of model deficiencies.

Given the limited number of hyperresolution hydrologic simulations, only a few studies have analyzed the scaling properties of model-derived θ (e.g., Gebremichael et al., 2009). The hyperresolution simulations obtained in Chapter 3 provide an extensive dataset of high-resolution θ fields spanning several years with different wetness states. Chapter 4 addresses the research question:

4. *Do soil moisture fields simulated by hyperresolution hydrologic simulations reproduce observed statistical properties including scale invariance?*

First, the spatial variability of 10-years simulated θ fields of the entire RSB and sixteen square domains of size 32 km by 32 km covering the entire RSB with different surface properties and conditions has been explored by comparing the relation between spatial standard deviation (σ_θ) and the coefficient of variation (CV_θ) of the spatial mean of the field ($<\theta>$). These relations have been compared to previous studies conducted in (i) SMEX04, Sonora (SON) (Mascaro and Vivoni, 2010) and (ii) two campaigns for 50-km scale; SMEX02 and SMEX03 (Famiglietti et al., 2008). Then, the presence of scale invariance properties in θ fields of the domains is investigated from 32 km to 1 km. In addition, the variability of the properties under different soil, vegetation, and terrain conditions is investigated and quantified, as well as the physical controls on the spatial variability of $<\theta>$ and scaling properties were investigated.

Finally, Chapter 5 summarizes general conclusion of the preceding chapters and outlines insights to future research. The results presented in chapters 2-4 correspond to journal articles that have been either published, submitted, or in preparation:

Chapter 2: Ko, A., G. Mascaro, and E.R. Vivoni. 2016. Irrigation impacts on scaling properties of soil moisture and the calibration of a multifractal downscaling

algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6): 3128-3142.

Chapter 3: Ko, A., G. Mascaro, and E.R. Vivoni. 2018. Strategies to Improve Physics-Based Hyperresolution Hydrologic Simulations at Regional Basin Scales. (In review, *Water Resources Research*).

Chapter 4: Ko, A., and G. Mascaro 2018. Statistical and Scaling Properties of Simulated Soil Moisture. (In Preparation).

CHAPTER 2

IRRIGATION IMPACTS ON SCALING PROPERTIES OF SOIL MOISTURE AND THE CALIBRATION OF A MULTIFRACTAL DOWNSCALING ALGORITHM

Introduction

Soil moisture (θ) is a crucial variable controlling and linking the water and energy fluxes at the land surface (Entekhabi, 1995) . Knowledge of the spatiotemporal variability of θ is thus important to inform weather forecasting, climate and hydrologic models (Entekhabi et al., 2010; Mascaro and Vivoni, 2012); monitor drought (Svoboda et al., 2002); and support agricultural and irrigation management (Tao et al., 2003). Spatial θ fields have become available globally from satellite remote sensing products, including the Advanced Microwave Scanning Radiometer (AMSR-E; Njoku et al., 2003), the Soil Moisture Ocean Salinity (SMOS; Kerr et al., 2001) mission, and the recently-launched Soil Moisture Active and Passive (SMAP; Entekhabi et al., 2010) mission. While satellite products have been extremely useful for applications at global and continental scales (e.g., Rebel et al., 2012; Reichle et al., 2007), their resolution is still too coarse to provide information at watershed scales (Crow and Wood, 2002). To increase the utility of satellite θ products, downscaling (or disaggregation) approaches have been proposed based on different methods, including: (i) application of hydrologic models (Pellenq et al., 2003; Merlin et al., 2006a), (ii) use of remotely-sensed products and ground meteorological data (Merlin et al., 2005, 2006b), (iii) use of high-resolution land surface properties (Shin and Mohanty, 2013), and (iv) approaches that reproduce the statistical

properties of θ (Hu et al., 1998; Kumar, 1999; Kim and Barros, 2002b; Perry and Niemann, 2008; Kaheil et al., 2008; Mascaro et al., 2010).

A class of statistical downscaling algorithms is based on the evidence that θ fields exhibit scale invariance and fractal or multifractal properties across a range of settings (Rodriguez-Iturbe et al., 1995; Hu et al., 1997; Kim and Barros, 2002a; Oldak et al., 2002; Das and Mohanty, 2008; Mascaro and Vivoni, 2010). For example, Kumar (1999) proposed a model that reproduces these statistical properties through a Kalman filtering algorithm, while Hu et al. (1998) and Das and Mohanty (2008) adopted techniques based on wavelet transforms. Kim and Barros (2002b) developed a technique that uses fractal interpolation informed by ancillary data to produce a unique disaggregated field. An alternative downscaling approach based on a stochastic generator of multifractal, homogeneous cascades has been adopted by Mascaro et al. (2011, 2010) to disaggregate θ from 25.6 km (~AMSR-E footprint) to 800 m (~an aircraft footprint), using θ data from the Southern Great Plains experiment in 1997 (SGP97) in Oklahoma, the Soil Moisture Experiment of 2002 (SMEX02) held in Iowa, and of 2004 (SMEX04) conducted in Arizona and Sonora, Mexico. This model depends only on two parameters that were calibrated as a function of coarse-scale predictors including the soil moisture mean and ancillary factors accounting for terrain, soil texture and land cover features in the satellite footprint. The applications of Mascaro et al. (2011, 2010) demonstrated the model robustness across different climate regimes and land surface characteristics. However, the model exhibited lower performances in agricultural areas where irrigated fields introduce spatial heterogeneity in the θ distribution.

In agricultural areas with irrigation systems, additional water input modifies the spatial and temporal variability that θ fields would typically have in natural landscapes, including their scale invariance and multifractal properties. As a result, if the presence of irrigation is not taken into account, the applicability of any type of downscaling algorithm used to increase the utility of satellite products (e.g., SMOS, SMAP) may be limited. Despite this, the effect of irrigated crops on θ spatial variability and downscaling algorithms has received very little attention. Mascaro et al. (2011) provided preliminary evidence that scale invariance properties of θ may be impacted differently by the presence of agricultural fields that have likely been irrigated, depending on climate of the region and the extent of the agricultural districts. Specifically, these authors showed that, in a large agricultural region in Iowa with homogenous land cover and a humid climate, θ fields are scale invariant in a single range from 25.6 km to 800 m with similar properties across the area. In contrast, in an arid climatic setting in Arizona, the presence of agricultural districts of limited extent modified the scale invariance properties introducing an additional scale regime. Hence, results from Mascaro et al. (2011) indicate that agricultural fields have a greater impact on the θ distribution in arid and semiarid areas surrounded by drier land. In that study, the authors identified the occurrence of irrigation based on reports of field visits and inspection of land cover maps (e.g., center-pivot irrigation systems were clearly visible in the map for Arizona), without proposing a technique to objectively identify the irrigated fields.

This study aims at further investigating the irrigation impacts on the scale invariance properties of θ fields and the application of a multifractal downscaling

algorithm. I use ground and aircraft-based (1-km) θ measurements from the National Airborne Field Experiments (NAFE) 2005 (NAFE05) and 2006 (NAFE06) field campaigns, undertaken in two regional sites in Australia to support the SMOS mission (Panciera et al., 2008; Merlin et al., 2008b). The two study sites have different characteristics in terms of climate (temperate in NAFE05 and semiarid in NAFE06), extent of agricultural districts, and crop types that require different irrigation schedules and amounts. First, I develop a simple method based on Landsat 5 Thematic Mapper (hereafter Landsat TM5) at 30-m resolution to detect irrigated areas in the two study regions. This information is then used to create three scenarios for the scale invariance analysis from the coarse scale of 32 km (~satellite footprint) to the fine scale of 1 km (aircraft footprint). In the first scenario (SC1), the analysis is conducted on the original θ data, thus considering potential spatial heterogeneity induced by irrigation activities. The other two scenarios aim at removing the source of spatial heterogeneity in θ distribution by replacing the soil moisture values in irrigated pixels with missing data (SC2) or with values interpolated from neighboring non-irrigated pixels (SC3). SC2 and SC3 represent then two possible strategies to simulate conditions where irrigation was not applied, thus mimicking natural settings. Metrics testing the presence of scale invariance are used in the three scenarios to evaluate how irrigated pixels affect these statistical properties in the two study regions.

Results of the scale invariance analysis are then used to apply the multifractal downscaling model of Mascaro et al. (2011) under SC3, where the effect of irrigation has been attenuated. After calibrating the model as a function of coarse-scale predictors, I

evaluate its performance by comparing an ensemble of disaggregated fields with (i) the observed small-scale θ distribution from the aircraft products and (ii) ground θ measurements. This is the first time that ground data are used to verify the skill of this type of downscaling model. Results from the model verification are discussed to identify the skills and limitations of the procedure to remove the effect of irrigated croplands. Note that, by using SC3, the disaggregated fields are generated in a spatially homogeneous framework and, in future work, methods should be developed to include the effect of the spatial heterogeneity due to irrigated croplands. Given the increasing extent of irrigated lands globally (Siebert et al., 2015), the findings of this study are useful to advance the general knowledge of soil moisture spatiotemporal variability, with focus on anthropogenic settings, and to support the development of a broad range of downscaling routines enhancing the utility of satellite products, including the SMOS and SMAP missions.

Study Sites and Datasets

NAFE05 and NAFE06 Campaigns

NAFE05 and NAFE06 are two intensive field campaigns conducted in southeastern Australia [see Figure 2.1(a)] during November 2005 and 2006, respectively, with the goal of supporting the development of the SMOS θ retrieval algorithm, as well as the design of downscaling and data assimilation techniques for future SMOS products (Panciera et al., 2008, 2009; Panciera, 2009; Merlin et al., 2008b, 2008a, 2009). For this aim, ground θ measurements were collected across a range of soil, vegetation and terrain

conditions during overpasses of the Polarimetric L-band Multi-beam Radiometer (PLMR) mounted on an aircraft. During NAFE05, the aircraft flights covered a region of 40 km x 40 km in the Golburn River catchment in the Australian state of New South Wales [see Figure 2.1(b)]. In this area with temperate climate (mean annual rainfall of 650 mm), elevation ranges from 174 to 802 m (average of 364 m) and the mean slope is 4.25˚ [Figure 2.1(b)]. The dominant soil types are clay (55.0% of the total area) and sand (35.5%), while the main land cover class is woodland (49.2%) followed by forest (19.4%), rainfed pasture (18.5%) and rainfed crop (12.8%). The predominant crop types are wheat, barley and lucerne with small amounts of sorghum and oat (Walker et al., 2008). Table 2.1 summarizes the terrain, soil texture and land cover features of the region, along with the source of each dataset.

NAFE06 was carried out in three separate sites with semiarid climate (mean annual rainfall of 300 mm) of the Murrumbidgee catchment in New South Wales (Merlin et al., 2008b). Here, I focused on the Yanco site, where aircraft measurements were collected on a region of 41 km x 59 km that includes the Coleambally Irrigation Area (CIA), an intensive agricultural district with more than 500 farms [Figure 2.1(c)]. Terrain is gently rolling, with elevation ranging from 114 to 164 m (average of 126 m) and a mean slope of 0.35˚ (see Figure 2.1(c) and Table 2.1). The soil type is mostly sandy loam (~99.0% of the area), while rainfed crop is the major land cover class (45.2%), followed by rainfed pasture (22.5%) and woodland (18.5%). This land cover classification, derived from Lymburner et al. (2011) and based on Moderate Resolution Imaging Spectroradiometer (MODIS) imagery at 250-m resolution, does not include irrigated

crops. However, in a description of the NAFE06 experiment, Walker et al. (2006) reported that, in the CIA district, flood irrigation is applied in November on rice fields that represent the dominant crop type. To address this, here I refined the classification of Lymburner et al. (2011) by developing a method to identify irrigated fields in the two study sites.

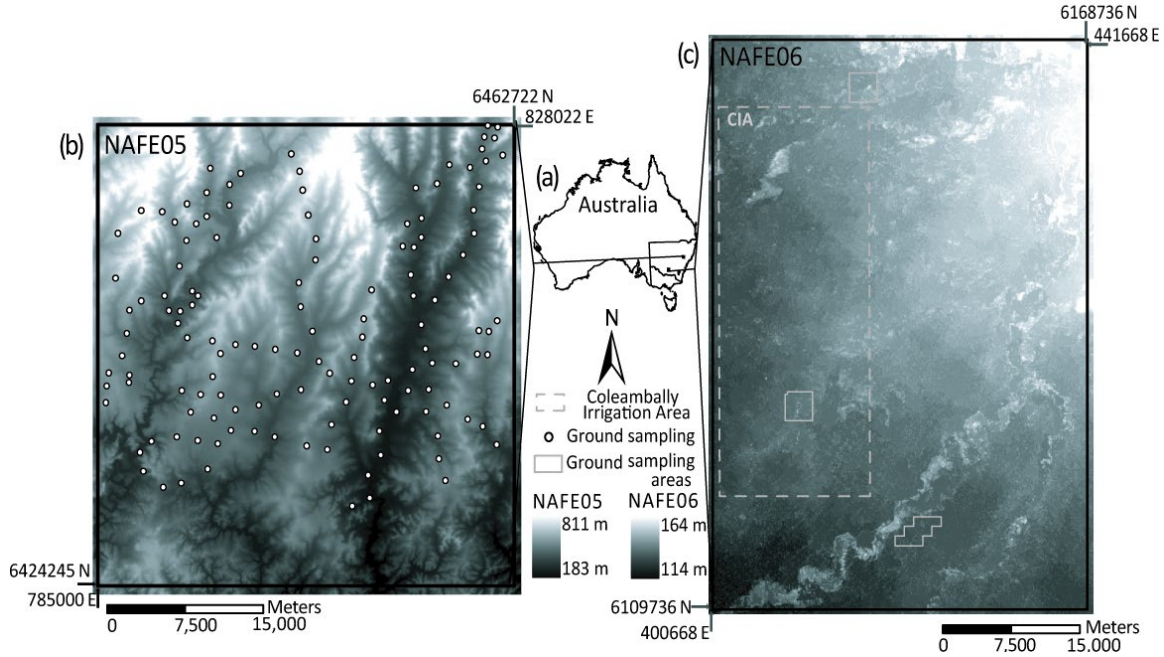

Figure 2.1. (a) Geographic location of the NAFE05 and NAFE06 study sites in Australia. Topography of (b) NAFE05 and (c) NAFE06 study areas derived from a 90-m DEM of SRTM, along with the boundaries of the aircraft acquisition shown with black lines and with examples of ground sampling points on Nov. 7 for NAFE05 and farm areas for NAFE06 shown with black dots or lines. In (c), the white dashed line indicates the boundary of the CIA area (see text for details).

Table 2.1. Summary of terrain, soil texture and land cover features for the NAFE05 and NAFE06 study regions with the source of each dataset. For topography, minimum, maximum and mean elevation (H_{min}, H_{max}, and H_{mean}), and mean slope (SL_{mean}) are listed. For soil texture and land cover, the main classes and their percentage at each site are reported.

Feature and Source	NAFE05	NAFE06
Topography from the 90-m Digital Elevation Model (DEM) of the Shuttle Radar Topography Mission (SRTM)	H_{min} = 174 m, H_{max} = 802 m, H_{mean} = 364 m, SL_{mean} = 4.5°.	H_{min} = 114 m, H_{max} = 164 m, H_{mean} = 126 m, SL_{mean} = 0.35°.
Soil Texture from *McKenzie et al*. (2000)	Loamy Clay (9.5%), Clay (55.0 %), Sand (35.5 %).	Sandy Loam (99.0%), Loamy Clay (1.0%).
Land Cover from *Lymburner et al*., (2011)	Rainfed Crop (12.8%), Rainfed Pasture (18.5%), Forest (19.4%), Woodland (49.2%).	Water Body (0.1%), Rainfed Crop (45.2%), Rainfed Pasture (22.5%), Tussock Grass (6.7%), Forest (7.0%), Woodland (18.5%)

Aircraft- and Ground-based Soil Moisture Data

During NAFE05, the PLMR sensor measured brightness temperature (T_B) on four days within a 22-day period (October 31, November 7, 14 and 21, 2005) over the region of 40 km x 40 km depicted in Figure 2.1(b). T_B was converted into maps of near-surface (top 3 cm) soil moisture at 1-km resolution according to the retrieval algorithm of Panciera (2009) and Panciera et al. (2009). In concomitance with the flights, ground θ data were collected over a regional network with a variable number of locations (45 to 152 depending on the day) and a mean spacing of 1 km (Panciera et al., 2008). An example of the regional sampling on November 7 is shown in Figure 2.1(b). For NAFE06, maps of θ at 1-km resolution were derived from PLMR T_B observations using the retrieval algorithm described by Merlin et al. (2008b). Maps were produced for eleven days (October 31, November 2-5, 7, 9, 13, 14, 16 and 18, 2006) over a 41 km x 59 km area and, during each day, ground measurements were taken at three farms with a spacing of 250 m in a variable number of locations ranging from 305 to 428 [Figure 2.2(c)] (Merlin et al., 2008b, 2009). Since, on some days, the values of θ retrieved in pixels in the CIA agricultural area were higher (>70%) than typical soil porosity, I replaced these values with the maximum of 58% adopted in NAFE05. Despite a slight discrepancy between the sampling depth of ground θ measurements (6 cm) and aircraft sensor (3 cm), the validation of the retrieval algorithms was adequate except for a slightly negative bias in irrigated pixels of NAFE06 (Merlin et al., 2009).

Time series of the soil moisture spatial mean ($<\theta>$) and standard deviation (σ_θ) computed from aircraft and ground data in each region are shown in Figure 2.2. In both

campaigns, ground and aircraft products are in good agreement in the estimation of $<\theta>$. In contrast, the values of σ_θ from the two sources differ, with ground data observing lower (negative bias of ~2%) and higher (positive bias of ~4%) spatial variability as compared to the aircraft data in NAFE05 and NAFE06, respectively. This discrepancy can be attributed to the different extent (regional vs. farm) and land cover type of the ground sampling areas, including data collection within farms during NAFE06. As indicated by the arrows in Figure 2.2(a), two heavy storms were observed on the first two days of NAFE05 that led to wet conditions, while in the other two days the region was considerably drier. During NAFE06, two rainfall events occurred on November 3 and 13 that caused two peaks of $<\theta>$ followed by dry down periods [see Figure 2.2(c)]. Overall, NAFE05 was characterized by wetter conditions than NAFE06 (average values of $<\theta>$ from aircraft of 28.2% and 11.2%, respectively). In both sites, σ_θ is linked with $<\theta>$ according to the bell-shaped relation found in other studies (e.g., Brocca et al., 2007; Lawrence and Hornberger, 2007; Mascaro and Vivoni, 2010).

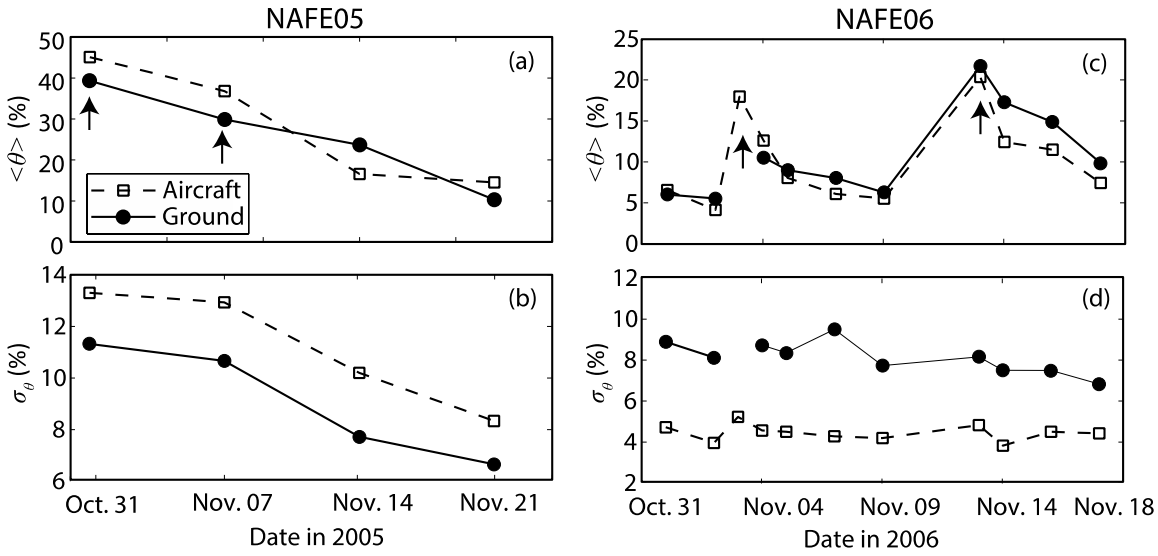

Figure 2.2. (a)-(b) Time series of aircraft- and ground-based (a) spatial mean soil moisture $<\theta>$ and (b) standard deviation σ_θ collected during NAFE05. (c)-(d) Same as (a)-(b) but for NAFE6. The arrows in (a) and (c) indicate rainy days.

Methods

Identification of Irrigated Areas

I designed a semiautomatic method to identify irrigated areas using Landsat TM5 images at 30-m resolution taken ~10 days before the beginning of each experiment (October 21, 2005 and October 22, 2006, respectively). True color maps in the two study regions are shown in Figure 2.3(a) and (d), where it is possible to visually detect the agricultural fields, forested areas in NAFE05 and a river floodplain in NAFE06. To support the analyses, boundaries of forested area and river floodplain were identified through visual analysis and the land cover maps of Table 2.1. The irrigated cropland detection method is based on two indices computed from the different bands of Landsat TM5. The first index is the Soil Adjusted Vegetation Index (SAVI), defined as:

$$SAVI = \frac{NIR - R}{NIR + R + M}(1 + M) \tag{2.1}$$

where NIR and R are the Near Infrared (band 4; wavelength range of 0.76-0.90 μm) and Red (band 3; 0.63-0.69 μm) bands and M is a factor accounting for vegetation density varying from 0 to 1. Huete (1988) suggested the adoption for M of an intermediate value of 0.5 to account for first-order soil background variations. SAVI is a modification of the Normalized Difference Vegetation Index (NDVI) that isolates vegetation from soil and water bodies, while minimizing the influence of soil brightness when vegetative cover is low. SAVI ranges from -1 to 1, with lower values associated with smaller amount and cover of green vegetation. Following Kaplan and Myint (2012), I assumed that areas with SAVI larger than a threshold SAVI* can be classified as an irrigated cropland. Specifically, Kaplan and Myint (2012) identified irrigated agricultural

areas in a semiarid region of Arizona by adopting $SAVI^{*}$= 0.4. Given the similarity in climate, this value of $SAVI^{*}$ was used for NAFE06. For the more vegetated NAFE05 area, I increased the threshold to $SAVI^{*}$ = 0.8 to separate agricultural fields from other vegetation types. The identification of this threshold was aided by visual comparison with the true color images.

The pixels where the condition $SAVI > SAVI^{*}$ is verified are depicted in Figure 2.3(b) and (e). Comparison with the true color maps of Figure 2.3(a) and (d) revealed that this single condition is not sufficient to identify all irrigated fields. Thus, I computed a second index defined as the ratio between bands 5 (1.55-1.75 μm) and 7 (2.08-2.35 μm) or band ratio (hereafter BR) that is able to separate water bodies from land and enhance the detection of saturated areas, an indication of irrigation (Quinn, 2001). Since lower values of BR are associated with higher moisture, I adopted the condition $BR < BR^{*}$, with BR^{*} being a threshold, to detect irrigated fields. To identify BR^{*} in the study regions, I initially determined a range of possible values based on Richards (Richards and Richards, 1999), who suggested the interval from 0.3 to 0.7. For NAFE06 where saturated conditions can occur due to flood irrigation, I adopted the middle value of the range (BR^{*} = 0.5). For NAFE05 where saturated croplands are more difficult to observe, I increased BR^{*} to 0.7. Figure 2.3(b) and (c) show the images with pixels where the condition $BR < BR^{*}$ is verified as well as with pixels where both conditions occur, that is $SAVI > SAVI^{*}$ and $BR < BR^{*}$.

The binary images at 30-m resolution produced by applying the two conditions were subsequently resampled at the same resolution (1 km) of the θ fields using a

majority algorithm. Based on visual comparisons, I found that this procedure identified pixels within cropland regions, as well as forested areas in NAFE05 and river floodplains in NAFE06 [see boundaries highlighted in Figure 2.3(a) and (d)]. Note that these are all cases of vegetated areas with high water content. The final images with irrigated croplands were then obtained by eliminating pixels meeting the conditions but located within forests and river floodplains, as shown in Figure 2.3(c) and (f) (labeled as discarded). A total of 104 (6.5% of total area) and 208 (8.6%) pixels at 1-km resolution were classified as irrigated cropland in NAFE05 and NAFE06. In the NAFE06 site, they are distributed in large connected areas within CIA district, while in NAFE05 they are more scattered across the region.

The accuracy of the detection method was tested using the soil moisture images available at each site. For each day, I sorted the θ values in decreasing order and computed the rank in each pixel (i.e., wetter pixels will have lower ranks than drier ones). This was repeated for all available days and the mean rank at each pixel was calculated. Irrigated pixels that are artificially wetter should have smaller ranks than other pixels. In NAFE05, I found that the mean rank of irrigated pixels is 816 out of a total of 1600. This outcome is influenced by the limited sample of the aircraft images (four) and by the fact that two of them are affected by large rainfall events (Figure 2.2), during which the fields may have not been irrigated. As such, in this study site I expect a lower impact of irrigation on the scale invariance properties of θ. In contrast, for NAFE06 the mean rank in irrigated pixels is 679 out of a total of 2419, an indication that pixels identified as irrigated were effectively wetter than the surrounding areas.

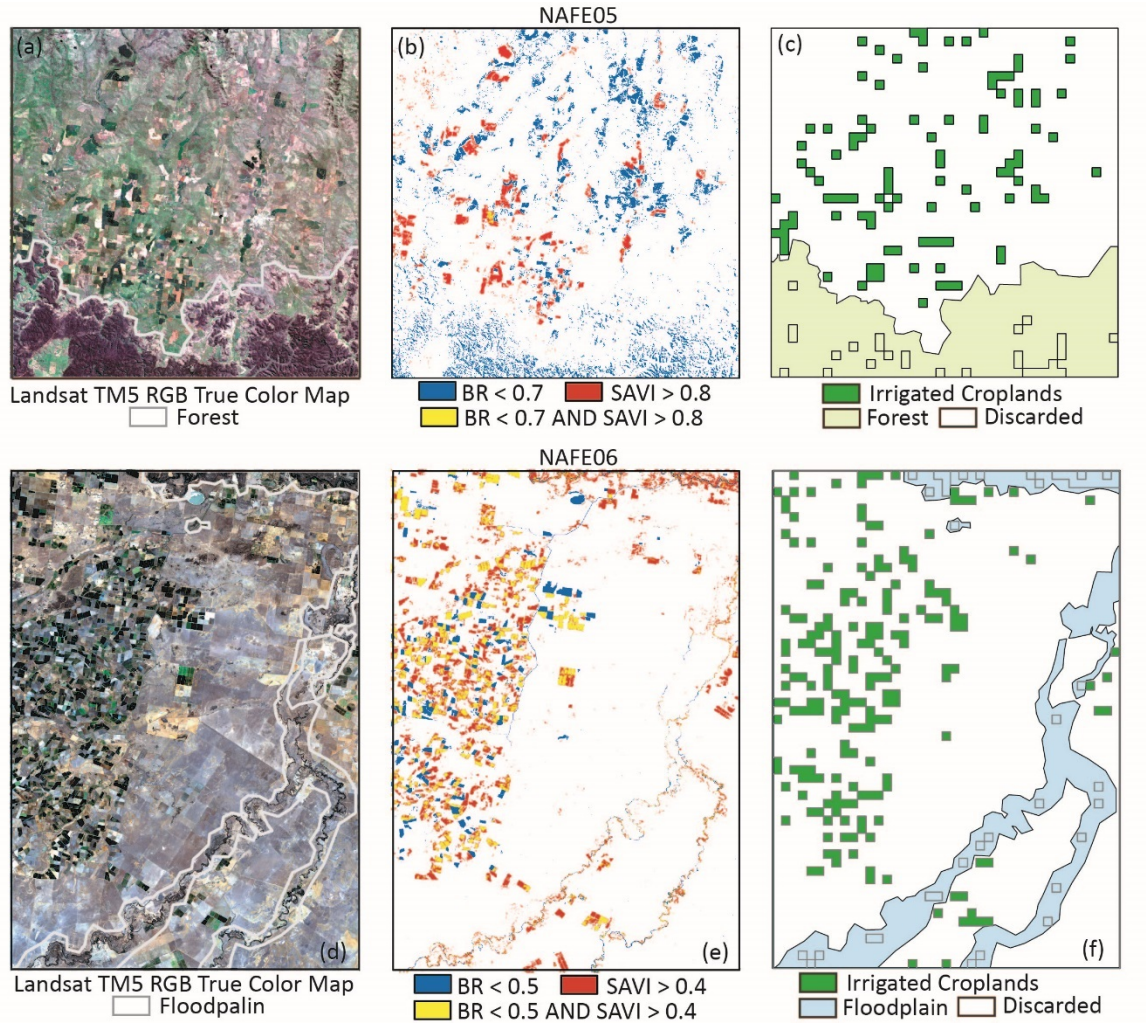

Figure 2.3. True color Landsat TM5 image at 30-m resolution of NAFE05; (b) map at 30-m resolution with pixels verifying one or both conditions SAVI > SAVI* and BR < BR*, with SAVI* = 0.8 and BR* = 0.7; and (c) map at 1-km resolution with pixels verifying the conditions, including irrigated and discarded pixels. (d)-(e)-(f) Same as (a)-(b)-(c) but for NAFE06, with SAVI* = 0.4 and BR* = 0.5.

Scale Invariance and Multifractal Analyses

The scale invariance and multifractal analysis was conducted following the approach of Mascaro et al. (Mascaro et al., 2010), which is based on Deidda (2000). In the following, I briefly summarize the theoretical framework. Scale invariance is tested between a fine (l) and coarse (L) scale linked by the binary rule $L = l \cdot 2^{N_{lev}}$, where N_{lev} is the number of the downscaling levels. In our case, the fine scale was fixed by the aircraft footprint, i.e. l = 1 km, while the coarse scale was selected as L = 32 km to approximate the resolution of most satellite products (N_{lev} = 5). As in Mascaro et al., 2011, I identified four and eight coarse-scale domains L x L within the NAFE05 and NAFE06 regions, respectively, by moving the square L x L on a grid with regular spacing of 8 km, as shown in Figure 2.4(a) and (d). Table 2.2 reports the terrain, land cover and soil texture characteristics for each domain, as well as the percentage of pixels identified as irrigated croplands.

In each domain, the analysis requires the computation of the structure function $S_q(\lambda)$ at a given scale λ included between l and L and for the moments q, defined as:

$$S_q(\lambda) = \frac{1}{N(\lambda)^2} \sum_{i=1}^{N(\lambda)} \sum_{i=1}^{N(\lambda)} \left[\theta_{i,j}(\lambda)\right]^q \tag{2.2}$$

where $\theta_{i,j}(\lambda)$ is the mean value of θ on a grid cell λ x λ in the location (i, j) within the domain and $N(\lambda)^2 = (L/\lambda)^2$ is the number of λ x λ cells embedded in the coarse domain. The presence of scale invariance is then assessed by verifying the existence of the power law:

$$S_q(\lambda) \sim \lambda^{-K(q)} \tag{2.3}$$

where $K(q)$ is the multifractal exponent. More simply, scale invariance is investigated by testing the linearity of the log-log transformation of (3):

$$\log S_q(\lambda) \sim K(q) \cdot \log \lambda \tag{2.4}$$

If (4) holds, the statistical information of the field can be transferred from the coarse scale L to any smaller scale $\lambda \geq l$ through the knowledge of $K(q)$. The type of relation linking $K(q)$ with q determines whether the spatial field is fractal (linear) or multifractal (non-linear).

Table 2.2. Terrain characteristics (mean elevation, H_{mean}, and mean slope, SL_{mean}), percentage of the main land cover and soil texture classes, and percentage of irrigated pixels for the coarse domains L x L identified in the NAFE05 and NAFE06 study regions.

NAFE05										
Domain	***H_{mean}*** **(m)**	***SL_{mean}*** **(°)**	**Rainfed crop**	**Rainfed pasture**	**Forest**	**Woodland** **(%)**	**Loamy clay**	**Clay**	**Sand**	**Irrigated pixels**
1	394.5	4.2	15.7	19.8	20.2	44.3	0.1	53.7	46.2	7.6
2	395.3	4.5	14.0	19.4	20.5	46.0	0.1	58.8	41.1	6.0
3	498.5	4.3	17.1	21.8	8.1	52.9	12.4	66.3	21.2	9.7
4	496.6	4.4	15.6	21.5	7.3	55.4	11.6	72.3	16.1	8.3
NAFE06										
Domain	***H_{mean}*** **(m)**	***SL_{mean}*** **(°)**	**Rainfed crop**	**Rainfed pasture**	**Tussock grass**	**Woodland** **(%)**	**Sandy loam**			**Irrigated pixels**
1	122.8	0.4	50.3	21.7	7.7	13.5	100.0			8.4
2	124.3	0.4	49.4	21.2	9.2	13.8	98.9			5.0
3	124.5	0.3	49.2	25.3	7.6	13.1	100.0			11.8
4	126.6	0.4	49.1	23.0	8.8	14.1	98.9			6.5
5	126.0	0.3	44.5	27.0	8.1	16.6	100.0			14.0
6	128.7	0.4	45.3	24.2	8.8	17.1	98.9			8.4
7	127.2	0.3	43.2	25.6	6.9	19.6	100.0			13.1
8	130.4	0.4	44.5	23.4	7.2	19.5	98.9			8.8

Scenarios for Testing the Effect of Irrigation on Scaling Properties

Irrigation activities may impact the spatial distribution of θ by introducing areas that tend to be consistently wetter than the surrounding regions. This effect is less evident when the entire region is wet (e.g., after rainfall events) and becomes more significant in dry conditions. From the statistical perspective, the presence of permanently wet areas creates spatial heterogeneity in the θ distribution that can affect its scale invariance and multifractal properties and limit the application of downscaling models based on the hypothesis of spatial homogeneity (i.e., the statistical distribution of θ is the same at each pixel). Here, I investigated the impact of irrigation through the three scenarios previously defined and labeled as SC1, SC2 and SC3. As an example, Figure 2.4 shows the θ fields created for the three scenarios for one day. In each panel, the spatial mean soil moisture ($<\theta>$) and the irrigated cropland pixels are shown: note that $<\theta>$ is highest in SC1 and lowest in SC2 (computed from non-missing values).

The effectiveness of SC2 and SC3 in removing the source of spatial heterogeneity in θ was tested by introducing two tests, labeled as SC2-R and SC3-R, where the same strategies were applied on the same number of irrigated pixels (N_{irr}), but distributed in random locations (-R is for random). These tests allow evaluating if the change in statistical properties caused by replacing θ in pixels identified as irrigated croplands is more significant than the change due to substituting θ in the same number of pixels randomly located in the region. For each soil moisture image, an ensemble of 100 fields was created by randomly selecting N_{irr} pixels (N_{irr} = 104 and 208 for NAFE05 and NAFE06, respectively) and θ values were replaced in each field with missing values in SC2-R or with interpolated values from neighboring pixels in SC3-R.

The scale invariance analysis was carried out for each scenario for the moments q = 1.5, 2, 2.5, 3, 3.5, and 4 and evaluated by calculating two metrics that test the linearity of (4), including the Correlation Coefficient (CC) and Root Mean Square Error (RMSE) of the linear regression for q = 4. I focused on the highest moment because it is more sensitive to the presence of high wetness associated with irrigated croplands: when q is high, wet areas lead to extremely high values of $S_q(\lambda)$ for smaller λ, and this may break the linear behavior of (2.4). As a result, I expect lower performances in SC1 where irrigated pixels are present. To compare SC2 with SC2-R and SC3 with SC3-R, I calculated the percent difference of CC (α_{cc}) and RMSE (α_{RMSE}) with the metrics obtained in SC1. For example, for SC3, α_{cc} and α_{RMSE} were calculated as:

$$= \frac{\mathrm{CC_{sc3}} - \mathrm{CC_{sc1}}}{\mathrm{CC_{sc1}}} \cdot 100 \tag{2.5a}$$

$$\alpha_{\mathrm{RMSE}} = \frac{-(\mathrm{RMSE_{sc3}} - \mathrm{RMSE_{sc1}})}{\mathrm{RMSE_{sc1}}} \cdot 100, \tag{2.5b}$$

where $\mathrm{CC_{SC1}}$, $\mathrm{CC_{SC3}}$, $\mathrm{RMSE_{SC1}}$ and $\mathrm{RMSE_{SC3}}$ are the metrics computed in SC1 and SC3. In both cases, positive values of α_{cc} and α_{RMSE} indicate an improvement in the linearity of (2.4) as compared to SC1. Similar expressions of α_{cc} and α_{RMSE} were used for the other scenarios.

The Multifractal Downscaling Model

SC3 was designed to attenuate the spatial heterogeneity of θ due to irrigation, a required condition for applying downscaling models based on the hypothesis of spatial homogeneity. Here, I used the soil moisture fields generated for SC3 to apply the multifractal downscaling model that Mascaro et al. (Mascaro et al., 2011, 2010) applied in diverse climate and landscape settings to disaggregate θ in space from 25.6 km to 800

m. The multifractal model is based on a log-Poisson stochastic generator of homogeneous, random binary cascades that reproduce the observed scale-invariance and multifractal properties. The generator depends on two parameters, c and β, that control the theoretical expectation for the multifractal exponents (Deidda et al., 1999; Deidda, 2000):

$$K(q) = c \cdot \frac{q(1-\beta) - (1-\beta^q)}{\mathrm{In}2} \tag{2.6}$$

To estimate c and β, scale invariance is first verified through (2.4) and the observed multifractal exponents $K(q)$ are computed for different values of q as the slope of the regression lines. Next, (2.6) is fitted to the observed $K(q)$ and used to derive c and β. A pair of c and β is estimated for each aircraft-based θ field in a coarse domain.

As a next step, empirical calibration relations are identified between c and β and coarse-scale predictors. Previously, Mascaro et al. (Mascaro et al., 2011, 2010) found β to be fairly constant, with values changing from region to region. In contrast, c was linked to the coarse-scale mean soil moisture $<\theta>$. In previous applications, an exponentially decreasing relation:

$$c = c_\infty + \alpha \cdot e^{-\gamma \cdot <\theta>} \tag{2.7}$$

with parameters c_∞, a and γ was used in sub-humid and semiarid regions, while an increasing linear relation was used in a humid site. The regression parameters were linked to the principal components of ancillary factors, including soil texture, land cover, and topography. Here, I calibrated the downscaling model using (2.7) to interpret the relation between c and $<\theta>$.

Given its stochastic nature, the algorithm generates an ensemble of equally probable small-scale fields that are consistent with the same coarse-scale condition. Thus, to test its performances, I compared the Empirical Cumulative Density Function (ECDF) of the small-scale (1-km) observed θ fields with the 90% confidence intervals derived from an ensemble of 100 disaggregated fields generated by the calibrated model, as done in previous applications (Mascaro et al., 2011, 2010). In addition, I designed a new verification method based on ground θ data. For this aim, I hypothesized that the scaling regime (2.3), tested up to the aircraft footprint (l = 1 km; N_{lev} = 5), extends up to a scale representative of the ground measurements (l = 31.25 m; N_{lev} = 10). The calibrated model was then used to create an ensemble of 50 disaggregated fields at resolution l = 31.25 m. For each field, I extracted θ in pixels overlying the locations of ground measurements collected in non-irrigated areas. In such a way, the sizes of observed and synthetic samples are the same, but their mean values are different. Finally, the synthetic samples were utilized to build the 90% confidence intervals that were compared against the ECDF of the ground θ observations.

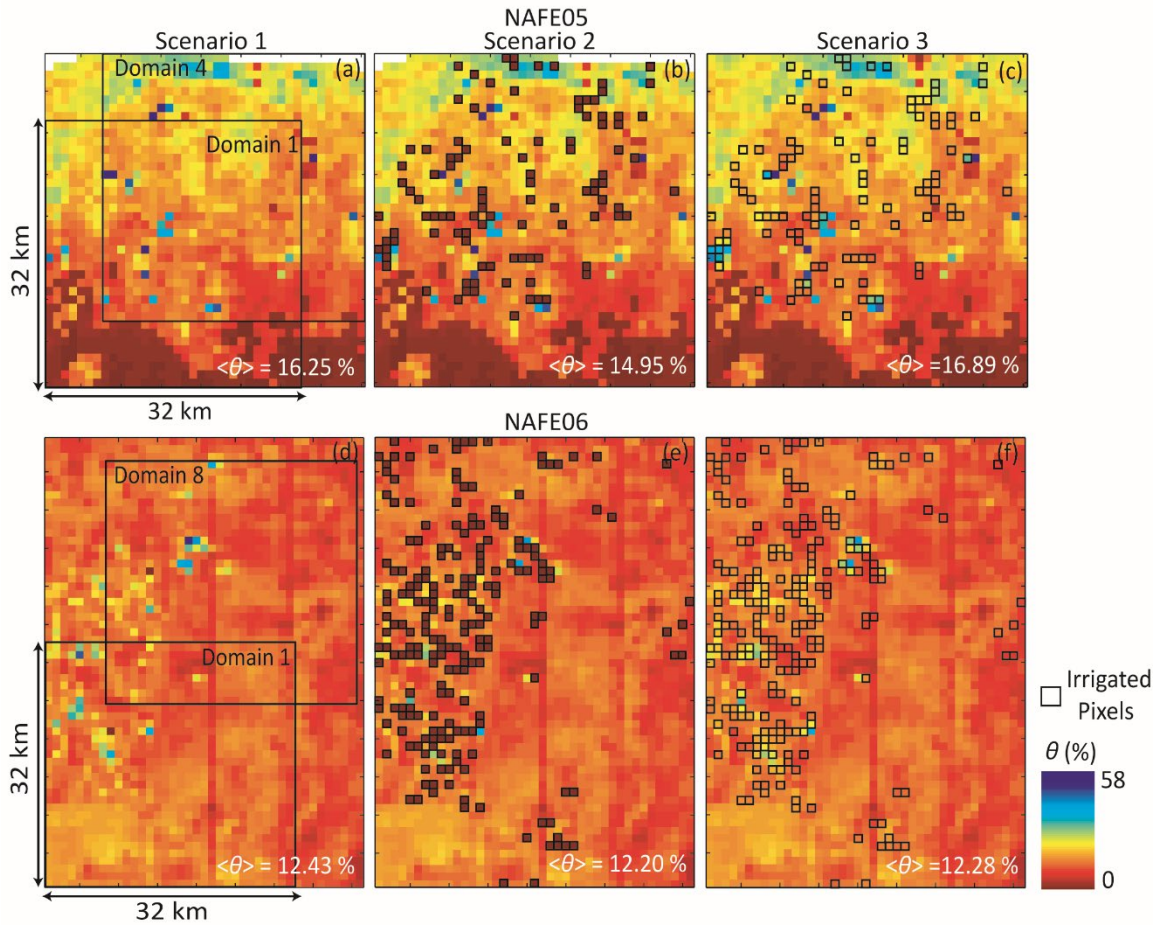

Figure 2.4. (a) SC1: original θ fields of NAFE05, with identification of four and eight coarse domains L x L (L = 32 km). (b) SC2: θ in irrigated pixels replaced with missing data. (c) SC3: θ in irrigated pixels replaced through interpolation from neighboring non-irrigated pixels. (d)-(f) Same as (a)-(c) but for NAFE06. The spatial mean soil moisture $\langle\theta\rangle$ is also reported in each panel.

Results

Effect of Irrigation on Scale Invariance Properties

I investigated the presence of scale invariance in the scale range from 1 to 32 km for four (eight) coarse-scale domains of NAFE05 (NAFE06). For NAFE05, I analyzed all four images to cope with the limited number of available days, while, for NAFE06, I excluded two days with rainfall (November 3 and 13) to better capture the effect of irrigation, thus retaining a total of nine images. Figure 2.5 shows examples of the scale invariance analysis for the SC1, SC2 and SC3 scenarios in domain 3 for each experiment. Similar findings were obtained for other days and domains. The presence of scale invariance was tested by verifying the linearity of (2.4) through the computation of RMSE and CC for q=4. As an example, Figure 2.6 summarizes results for domain 3 by showing the relations between these two metrics and $<\theta>$ in all available days. Table 2.3 reports the mean RMSE and CC across all days for each domain and scenario. Overall, the metrics indicate that: (i) performances of the regression quantified by CC are acceptable in all scenarios, with values larger than 0.9 as found in past studies [e.g., Hu et al., 1997; Mascaro et al., 2010, 2011)]; (ii) performances of the regression in SC1 are better in NAFE05 than NAFE06, especially in terms of RMSE, an indication that irrigation has a larger impact on the scale invariance analysis in the NAFE06 region; and (iii) performances tend to improve with increasing $<\theta>$, especially when focusing on RMSE (shown in Figure 2.6).

When comparing the three scenarios in NAFE06 [Figure 2.6(b) and (d) and Table 2.3], it is apparent that SC1 is characterized by degraded performances for the linear regression as compared to SC2 and SC3. This suggests that the presence of irrigated

pixels in this site impacts the scale invariance properties and can erroneously indicate the potential existence of two scaling regimes [e.g., on November 9 and 18 in domain 3 where a breaking point appears at $\lambda = 4$ km, arrows in Figure 2.5(d)]. If the irrigation pixels are removed as in SC2 and SC3, the linearity of (2.4) is more apparent and a single scaling regime emerges, as found in natural landscapes of other regions (Hu et al., 1997; Oldak et al., 2002; Mascaro et al., 2011, 2010). In contrast, in NAFE05 the metrics are quite similar across the three scenarios [see Figure 2.6(a) and (b) and Table 2.3]. This can be explained by: (i) sparseness of irrigated pixels; and (ii) overall wet conditions throughout the region that mask the effect of irrigated pixels, especially during rainy days. This is consistent with the preliminary analyses of Mascaro et al. (2011), who found that irrigation impacts more significantly the scale invariance of θ in semiarid areas with drier background (e.g., Arizona) as compared to more humid sites that are wetter throughout the region (e.g., Iowa).

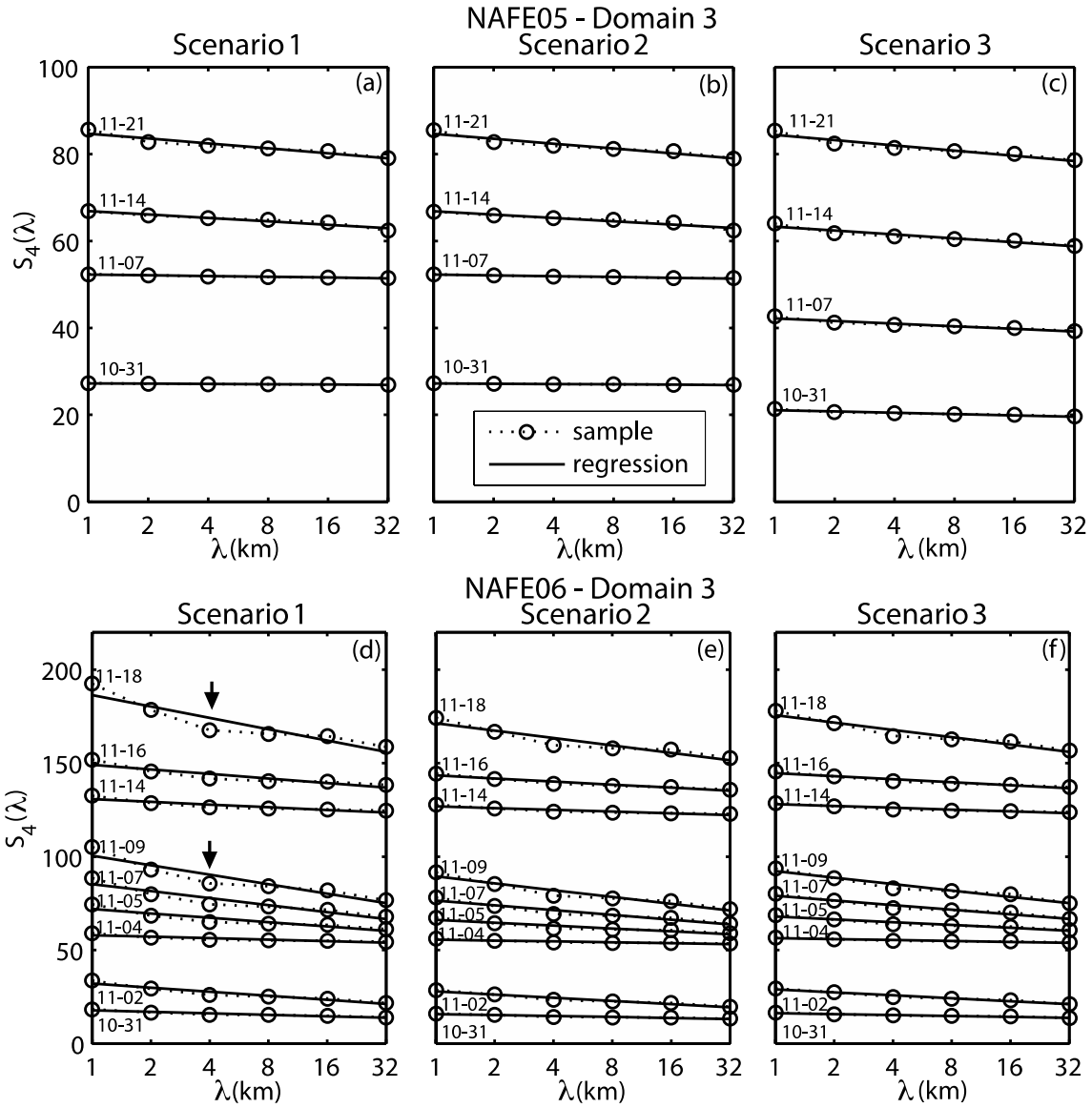

Figure 2.5. Scale invariance analysis for $q = 4$ in the three scenarios for domain 3 of NAFE05 (4 days; a-c) and NAFE06 (9 days; d-f). $S_4(\lambda)$ estimated on the θ fields are shown with circles, while regression lines are plotted in black. Arbitrary values in the y axis were used to display the lines of all days in a single plot. The arrows in (d) highlight the breaking points emerging in SC1 of NAFE06 on November 9 and 18. Dates are represented as MM-DD.

Table 2.3. Mean across all days of CC and RMSE computed from the scale invariance analysis in the three scenarios (SC1, SC2, SC3) for each domain of NAFE05 and NAFE06.

	CC (-)			RMSE (%)		
	SC1	**SC2**	**SC3**	**SC1**	**SC2**	**SC3**
Domain			**NAFE05**			
1	0.96	0.95	0.96	0.14	0.15	0.14
2	0.94	0.94	0.94	0.16	0.17	0.15
3	0.98	0.98	0.98	0.06	0.06	0.06
4	0.97	0.98	0.98	0.06	0.06	0.06
Domain			**NAFE06**			
1	0.93	0.96	0.97	0.32	0.04	0.03
2	0.95	0.97	0.98	0.30	0.04	0.02
3	0.93	0.97	0.97	0.30	0.04	0.02
4	0.96	0.98	0.99	0.24	0.03	0.04
5	0.94	0.95	0.97	0.37	0.07	0.03
6	0.95	0.96	0.97	0.35	0.07	0.06
7	0.94	0.96	0.97	0.39	0.06	0.04
8	0.95	0.97	0.98	0.35	0.06	0.03

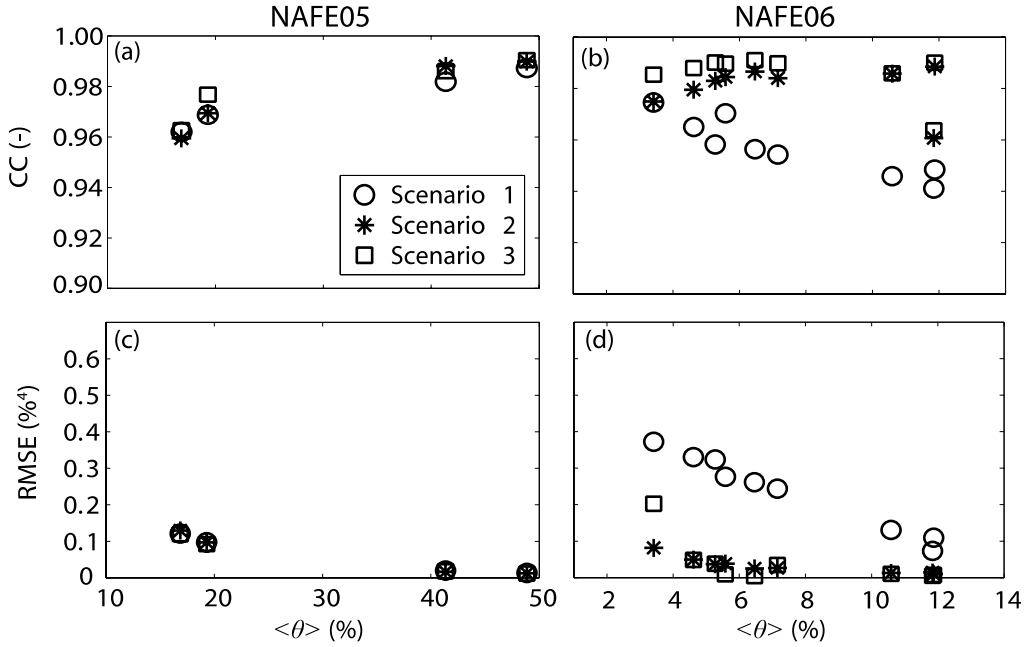

Figure 2.6. Relations CC versus <θ> and RMSE versus <θ> for domain 3 of (a) and (c) NAFE05 and (b) and (d) NAFE06. <θ> computed for SC2 is plotted in the *x*-axis for all scenarios.

Next, I tested the ability of SC2 and SC3 to reduce the spatial heterogeneity in θ as compared to the random removal of pixels performed in SC2-R and SC3-R. Figure 2.7 presents the relations between α_{RMSE} and $<\theta>$ for these cases, with the 90% confidence intervals of α_{RMSE} displayed for SC2-R and SC3-R. In NAFE06, α_{RMSE} for SC2 and SC3 lies always outside the 90% confidence intervals of SC2-R and SC3-R. This is a clear indication that replacement of θ in irrigated pixels, either via substitution with missing data (SC2) or interpolation (SC3), leads to a statistically significant improvement in the linear regression as compared to removals in random locations. In contrast, the improvement in NAFE05 is not statistically distinguishable. Similar results were found for α_{cc} (not shown). These findings demonstrate the utility of a quantitative framework based on statistical metrics and tests for evaluating the effect of irrigation on θ spatial variability. This approach can also be applied to evaluate the impact of other sources of spatial heterogeneity.

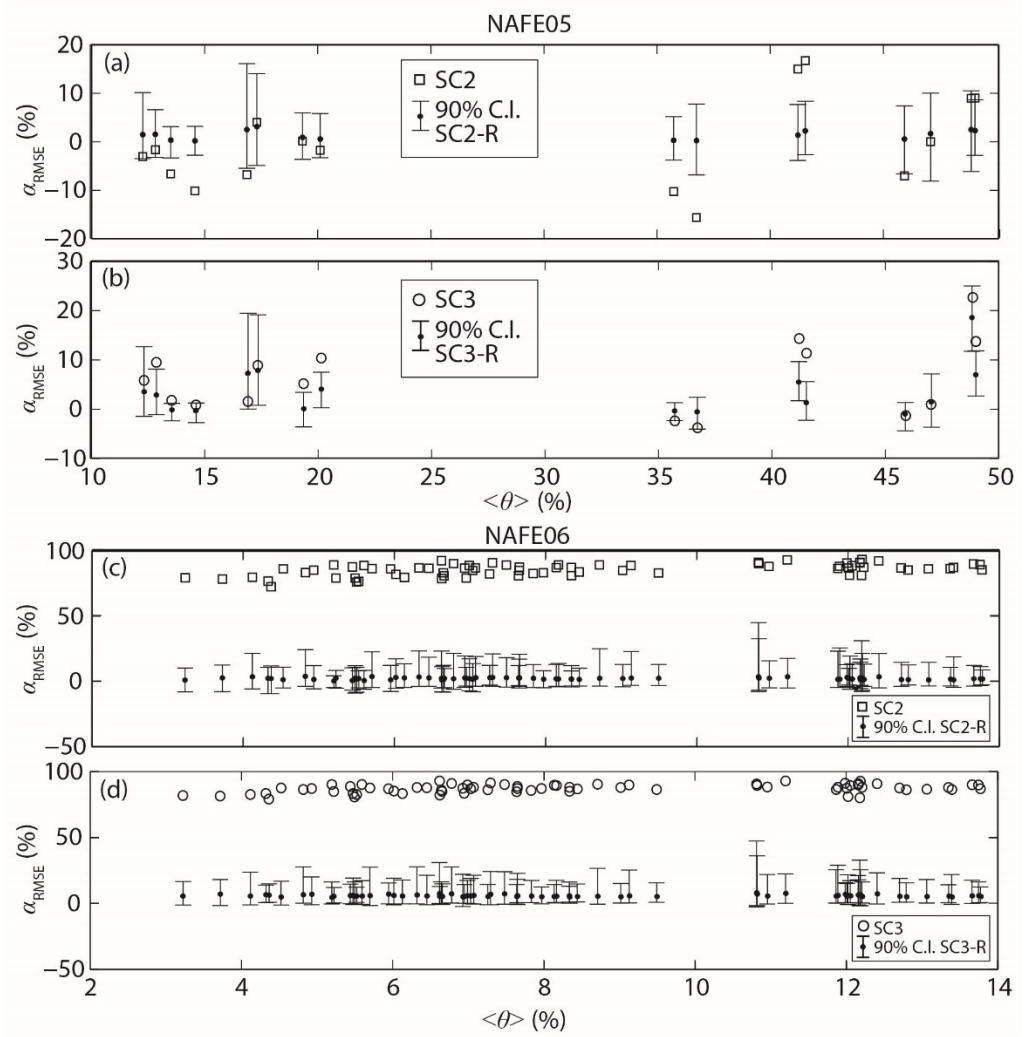

Figure 2.7. (a) and (b) Relations α_{RMSE} vs. $\langle\theta\rangle$ for RMSE in (a) SC2 and SC2-R and (b) SC3 and SC3-R of NAFE05. (c) and (d) Same as (a) and (b) but for NAFE06. In all panels, bars are the 90% CIs of the ensemble values of α_{RMSE} computed for SC2-R and SC3-R. $\langle\theta\rangle$ computed for SC2 is plotted in the *x*-axis for all scenarios.

Downscaling Soil Moisture Fields with Removal of Irrigated Croplands

1) Model Calibration: For each day and domain, parameters c and β of the downscaling model were estimated by fitting the theoretical expectation for $K(q)$ to the observed values. Figure 2.8 presents results for all days in two representative domains with contrasting behavior in each experiment. In all cases, the relations between $K(q)$ and q are non-linear, thus indicating the presence of multifractality. In addition, Figure 2.8 shows that θ fields with lower $<\theta>$ have higher K(q), both when comparing: (i) different domains within each experiment, and (ii) the wetter conditions of NAFE05 with the drier states of NAFE06. Based on (2.4), one can notice that larger $K(q)$ are obtained when $S_q(\lambda)$ increases faster as λ decreases, a circumstance that occurs when the field sampled at smaller scales is characterized by uneven and sudden variations (i.e., more intermittent). As a result, our findings reveal that the θ spatial distribution in drier (wetter) domains tends to be more intermittent (smoother), consistent with previous studies by Hu et al. (1997), Oldak et al. (2002), Das and Mohanty (2008), and Mascaro and Vivoni (2010).

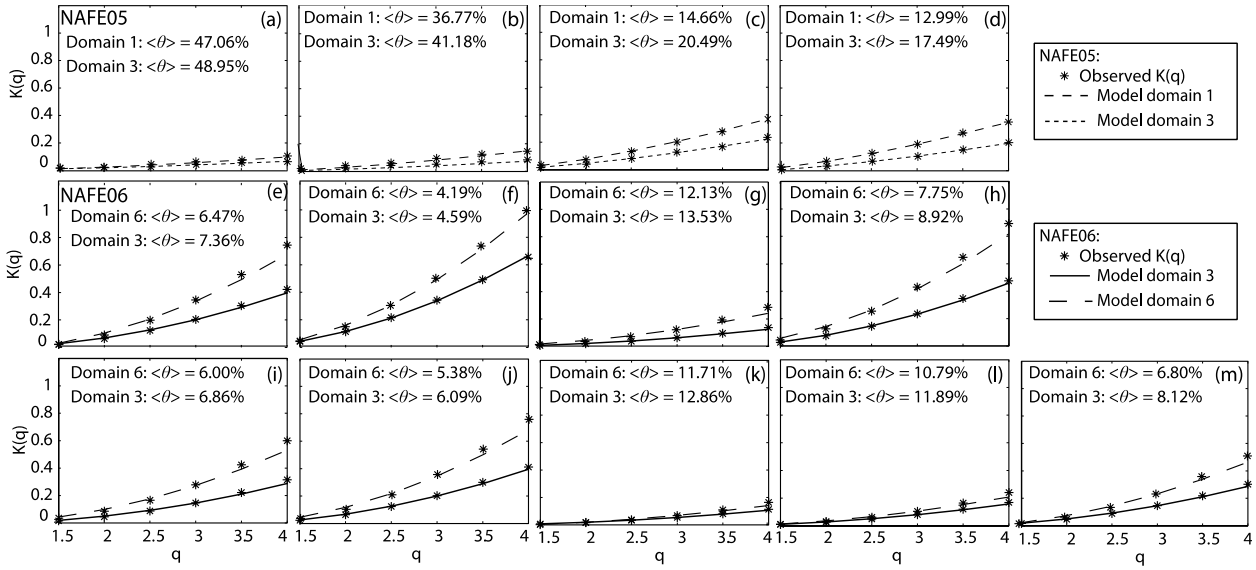

Figure 2.8. Relation between multifractal exponents $K(q)$ and moments q in SC3 for two representative domains of NAFE05 [domains 1 and 3; panels (a)-(d)] and NAFE06 [domains 3 and 6; panels (e)-(m)] for all available days. $\langle\theta\rangle$ is also reported in each panel.

For each experiment, the values of β estimated in all domains and days were found to be similar, while the c estimates were more variable. In NAFE06, β averages 0.96 with a standard deviation of 0.07, while in NAFE05 the mean β is 0.68 with higher standard deviation of 0.16, likely due to the marked differences in wetness conditions of the four images. To reduce the dispersion of the c estimates, I assumed a single value of β in each experiment, equal to the corresponding mean, and re-estimated c. The resulting values of c ranged from 0.09 to 0.56 in NAFE05 and from 7.00 to 72.57 for NAFE06. These findings are similar to other study areas (Mascaro et al., 2011, 2010), with values of c and β varying with the region. Specifically, β was close to 0.85 in a sub-humid climate in Oklahoma, 0.71 in the semiarid sites in Arizona and Sonora, and 0.89 in the humid region in Iowa. The range of c in NAFE05 is comparable with the sub-humid and semiarid sites, whereas the c values in NAFE06 vary in a larger range. This can be due

to: (i) the specific land use and soil properties that affect the soil moisture distribution and retrieval algorithm, and (ii) the numerical effect of assuming a relatively large β (0.96) in (2.4) to estimate c.

The downscaling model was calibrated by assuming a constant, regional value of β and using (2.7) to interpret the relation between c and $<\theta>$. In NAFE05, a single relation is able to represent the variability of the c estimates for all domains and days (see Figure 2.9) and was then labeled as the regional (REG) calibration approach. The single set of parameters, c_∞, a and γ, estimated through non-linear fitting, is reported in Table 2.4. In NAFE06, a single calibration relation is not able to fully capture the dispersion of the c values of all domains (i.e., different values of c_∞, a and γ). Following Mascaro et al. (2011), I used another calibration strategy based on ancillary data and labeled as ANC. Based on empirical evidence regarding the variation of c_∞, I also assumed a fixed c_∞ = 2.5 and re-estimated a and γ. Table 2.2 summarizes results of the estimation of c_∞, a and γ in NAFE06 with free and fixed c_∞. As a next step, I applied principle component analysis (PCA) to a total of six ancillary factors computed for each domain (Table 2.2), including two topographic variables (mean elevation and slope) and four vegetation classes (percentage of rainfed crops, rainfed pasture, tussock grass, and woodland). Soil properties were not used since a single soil type occupies almost the entire region. By an orthogonal rotation of the analyzed dataset, the PCA identifies a minimal number of principal components (PCs) that explain the largest variability. I found that the first 3 PCs are able to explain 98% of the variability. The scores of the PCs (PCS_j, j = 1, 2, 3) were linked to a and γ with a multilinear regression as:

$$\begin{cases} \alpha = t_0 + t_1 \mathrm{PCS}_1 + t_2 \mathrm{PCS}_2 + t_3 \mathrm{PCS}_3 \\ \gamma = v_0 + v_1 \mathrm{PCS}_1 + v_2 \mathrm{PCS}_2 + v_3 \mathrm{PCS}_3 \end{cases} \quad (2.8)$$

with parameters t_j and v_j that were estimated by substituting (2.8) in (2.7). Estimates of the regional parameters t_j and v_j are reported in Table 2.5, while results of the ANC calibration are reported in Figure 2.10, along with the regression lines estimated locally (LOC) at each domain using parameters of Table 2.5 with c_∞ = 2.5. Even if ANC has lower performances than LOC, the results show the ANC calibration method captures the local behavior at all domains well, with lower skills in domains 2, 4 and 5 (see RMSE reported in each panel).

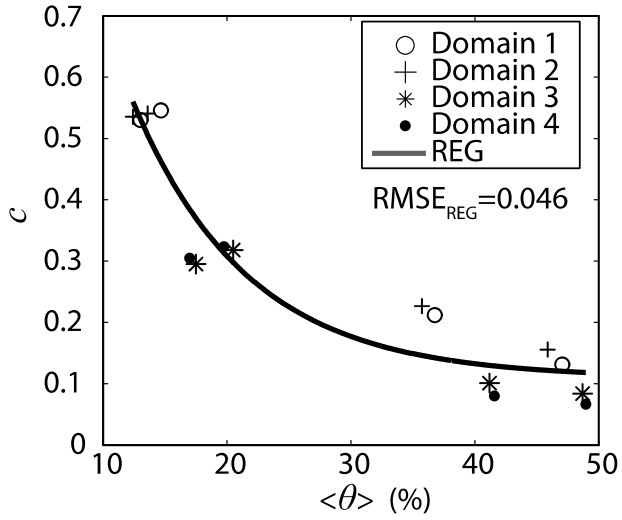

Figure 2.9. Regional (REG) calibration relation for NAFE05, with the RMSE between REG line and points.

Table 2.4. Estimates of c_∞, a, and γ. For NAFE05, the values are reported for the case with all domains pooled together. For NAFE06, the values are reported for each domain with free c_∞ and fixed c_∞.

	Free c_∞			Fixed c_∞		
Domain	c_∞	γ	a	c_∞	γ	a
		NAFE05				
All	0.1097	0.1079	1.717	-	-	-
		NAFE06				
1	5.10	0.28	137.9	2.5	0.24	125.0
2	6.34	0.31	133.3	2.5	0.26	118.1
3	2.56	0.23	132.5	2.5	0.23	132.3
4	3.26	0.26	133.4	2.5	0.25	130.6
5	3.75	0.18	165.4	2.5	0.21	183.9
6	-3.80	0.19	174.1	2.5	0.23	188.7
7	-2.85	0.19	161.4	2.5	0.22	178.2
8	-2.64	0.21	173.6	2.5	0.24	187.6

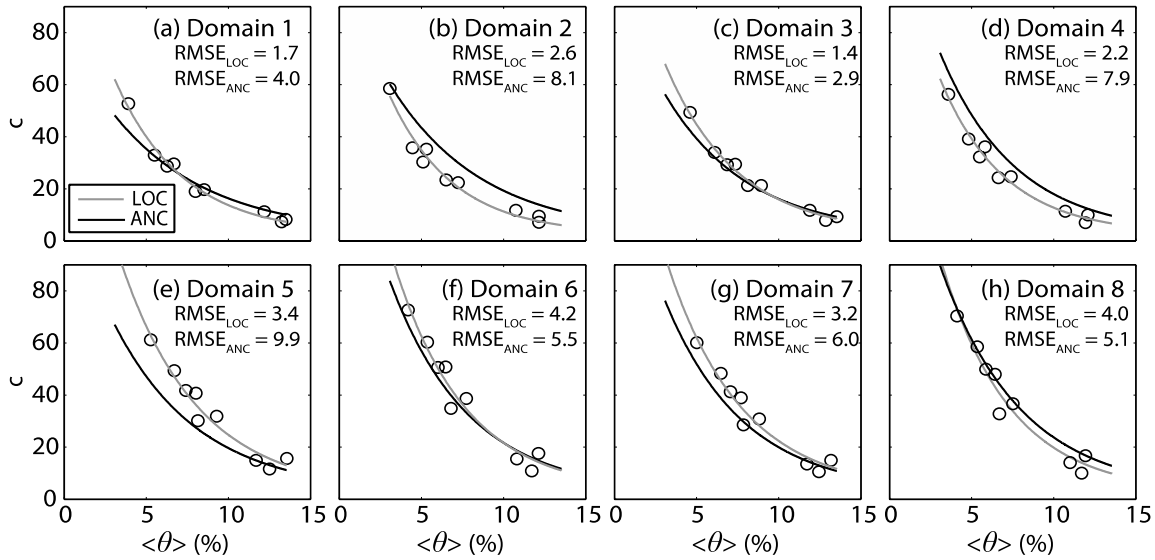

Figure 2.10. Ancillary (ANC) and local (LOC) calibration relations for NAFE06, with RMSEs between LOC and ANC lines and points ($RMSE_{LOC}$ and $RMSE_{ANC}$ respectively).

Table 2.5. Parameters of the ANC calibration relation.

Parameter	Value
t_0	123.633
t_1	4.600
t_2	0.925
t_3	8.241
v_0	0.198
v_1	0.002
v_2	-0.003
v_3	-0.003

2) Model Validation Against Aircraft Data: Model performances were assessed by comparing the observed ECDF of the small-scale aircraft-based θ fields in SC3 with the 90% confident intervals estimated from an ensemble of 100 synthetic fields. To help interpreting this verification approach, Figure 2.11 shows four examples where the model exhibits different performances. Given the statistical nature of the downscaling scheme, the test evaluates the hypothesis that observed and disaggregated fields are equally probable realizations of the same statistical distribution. Such hypothesis cannot be discarded if the exceedance probability $S(\theta) = 1 - F(\theta)$ of the observed distribution, S_{obs}, is included between the minimum and maximum exceedance probabilities of the 90% confidence intervals, S_{min} and S_{max}, at least 90% of the time. This statistical test is described in detail in Mascaro et al. (2008). Figure 2.11(a) shows two cases where the condition $S_{min} \leq S_{obs} \leq S_{max}$ is verified in about 95% of the cases, indicating high model performances. In case B, note that: (i) ~3% of the observed θ values are equal to the maximum of 58%, and (ii) the observed soil moisture value is smaller than the maximum value of the left confidence interval, so that $S_{min} = 0$. In contrast, Figure 2.11(b) and (c) present two cases with low performances. In Figure 2.11(b), while $S_{min} \leq S_{obs} \leq S_{max}$ in

about 60% of the cases, the model fails to capture the right, heavy tail of the observed distribution, where $S_{obs} > S_{min} = S_{max} = 0$. The presence of a heavy tail is indicative of a heterogeneous observed field with wet irrigated areas. Finally, Figure 2.11(c) shows a case where the observed and disaggregated fields do not have the same mean soil moisture value at the coarse scale and the model is positively biased, so that $S_{obs} > S_{min} = S_{max} = 0$.

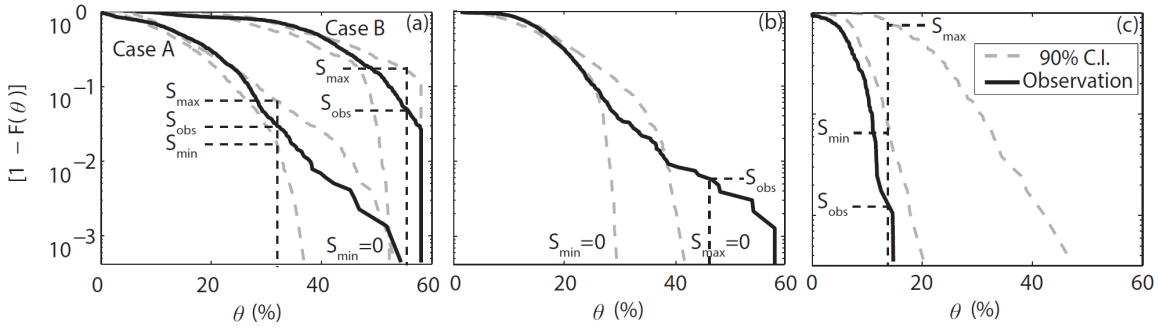

Figure 2.11. Examples of comparison between the ECDF of observed small-scale θ against the 90% CIs derived from the downscaling model in cases with (a) high and (b) and (c) low performances. In each case, an example of the derivation of the exceedance probabilities for the observed distribution, S_{obs}, and the 90% CIs, S_{min} and S_{max} is also reported.

For NAFE05, the synthetic fields were generated using the REG calibration relation: for each domain, the coarse-scale mean soil moisture $<\theta>$ was used to derive c from (2.7) with parameters of Table 2.4 and 100 disaggregated fields were produced through the Log-Poisson stochastic generator. Figure 2.12(a)-(h) show examples of the ECDF comparison of observed and disaggregated θ fields for all four days in domain 1 and 3, selected because of contrasting land cover conditions. In wet days (October 31 and November 7), the downscaling model captures very well the shape of the observed ECDF that contains several values equal to the maximum threshold of 58%. Model performance degrades in drier conditions (November 15 and 21), especially in domain 3. For NAFE06,

the synthetic disaggregated fields were generated with the ANC calibration relation. For each domain, the coarse predictors $<\theta>$ and PCS_j derived from the ancillary data were utilized to estimate c from (2.7) and (2.8) and produce an ensemble of 100 disaggregated fields. Figure 2.12(i)-(p) present examples for four days with different wetness conditions in domain 2 and 5. Overall, the model is able to capture the observed ECDFs adequately in most cases, with a slightly better performance in drier days [e.g., Figure 2.12(i), (k), and (m)].

A more in-depth comparison between observed and downscaled ECDFs reveals that model performances tend to be lower in domains with a larger percentage of irrigated pixels. This is true for NAFE06 independently of the mean wetness, while it occurs in drier conditions for NAFE05. In these cases, the model is not able to capture the right tail of the observed distribution that is characterized by anomalously larger θ, as in the illustrative case of Figure 2.11(b). For example, this occurs for domain 3 of NAFE05 (9.7% of irrigated pixels and low $<\theta>$) on November 14 and 21 and for domain 5 of NAFE06 (14.0% of irrigated pixels) on November 4 and 9. Most likely, lower performances in these cases are due to the fact that the method proposed to detect irrigated croplands may have not been fully effective, so that pixels affected by irrigation are still present in SC3. Possible reasons limiting the skill of this approach are: (i) deficiencies of SAVI and BR indices and selected thresholds; (ii) scale mismatch between Landsat TM5 (30 m) and soil moisture (1 km) images, and (iii) difference of more than 10 days between Landsat TM5 and soil moisture images.

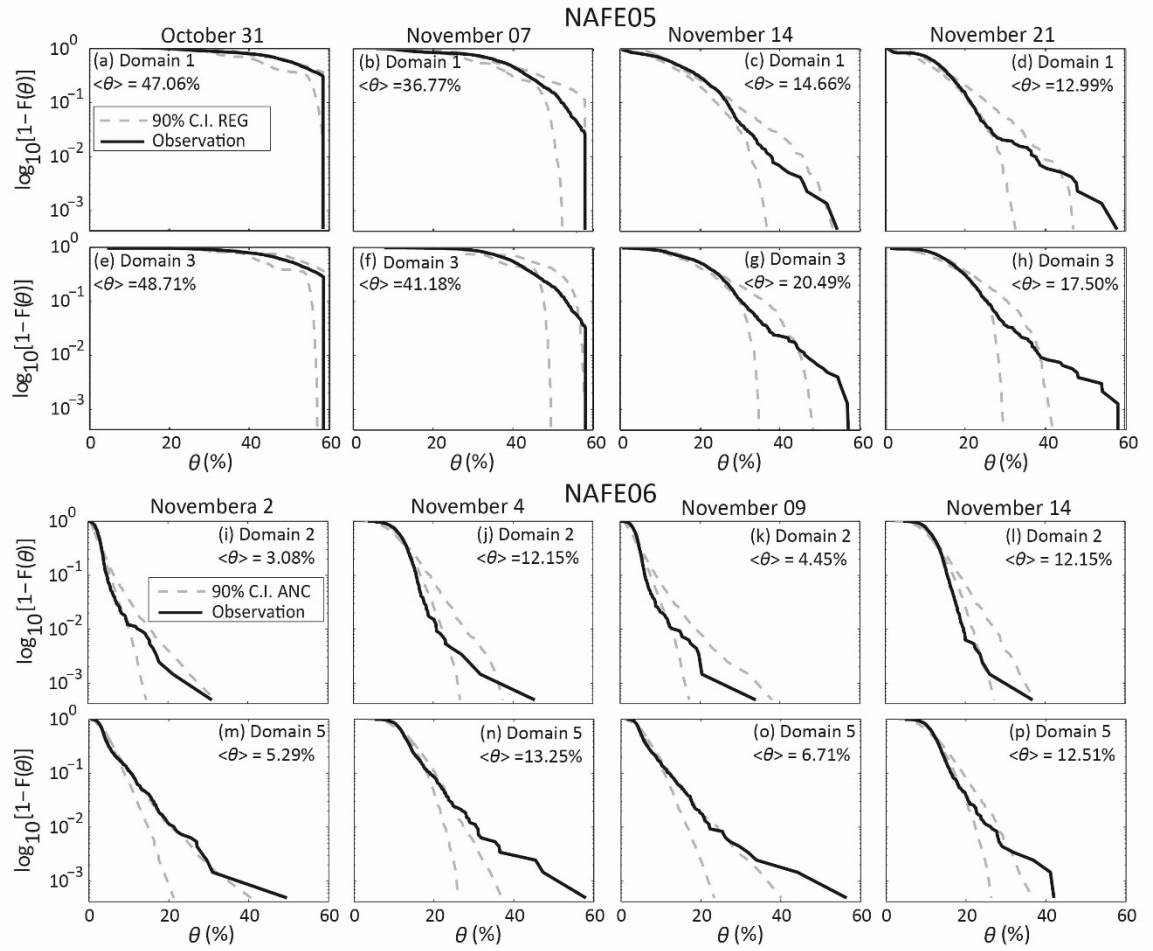

Figure 2.12. Comparison of the ECDFs of observed small-scale θ against the 90% CIs derived from the calibrated downscaling model in NAFE05 (REG calibration) and NAFE06 (ANC calibration) for selected domains and dates.

3) Model Validation Against Ground Data: An additional verification was conducted by comparing the ECDF of ground θ measurements with the 90% confidence intervals estimated from an ensemble of 50 synthetic fields generated under the assumption that the scaling regime extends up to l = 31.25 m, a scale representative of the ground samples. Figure 2.13(a)-(d) show results for NAFE05 in domain 4 that encompasses the largest number of ground samples. In the wet days of October 31 and November 7, the model captures about 50% of the data in the left tail of the distribution, while it overestimates the larger θ values. On November 14, the model fails in capturing

the shape of the observed ECDF, while on November 21 it overestimates the θ observations, as in the example of Figure 2.11(c). The poor performances revealed by this test indicate that extending the presence of scale invariance up to l = 31.25 m is not a valid assumption in this area. Here, the downscaling model is able to adequately reproduce the θ variability only up to l = 1 km [Figure 2.12(a)-(h)]. In contrast, in NAFE06 the model exhibits a very good ability to capture the observed ECDFs across different wetness conditions, as reported in the examples of Figure 2.13(e)-(h) for domain 1. These findings sustain the hypothesis that the scaling regime can be extended in this region and confirm the high performances of the downscaling model across a wide range of scales from the satellite to the ground measurement footprints.

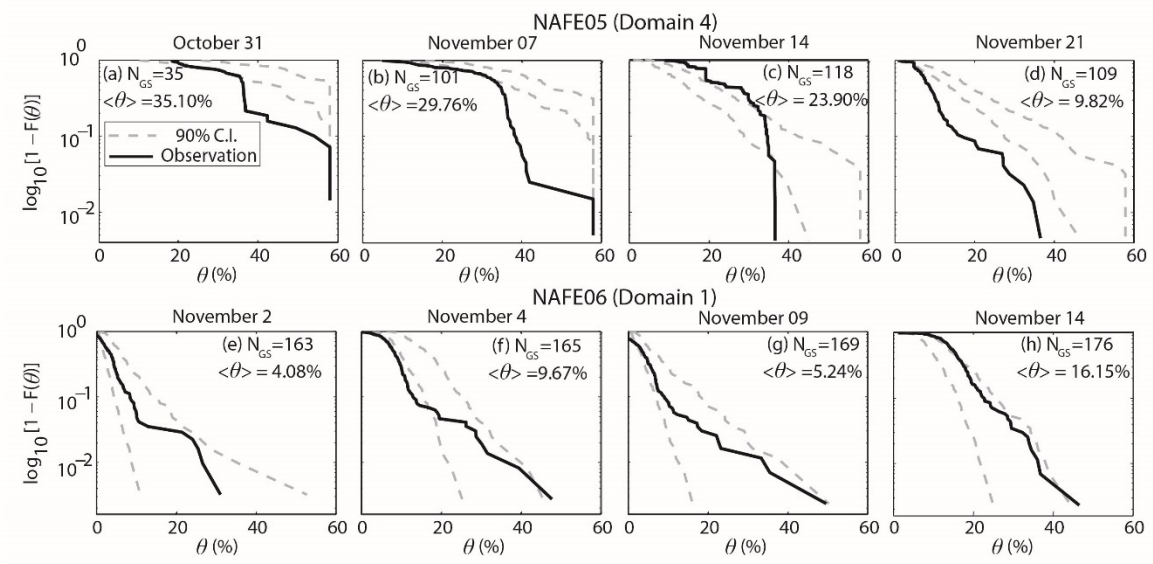

Figure 2.13. Comparison of the ECDFs of observed ground θ against the 90% CIs derived from the calibrated downscaling model in NAFE05 (REG calibration) and NAFE06 (ANC calibration) for selected domains and dates. In each panel, the number of ground samples N_{GS} is also reported.

Discussion

The analyses conducted in this study provide three main novel contributions. First, I proposed a simple method for the time-varying detection of irrigated croplands from remote sensing imagery, a necessary requirement to account for the presence of irrigation in any type of downscaling schemes. Second, I developed a quantitative framework based on statistical metrics for evaluating the impact of irrigated areas on the statistical properties of soil moisture and attenuating their effect on the calibration of a downscaling model. This framework can be applied or easily adapted to investigate the effect of different sources of spatial heterogeneity, such as the presence of built areas. By applying this framework, I found that the climate of the region and the extent and spatial distribution of the irrigated areas control the impact of irrigation on the scale invariance properties of θ. In the NAFE06 region, climate is semiarid and irrigated pixels tend to be distributed in large, connected areas. Under these conditions, larger θ values in irrigated croplands contrast with a drier background, thus affecting the scale invariance analysis. In contrast, in the more humid NAFE05 site, conditions are wetter throughout the region and irrigated pixels are more scattered, leading to negligible impacts on the scale invariance analysis. These analyses confirm preliminary results of Mascaro et al. (2011). Third, I found that incorporating the presence of irrigation in downscaling schemes of satellite products is more important in semiarid regions, where irrigation creates a drastic soil moisture contrast with the surrounding areas.

Previous studies of Mascaro et al. (2010, 2011) and findings of this work suggest that scale invariance and multifractality are statistical properties exhibited by the spatial distribution of θ in a wide variety of natural landscapes and wetness conditions, over the

scale range that goes from the satellite (~30 km) to the aircraft (~1 km) footprints. As such, downscaling approaches that reproduce these properties can be potentially applied at all locations and times. For example, the multifractal downscaling model used here has been successfully applied in areas with markedly different climatic and environmental settings, where high-resolution, aircraft-based θ data were available (2010, 2011), If one wants to apply the downscaling model at a new site, the calibration relations published for a region with similar conditions could be adopted as a first level of approximation. Clearly, as for any type of downscaling method, additional high-resolution θ data would be highly desirable to improve the calibration relations at the new location. A strategy to generate these datasets could be the use of distributed hydrologic models applied at high resolution over large basins (e.g., Mascaro et al., 2015). In addition, in regions with the presence of irrigation, it is necessary to account for this source of spatial heterogeneity. The framework proposed in this chapter for detecting the irrigated croplands via remote sensed imagery and attenuating their effect can be used for this goal.

Conclusion

I investigated the impact of irrigated cropland on the scaling properties of θ and the application of a multifractal downscaling model using aircraft- and ground-based data from the NAFE05 and NAFE06 campaigns. To our knowledge, this is the first study to focus on the effects of irrigation on soil moisture variability and disaggregation. I proposed a relatively simple approach to detect irrigated areas based on a single Landsat TM5 scene taken a few days before the beginning of each experiment. Through this information, I attenuated the source of spatial heterogeneity in the θ distribution. I found that irrigation

has a larger impact on the scale invariance properties of θ in the semiarid NAFE06 region, where irrigated pixels are distributed in a large high-density district, while it is negligible in the more humid NAFE05 site, where irrigated pixels are more scattered. Subsequently, the θ fields that were modified to attenuate the impact of irrigated croplands were utilized to calibrate a multifractal downscaling model based on the hypothesis of spatial homogeneity. The model was calibrated as a function of readily available coarse-scale predictors using regional and ancillary factor approaches. A verification test based on aircraft θ data revealed satisfactory performances in most cases in both experiments, except in conditions where the method to identify irrigated croplands may have not been fully effective. In these cases, anomalously larger θ values due to irrigation are still present in the observed distribution and cannot be reproduced by the homogeneous multifractal cascades of the disaggregation scheme. The model was further verified against ground θ observations according to a new method based on the hypothesis that the scaling regime extends up to a scale representative of these measurements. Performances were very good in the NAFE06 region, implying that the downscaling model is potentially capable to reproduce the θ variability from the satellite to the ground measurement footprints. In contrast, the model fails in simulating the ground θ variability in the NAFE05 area, thus suggesting its applicability only up to the aircraft footprint.

CHAPTER 3

STRATEGIES TO IMPROVE PHYSICS-BASED HYPERRESOLUTION HYDROLOGIC SIMULATIONS AT REGIONAL BASIN SCALES

Introduction

During the last decade, the increasing availability, coverage and resolution of hydrometeorological data (Mitchell et al., 2004; Dee et al., 2011; Harris et al., 2014; Maidment, 2016; Do et al., 2018), geospatial datasets characterizing land surface properties (Lehner et al., 2008; Hengl et al., 2014), and remotely-sensed products from numerous satellite missions have stimulated the feasibility of numerical simulations of the water cycle at hyperresolutions, defined by (Wood et al., 2011) as <100 m for regional scales and <1 km for continental and global scales. Achieving these high spatial resolutions in hydrologic simulations is expected to support a large range of water-related applications, including weather forecasting (Senatore et al., 2015), climate prediction (Baker et al., 2017), irrigation scheduling (Gibson et al., 2017), quantification of greenhouse gas fluxes (Franz et al., 2017), flood prediction (Maidment, 2017), and estimation of water scarcity (Zhou et al., 2016). Moreover, hyperresolution simulations have the potential to improve our understanding of hydrologic processes, by identifying the physical controls on the spatiotemporal dynamics of fluxes and state variables of the water and energy balances (Bierkens et al., 2015; Mascaro et al., 2015).

Despite the expected benefits, a number of challenges still limit the ability to carry out accurate hydrologic simulations at hyperresolutions, as described by Wood et al. (2011) and Bierkens et al. (2015). Three critical challenges are here summarized. First, it is necessary to evaluate whether the representation of physical processes in current land

surface and hydrologic models, designed to operate at coarser scales, are appropriate at hyperresolutions. Second, as pointed out by Beven and Cloke (2012), increasing the resolution of hydrologic simulations leads to higher epistemic uncertainties, because the available information on hydrometeorological forcings and land surface properties decreases as I transition to smaller scales. Strategies are needed to (i) generate reliable high-resolution fields of hydrometeorological variables, as well as of terrain, soil and vegetation properties; and (ii) thoroughly test the large datasets of simulated time series and spatial patterns against independent observations. Third, increasing the domain resolution (i.e., the number of cells) rapidly increases the demand of computational resources. An effort should be placed to (i) take advantage of high performance computing clusters; and (ii) optimize storage of and access to the large datasets generated by these simulations.

During the last few years, a significant effort to address these challenges and achieve hyperresolutions has been devoted to the development of new or adaptation of existing land surface and global hydrologic models, which have been applied at continental and global scales often as part of climate and numerical weather predictions models (e.g., Singh et al., 2015; Maidment, 2016; Cai et al., 2017). Physics-based distributed hydrologic models (DHMs), which were originally designed to represent hydrologic processes at high spatial and temporal resolutions (e.g., Ivanov et al., 2004a; Kollet and Maxwell, 2006; Camporese et al., 2010), have been mainly applied at field, hillslope and small catchment scales (e.g., Mahmood and Vivoni, 2011; Pierini et al., 2014; Paniconi and Putti, 2015; Fatichi et al., 2016). The recent adoption of parallel computing in hydrology has opened the door to the use of DHMs at hyperresolutions over

regional watersheds ($O(10^3)$ km^2 and higher). For example, Kollet et al. (2010) demonstrated on synthetic domains the feasibility of hyperresolution simulations at regional scales with a parallel, three-dimensional, variably saturated groundwater flow model with integrated overland flow. Despite these promising results, most simulations conducted with DHMs in real-world regional basins have been at resolutions of ~500 m (e.g., Shen and Anagnostou, 2017; Koch et al., 2017), thus not reaching the target of 100 m defined by Wood et al. (2011). To our knowledge, only Mascaro et al. (2015) have applied a DHM to a real-world regional basin of 3796 km^2 in northwest Mexico–the Río San Miguel basin–at the hyperresolution of ~78 m over a period of 7 years.

The main objective of this study is to improve the accuracy of hyperresolution simulations of physics-based DHMs over regional watersheds. For this aim, I build upon results of Mascaro et al. (2015) (MA15, hereafter) and expand their work to design strategies that address some of the challenges mentioned above. I significantly increase the size of the study basin of MA15 by simulating the hydrologic response of the Río Sonora basin (RSB), a regional watershed of 21,237 km^2 that includes the Río San Miguel basin as one of its sub-watersheds. The RSB represents a compelling test case for hyperresolution simulations due to the significant climate seasonality dominated by the North American monsoon, its complex topography, the high spatiotemporal variability of vegetation properties, and the limited and sparse datasets available to setup and apply DHMs. I conduct simulations over 10 years at a nominal resolution of ~88 m on a domain of more than 2.6 million nodes, using the physics-based DHM known as TIN-based Real-time Integrated Basin Simulator (tRIBS, Ivanov et al., 2004a, 2004b). MA15 showed that the tRIBS model is a computational efficient physics-based model suitable

for hyperresolution simulations, since (i) it is based on an irregular domain, which allows capturing the details of complex topography with minimal loss of information as compared to grid-based models (Vivoni et al., 2005); (ii) it uses approximate solutions in the infiltration and routing schemes (Cabral et al., 1992); and (iii) its code has been parallelized to be run in high performance computing clusters (Vivoni et al., 2011).

I pursue three specific tasks to address our main objective. First, I generate reliable high-resolution meteorological forcings by (i) integrating reanalysis products at 12-km resolution with ground observations, and (ii) applying downscaling techniques that account for terrain properties to increase the resolution to 1 km. Second, I develop soil parameter grids through a strategy that involves model calibration against soil moisture data at distributed locations, use of pedotransfer functions, and the integration of coarse (~6 km) local and high-resolution (250 m) global soil datasets. Third, I extensively test the model performance by comparing simulated outputs against independent observations, including time series of soil moisture recorded at a network of stations and spatial patterns of land surface temperature (LST) estimated by remote sensors. This is a crucial step to assess the ability of DHMs to accurately represent physical processes (e.g., Melsen et al., 2016) and to implicitly evaluate our effort to increase quality and resolution of forcings and parameters. I place particular emphasis on the systematic validation of the simulated spatial patterns of LST, which so far has been done in a limited number of cases, often focusing on small basins (e.g., Xiang et al., 2014; Koch et al., 2015) or using only a few images (e.g., MA15; Xiang et al., 2017), but that has been receiving increasing attention (Zink et al., 2018). In doing so, I apply a set of tools, which have been recently adopted to validate and interpret spatial outputs of DHMs (MA15; Koch et

al., 2016a), to quantify model skill and diagnose the potential causes of model deficiencies. While our analyses are based on a specific DHM, the strategies proposed here can be adopted for the application of other hydrologic models in different settings and, thus, they provide valuable support for the scientific community focused on hyperresolution hydrologic modeling.

Study Area and Datasets

Climate and Physiographic Properties of the Río Sonora Basin

The hyperresolution simulations were conducted in the Río Sonora basin (RSB) located in northwest Mexico [Figure 3.1(a)]. The RSB drains an area of 21,237 km^2 within a rural region in the state of Sonora characterized by complex terrain with mountain ranges and ephemeral rivers (Vivoni et al., 2007b; Méndez-Barroso et al., 2009). As shown in Figure 3.1(b), the main river, Río Sonora, flows from northeast to southwest up to the city of Hermosillo, after receiving the contribution of the two major tributaries, Río San Miguel and Río Zanjón. The basin includes areas with climate classified as arid and semiarid, with mean annual temperature (precipitation) varying from 11 to 29 ºC (350 to 750 mm) depending on the location. The precipitation regime is characterized by: (i) large interannual variability, which is partially explained by teleconnections (Gochis et al., 2006); (ii) high spatiotemporal variability within each season due to the interactions of different weather systems and complex orography (Gebremichael et al., 2007; Johnson et al., 2007; Rowe et al., 2008); and (iii) marked seasonality, with a wet summer season from July to September dominated by the North American Monsoon (NAM), when approximately 40-80% of the total annual

precipitation falls mainly in form localized convective thunderstorms (Vivoni et al., 2010; Mascaro et al., 2014), and a relatively dry winter period characterized by the occurrence of a few widespread precipitation events caused by frontal systems of subtropical or polar origin (Brito-Castillo et al., 2003).

As shown in the digital elevation model (DEM) of Figure 3.1(b) derived from the Shuttle Radar Topography Mission (SRTM), elevation varies significantly between a minimum of 200 m and a maximum of 2618 m (mean of 858 m). Topography is rugged, as indicated by the slope ranging from 0° to 77° (mean of 8°) and the presence of incised stream channels. Due to the complex terrain properties and the effect of the NAM, ecosystems in the RSB exhibit significant spatial variability and strong seasonality. The high spatial heterogeneity of the vegetation characteristics is shown in Figure 3.1(c), which presents the land cover map provided by the Instituto Nacional de Estadística y Geografía (INEGI). As reported in Table 3.1, the dominant classes are shrubland (22.1%), desert scrub (20.7%), mesquite (18.4%), and pasture (14.0%). The majority of the ecosystems are characterized by leaf greening in mid-July and senescence in late September following the onset and decay of the NAM, as documented in detail by Vivoni et al. (2007b, 2008b) and Méndez-Barroso and Vivoni (2010).

Figure 3.1(d) shows the soil map provided by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) that follows the Food and Agriculture Organization (FAO) soil classification. The major soil types are Regosol, Lithosol, and Xerosols accounting for ~76% of the basin area. The CONABIO soil map has been used in previous studies in the region to parameterize the tRIBS (e.g., Robles-Morua et al., 2015; MA15) and WRF (Weather Research and Forecasting)-Hydro (Xiang

et al., 2017) models. Since this map has an equivalent spatial resolution of ~6 km, which is rather coarse for hyperresolution simulations, in this work I integrated the information of the CONABIO soil map with the high-resolution soil dataset recently released by the International Soil Reference Information Centre (ISRIC)-World Soil Information. This dataset, named SoilGrids250m (Hengl et al., 2017), consists of global maps of soil physical properties, including bulk density and content mass fractions of clay, silt and sand at multiple depths, which were generated by merging soil profiles and remote sensing imagery (see Hengl et al., 2014, 2017 for details).

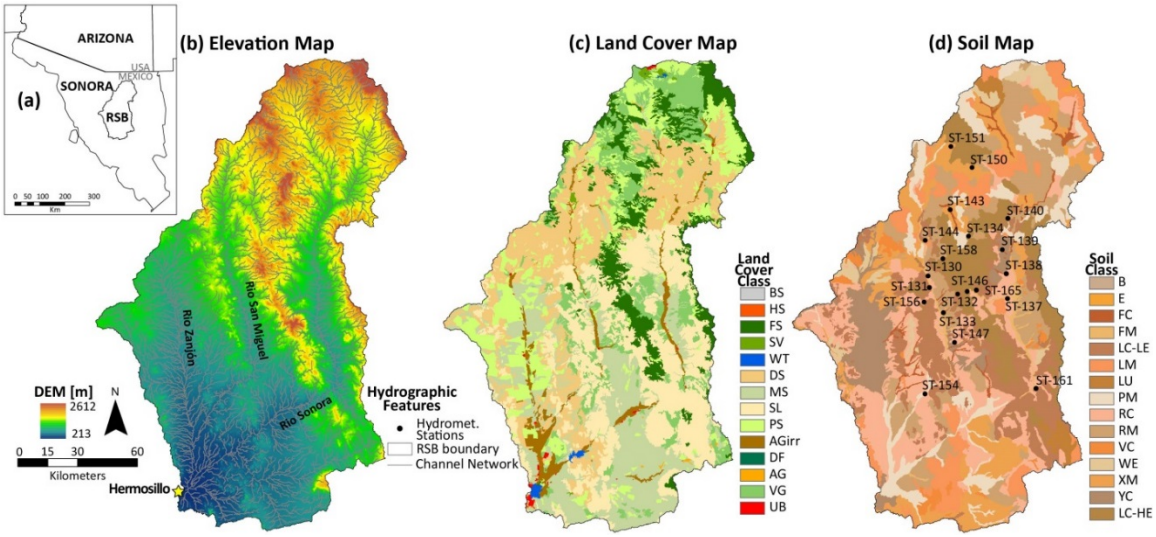

Figure 3.1. (a) Location of the RSB in the state of Sonora, Mexico. (b) Elevation map with the channel network represented in the model along with the main rivers and location of the state capital, Hermosillo. (c) Land cover map from INEGI. (d) Soil map from CONABIO with location of the stations used for model testing (see Table 3.2). Acronyms of the soil and land cover classes are described in Table 3.1.

Table 3.1. Percentage of area (A_f) of the RSB covered by each soil and land cover class along with the IDs used in Figure 3.1 and Table 3.2.

Land cover class	ID	A_f [%]	Soil class	ID	A_f [%]
Bare soil	BS	<0.01	Cambisol	B	0.02
Human Settlement	HS	0.07	Rendzina	E	0.1
Forest	FS	9.1	Fluvisol-Coarse	JC	1.4
Sparse Vegetation	SV	0.3	Fluvisol-Medium	JM	1.3
Water	WT	0.3	Lithosol-Coarse, Low Elevation	IC-LE	12.5
Desert Scrub	DS	20.7	Lithosol-Coarse, High Elevation	IC-HE	10.4
Mesquite	MS	18.4	Lithosol-Medium	IM	9.0
Shrubland	SL	22.1	Luvisol	L	1.4
Pasture	PS	14.0	Phaeozem-Medium	HM	9.2
Agricultural area with irrigation	AGirr	3.3	Regosol-Coarse	RC	17.0
Deciduous Forest	DF	0.01	Regosol-Medium	RM	15.0
Agricultural area without irrigation	AG	0.1	Vertisol-Coarse	VC	1.5
Vegetation	VG	11.4	Planosol	WE	2.9
Urban Area	UB	0.2	Xerosol-Medium	XM	12.0
			Yermosol-Coarse	YC	6.5

Table 3.2. Summary of the datasets used in the hyperresolution simulations. Sources are described in the text. The type of dataset can be model outputs (MO); remote sensing products (RS); or ground observations (GO). The variables are: elevation; FAO soil classes; land cover classes; percentage (%) of sand, silt, and clay; bulk density; precipitation (P); surface pressure (SP); incoming solar radiation (IS); relative humidity (RH); air temperature (T); wind speed (US), soil moisture (SM); land surface temperature (LST); normalized difference vegetation index (NDVI); leaf area index (LAI); and albedo (AL). For each dataset, a short description of the usage in the hyperresolution simulations is also provided.

Source	Type	Variables	Spatial, Temporal Resolution	Usage
SRTM	RS	elevation	30 m	TIN creation NLDAS downscaling Bedrock depth calculation
CONABIO	RS, MO	FAO soil classes	~6 km	Derivation of soil parameters
INEGI	RS	Land cover classes	~4 km	Derivation of vegetation parameters
SoilGrids250m	RS, MO	% sand, % silt, % clay, bulk density	250 m	Derivation of soil parameters (K_s, θ_r, θ_s)
NLDAS-2	MO	P, SP, IS, RH, T, US	12 km, 1 h	Meteorological forcings
CEA	GO	P, SP, RH, T, US	Point, daily	NLDAS bias correction
CONAGUA	GO	P	Point, daily	NLDAS bias correction
ASU-UNISON	GO	P, SM, LST, IS (ST-147)	Point, hourly	NLDAS bias correction, model calibration and validation
MODIS	RS	NDVI	250 m, 16-day composite	Derivation of vegetation parameters
		LAI	1000 m, 8-day composite	
		AL	500 m, 16-day composite	
		LST	1000 m, daily	Model validation

Ground Observations, Model Forcings, and Remotely-sensed Datasets

To capture the spatiotemporal variability of hydrometeorological forcings and vegetation properties, I combined ground observations and remotely-sensed datasets from multiple sources. Table 3.2 provides a summary of all datasets used in this study, including variables and their abbreviations, spatial and temporal resolutions, and usage within the hyperresolution simulations. In the RSB, two government agencies, including Comisión Nacional del Agua (CONAGUA) and Comisión Estatal del Agua (CEA), collect meteorological data (listed in Table 3.2) at daily resolution at a few locations. In addition, a regional network of stations measuring hourly P and surface (5-cm) SM and temperature (used as a proxy of LST, variables described in Table 3.2) has been installed and managed by Arizona State University and Universidad de Sonora (ASU-UNISON), with the aim of investigating the impacts of the NAM on regional ecosystems and hydrology (Vivoni et al., 2007b). Here, I used P data from a total of fifty-eight stations, including six gauges from CEA, nineteen from CONAGUA, and thirty-three from ASU-UNISON. Meteorological data were obtained from seven weather stations from CEA and an eddy covariance tower installed by ASU (Méndez-Barroso et al., 2014). In addition, SM (LST) data recorded by twenty (nine) stations of the ASU-UNISON network were used to parameterize and validate the model. The location of these stations is reported in Figure 3.1(d), while their characteristics including elevation, soil and land cover classes are presented in Table 3.3.

Table 3.3. Locations (latitude, longitude and elevation) and characteristics (vegetation and soil classes) of the hydrometeorological stations of the ASU-UNISON network, including the years used to calibrate and validate the tRIBS hydrologic model against SM observations. Stations marked with an asterisk were used by MA15 to parameterize the tRIBS model.

Station ID	Lat. [deg.]	Lon. [deg.]	Elev. [m]	Soil Class	Land Cover Class	Calibration Year	Validation Years
ST-130*	30.04	-110.67	718	RC	VG	2004	2007
ST-131	29.99	-110.67	728	LC-LE	VG	-	2013
ST-132*	29.96	-110.52	897	LC-LE	SL	-	2004, 2007, 2013
ST-133*	29.88	-110.59	622	RM	PS	2005	2007
ST-134	30.22	-110.46	1183	RM	FS	2004	
ST-137	29.94	-110.26	660	RC	MS	2004	
ST-138	30.05	-110.27	724	FC	MS	2004	
ST-139	30.16	-110.29	747	YC	MS	2004	
ST-140	30.3	-110.26	1014	LC-HE	PS	2004	
ST-143*	30.34	-110.56	962	FC	AGirr	2004	2005
ST-144*	30.2	-110.69	794	FM	AGirr	-	2005, 2007
ST-146*	29.97	-110.47	1373	LC-HE	SL	2004	2007, 2013
ST-147*	29.74	-110.54	634	YC	VG	2004	2006
ST-150	30.53	-110.44	1508	XM	VG	-	2013
ST-151*	30.62	-110.55	1427	XM	VG	-	2007, 2009, 2013
ST-154*	29.51	-110.69	413	PM	AGirr	-	2007, 2008
ST-156	29.92	-110.69	912	RC	VG	-	2013
ST-158	30.12	-110.6	1218	LC-HE	VG	-	2013
ST-161	29.53	-110.12	522	LC-LE	UB	-	2013
ST-165	29.98	-110.42	880	LC-LE	SL	-	2013

In addition to ground data, which suffer from sparse areal and time coverage, hydrometeorological forcings for the model were obtained from reanalysis dataset of the North American Land Data Assimilation System (NLDAS-2), which provides a long-term and consistent set of spatially-distributed variables (Mitchell et al., 2004). As explained in Methods section, NLDAS data were bias corrected with ground observations and downscaled in space. Time-varying maps of vegetation parameters were derived from the composite products of remotely-sensed imagery of the Moderate Resolution

Imaging Spectroradiometer (MODIS) sensor, as described in Xiang et al. (2014). To assess the accuracy of the simulated spatial outputs, daytime daily LST imagery from MODIS aboard the Aqua satellite were obtained for scenes with less than 10% of missing data, which are mostly due to cloud cover during the NAM. The Aqua satellite (overpassing at 1:30 pm) was preferred over Terra (overpassing at 10:30 am), because its LST estimates were found to compare more closely with ground observations of the ASU-UNISON network (not shown), consistent with Xiang et al. (2014).

Methods

Model Overview and Setup

tRIBS is a physically-based, distributed hydrological model capable of reproducing the coupled water-energy balance in a continuous fashion (Ivanov et al., 2004b, 2004a, Vivoni et al., 2007a, 2011). This model provides a spatially explicit representation of terrain, soil and vegetation properties through a TIN containing elevation, channel, and basin boundary nodes (Vivoni et al., 2004, 2005). The Voronoi polygons associated to the TIN serve as individual computational domains where the model solves the mass and energy balances. The model accounts for the spatial variability of meteorological forcings and vegetation parameters, which are provided as time-varying grids. A range of hydrologic processes are simulated by tRIBS. Infiltration, lateral movement of water, and soil moisture redistributions are modeled through a modified Green-Ampt scheme (Cabral et al., 1992). A simplified groundwater model is used to model saturated flow in the aquifer. The Penman–Monteith approach (Penmen,

1948; Monteith, 1965) is applied to estimate potential evapotranspiration when solving the energy balance. The actual evapotranspiration is then computed based on the available soil moisture and partitioned into plant transpiration and evaporation from bare soil and wet canopy. A hydrologic routing scheme is used to simulate overland flow in the hillslope, while the kinematic wave approximation is adopted to route water in the channel network. Details on each process are provided by Ivanov et al. (2004a). In the RSB, a TIN of 2,612,662 Voronoi polygons or 11.4 % of the original DEM pixels was created, corresponding to an average cell size of ~88 m. This represents the largest domain that has been modeled to date with tRIBS. In addition, the domain resolution is well within the definition of hyperresolution provided by Wood et al. (2011) for applications in regional watersheds. A spatially-variable soil depth was calculated following the method of Saulnier et al. (1997) based on the assumption of a linear decreasing function between soil thickness and elevation (see Méndez-Barroso et al., 2016 for details).

Hydrometeorological Forcings and Vegetation Parameters

A challenge of hyperresolution simulations is the availability of reliable high-resolution hydrometeorological forcings. To address this, I combine the sparse ground-based observations with hourly NLDAS products. Since these gridded datasets are provided at a relatively coarse spatial resolution of ~12 km, the meteorological variables were downscaled at 1-km resolution. Specifically, as done in previous applications in the region (e.g., Robles-Morua et al., 2012; Xiang et al., 2014, 2017; Mascaro et al., 2015), the NLDAS data were first bias corrected using ground data. A minimum of seventeen

rain gauges from the various networks were used to correct the bias of P by computing a daily factor based on the mean areal precipitation. Observations from seven weather stations of the CEA and ASU-UNISON networks were used to remove the bias of IS, RH, and US through linear regression equations. Subsequently, the methods proposed by Liston and Elder (2006) were used to downscale T, PA, RH, US, and IS at 1-km resolution. These techniques combine the Barnes objective analysis scheme (Barnes, 1994; Koch et al., 1983) with terrain information to spatially interpolate meteorological variables into a high-resolution grid. Terrain, slope, and curvature grids were generated at 1-km resolution and used to apply the downscaling method: elevation was used for T, PA, RH; slope for IS; and slope and curvature for US. As an example, Figure 3.2 shows results of the bias correction of the mean annual P and of the downscaling procedure for the mean annual T across the simulation period 2004-2013.

In regions with strong seasonality, the ability to account for spatiotemporal variations of vegetation in hyperresolution simulations is key to model evapotranspiration and canopy interception. The tRIBS model has this ability by allowing the ingestion of time-varying maps of vegetation properties. As in MA15, the vegetation parameters were derived from MODIS imagery (see Table 3.2) by applying the empirical relationships proposed by Méndez-Barroso et al. (2014) and Xiang et al. (2014). MODIS-derived vegetation products have been shown to represent vegetation conditions well in semiarid regions against ground observations (Fensholt et al., 2004; Ryu et al., 2012). The 16-day composites of NDVI and AL, and the 8-day composites of LAI (variables described in

Table 3.2) were used to obtain the model vegetation parameters, which were then linearly interpolated in time to create the daily fields.

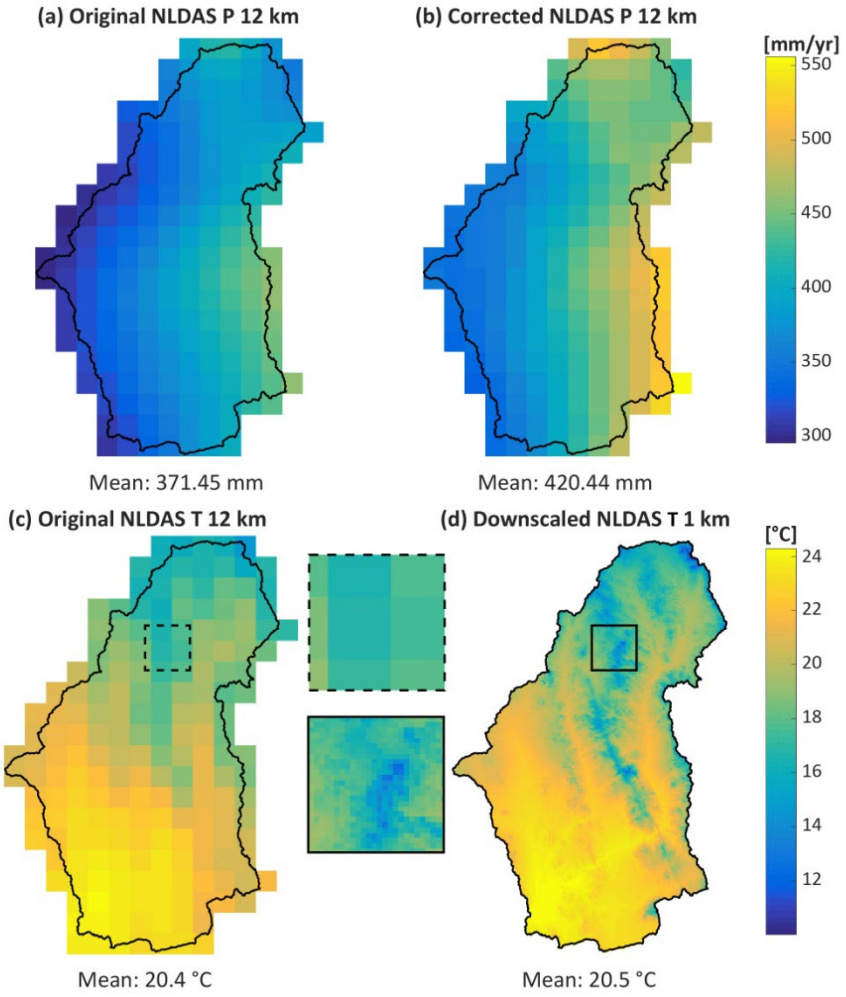

Figure 3.2. (a) NLDAS and (b) bias corrected mean annual P across the period 2004-2013. (c) NLDAS and (d) downscaled mean annual T across the period 2004-2013. Boxes show details of the same area of the original and downscaled T maps to visualize difference in spatial variability.

Model Parameterization and Generation of Soil Maps

The model parameters for the hyperresolution simulations were identified through a two-step strategy. First, soil parameters were derived for the CONABIO soil classes (Figure 3.2(d) and Table 3.1) by adjusting values published in previous studies (MA15; Xiang et al., 2014; Vivoni et al., 2009) through calibration against SM observations at the ASU-UNISON stations. MA15 calibrated the soil parameters of nine classes of the CONABIO map that cover most (85%) of the basin area, using SM observations of nine co-located stations (marked with an asterisk in Table 3.3). Here, I updated the calibrated parameters of these nine classes by including new observations of additional eleven stations and different time periods (see Table 3.3). For this aim, the tRIBS model was setup at each station, creating a domain with a single Voronoi polygon (area of 100 m^2) as in MA15. Individual station simulations were conducted from 1 June to 30 September of different years, including a dry month used as spinup and the wet summer season. The model was initialized assuming a 25% saturation of the total soil depth. Forcings for the simulations included precipitation observed at the stations, downscaled meteorological data from NLADS, and vegetation parameters from MODIS in the co-located pixels. I initially adopted the soil parameters obtained by MA15 and, then, performed minimal adjustments to saturated hydraulic conductivity (K_s) and/or stress thresholds for transpiration (θ^*_t) with the aim of improving model performance based on new available observations. For the six soil classes where no station was present, I (i) used the ROSETTA pedotransfer function (Schaap et al., 2001) to estimate the dominant soil texture, and (ii) adopted the parameter values reported by Xiang et al. (2014) and Vivoni

et al. (2009) for the same soil texture class. The calibrated soil parameters are presented in Table 3.4. For the soil heat conductivity (k_s) and soil heat capacity (C_s), I used dry and wet values (k_s of 1.2 and 1.3 J/ms K; C_s of 6 x 10^5 and 1.47 x 10^6 J/m^3 K), depending on a soil moisture threshold of 0.021 m^3/m^3 as in Xiang et al. (2014).

In the second step of the strategy, I increased the spatial variability of K_s, residual moisture content (θ_r), saturated moisture content (θ_s) and porosity (n) by incorporating the SoilGrids250m maps. I first applied the ROSETTA pedotransfer functions to the datasets to generate maps of K_s, θ_r, θ_s and n at 250-m resolution. Since the parameter values provided by the pedotransfer functions were affected by significant bias compared to those calibrated at the stations, I applied a bias correction procedure that is illustrated using K_s as example. For the jth soil class of CONABIO, I computed the mean value of K_s returned by the pedotransfer functions in all 250-m pixels of that class. This is defined as $\overline{K}_{s,j}^{p} = \frac{1}{N_j}\sum_{i=1}^{N_j} K_{s,ji}^{p}$, where $K_{s,ji}^{p}$ is the value of K_s returned by the pedotransfer function in pixel i of the map at 250-m resolution located in class j, and N_j is the total number of pixels located in class j. Next, I estimated a single value of the bias for the class j as $b_j = (K_{s,j}^{c} - \overline{K}_{s,j}^{p})$, where $K_{s,j}^{c}$ is the value of K_s calibrated for class j (Table 3.4). Finally, I computed the bias corrected value in pixel i as $K_{s,ji}^{bc} = (K_{s,ji}^{p} + b_j)$. The procedure is illustrated in Figure 3.3 for K_s and θ_s, where values in the CONABIO map derived through calibration at the stations, those returned by pedotransfer functions, and the bias corrected values are shown. Summarizing, the basin simulations consisted of: (i) bias-corrected values of K_s, θ_r, θ_s and n at 250-m resolution; and (ii) values reported in Table 3.4 for other parameters with the spatial distribution of CONABIO classes.

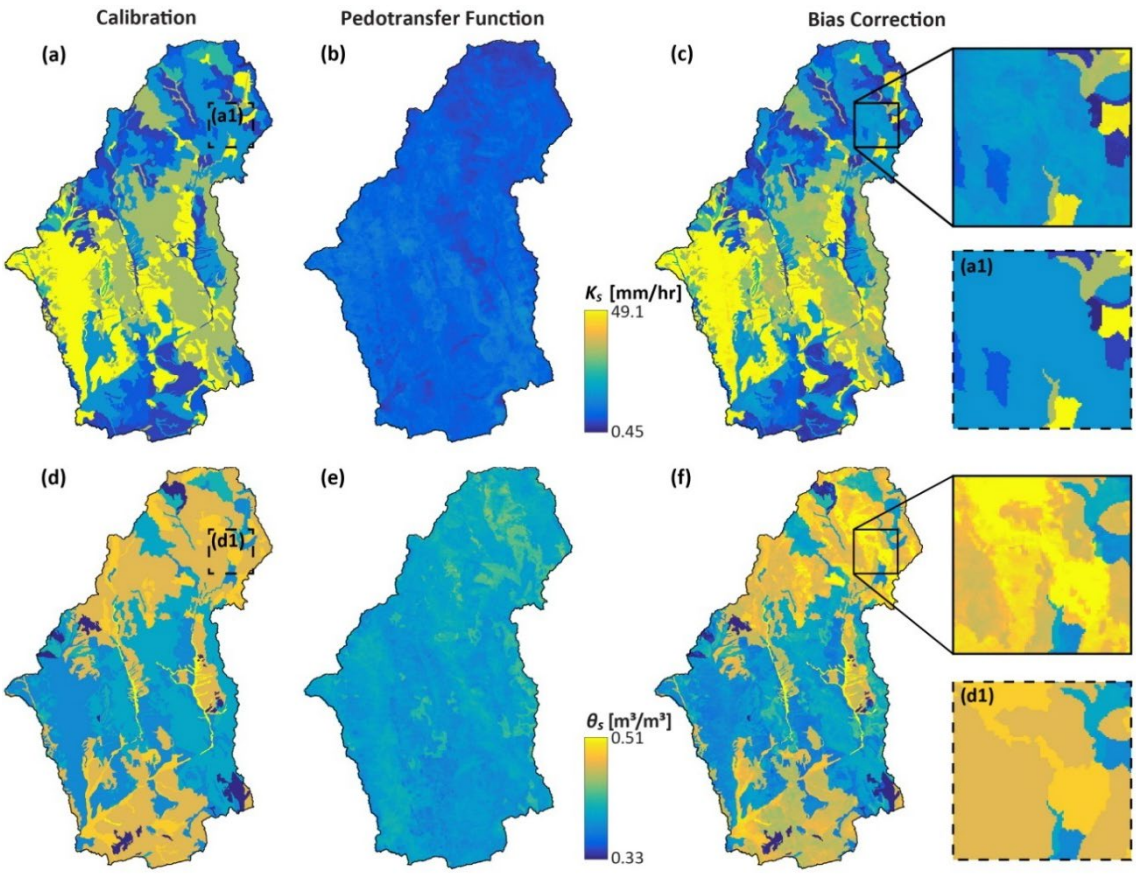

Figure 3.3. Generation of high-resolution (250-m) maps of K_s and θ_s. (a, d) Parameter values calibrated for the CONABIO soil classes. (b, e) Parameter values estimated through the ROSETTA pedotransfer functions applied to SoilGrids250m. (c, f) Bias corrected values. Boxes show zooms on the same area of the CONABIO and bias corrected maps to better visualize the difference in spatial variability.

Table 3.4. Calibrated values of model parameters for the CONABIO soil classes. Parameters are porosity (n), saturated hydraulic conductivity (K_s), saturated soil moisture content (θ_s), residual soil moisture content (θ_r), pore size distribution index (m), conductivity decay parameter (f), air entry bubbling pressure (Ψ_b), conductivity anisotropy ratio (A), and stress thresholds for evaporation (θ^*_e) and transpiration (θ^*_t). NA is not available.

	Station	K_s [mm/hr]	θ_s []	θ_r []	m []	Ψ_B [mm]	f [mm^{-1}]	A []	n []	θ^*_e []	θ^*_t []
B	NA	24.2	0.41	0.025	1.5	0	0.001	1	0.417	0.298	0.298
E	NA	0.5	0.39	0.077	0.2	0	0.007	1	0.417	0.297	0.297
FC	138, 143	33.6	0.4	0.05	1.5	0	0.05	1	0.42	0.28	0.13
FM	144	5	0.51	0.06	2	0	0.0009	1	0.53	0.2	0.13
LC-LE	131, 132, 161, 165	32.4	0.4	0.01	2	0	0.0009	1	0.42	0.07	0.06
LC-HE	140, 146, 158	32.4	0.4	0.01	0.6	0	0.008	1	0.42	0.08	0.04
LM[1]	NA	6.3	0.4	0.063	1.5	0	0.001	1	0.419	0.299	0.299
LU[2]	NA	9.8	0.33	0.068	0.3	0	0.001	1	0.418	0.299	0.299
PM	154	15	0.49	0.01	1.5	0	0.0009	1	0.51	0.19	0.19
RC	130, 137, 156	49.1	0.38	0.002	0.8	0	0.0003	1	0.4	0.04	0.02
RM	133, 134	15.3	0.47	0.002	2	0	0.0009	1	0.49	0.24	0.3
VC	NA	0.5	0.39	0.077	0.2	0	0.007	1	0.417	0.297	0.297
WE	NA	24.2	0.41	0.025	1.5	0	0.001	1	0.417	0.298	0.298
XM	150, 151	3.9	0.47	0.01	2	0	0.0006	1	0.49	0.28	0.28
YC	139, 147	49.1	0.38	0.002	1.5	0	0.0009	1	0.4	0.19	0.13

Hyperresolution Basin Simulations on High-performance Computing Clusters

Basin simulations were conducted from 1 January 2004 to 31 December 2013 using the parallel version of tRIBS (Vivoni et al., 2011). A spinup period of 60 months was used, chosen after verifying the numerical stability of the model outputs. Combining domain size and modeling period, these are the most computationally intensive simulations that have been conducted to date with the tRIBS model. To satisfy these computational needs, high-performance computing clusters available at ASU were used. The model domain was partitioned among different processors using a criterion that optimizes size of the sub-domains and message passing across processors (Karypis and Kumar, 1998). Since each processor simulates a portion of the domain and returns its

associated spatial outputs, spatial maps for the entire basin were created by merging the outputs generated by all processors. A total of 3653 spatial outputs were obtained at the overpass times of MODIS during the simulation period. LST outputs were interpolated from the irregular mesh to MODIS grid to allow the application of techniques to compare spatial patterns. Additionally, time series of hydrologic variables were outputted by the model at the Voronoi polygons containing the stations of the ASU-UNISON network and at their neighbors, for a total of 162 nodes. Overall, the size of the generated output files was 0.5 TB.

Evaluation of Simulated Spatial Patterns

I compared the LST simulated spatial patterns with those estimated by MODIS at 1-km resolution using three tools, including Taylor diagram, connectivity analysis, and empirical orthogonal function (EOF) analysis. These tools test different properties of spatial patterns and have proven to be effective for the analysis of large and high-dimensional datasets.

1) Taylor diagram

The Taylor diagram (Taylor, 2001) has been widely used in the intercomparison of spatial outputs of climate models (e.g., Gautam and Mascaro, 2018; Ahmadalipour et al., 2018) and recently for hydrologic models (Orth et al., 2015; Koch et al., 2016a). Taylor diagrams provide a concise overview of the pixel-to-pixel correspondence between two spatial patterns, by plotting in the same space three metrics measuring linear correlation, mean error, and spatial variability. The metrics include the coefficient of correlation (CC) between the patterns; the centered root mean square error (RMSE),

defined as the RMSE of the anomalies (i.e., the variables of interest minus the corresponding mean); and the spatial standard deviations (STDs) of the two maps. To compare several pairs of LST patterns, I used the normalized Taylor diagram, which is obtained by dividing the standard deviation and centered RMSE by the reference STD. The corresponding normalized metrics are indicated with nRMSE and nSTD.

2) Connectivity analysis

Connectivity analysis was developed in groundwater hydrology to characterize aquifer heterogeneity (e.g., Renard and Allard, 2013; dell' Arciprete et al., 2012). A few studies have applied this tool to analyze spatial patterns of land surface variables such as SM (Western et al., 2001; Grayson et al., 2002; Koch et al., 2016a), LST (Koch et al., 2016b, 2017; Koch and Stisen, 2017) and evapotranspiration (Koch et al., 2017; Koch and Stisen, 2017). Here, the method proposed by Renard and Allard (2013) for continuous variables was used, as in Koch et al. (2016b). Let $\mathbf{X}$ be the matrix of a spatial pattern for the analyzed variable, and x_F the quantile calculated for the non-exceedance (for the high-phase analysis) or exceedance (for the low-phase analysis) probability F from the empirical cumulative distribution function of all values in $\mathbf{X}$. The method involves: the generation of the binary map $\mathbf{X_F} = (\mathbf{X} \geq x_F)$ (high-phase) or $\mathbf{X_F} = (\mathbf{X} \leq x_F)$ (low-phase); the identification of the clusters of connected pixels in $\mathbf{X_F}$; and the computation of the metric $\Gamma(F)$ defined as:

$$\Gamma(F) = \frac{1}{n_F^2} \sum_{k=1}^{N(\mathbf{X_F})} n_k^2, \quad (1)$$

where n_F is the total number of pixels in the binary map $\mathbf{X_F}$ above x_F, n_k is the number of pixels in the kth cluster in $\mathbf{X_F}$, and $N(\mathbf{X_F})$ is the number of clusters. This metric quantifies

the degree of connectivity of the field: when $F = 0$, all pixels are connected, $N(\mathbf{X}_F) = 1$, $n_k = n_F$, and $\Gamma(F) = 1$; if most pixels are disconnected, $N(\mathbf{X}_F)$ is large, the values of n_k are similar across all clusters, and $\Gamma(F)$ becomes small. $\Gamma(F)$ can be calculated for a range of F from 0 to 1 to study the transition from a large cluster to smaller disconnected clusters. For simplicity, the relation $\Gamma(F)$ versus F is named here the connectivity function. When comparing spatial patterns of two variables, the use of frequencies makes $\Gamma(F)$ insensitive to the presence of bias. Thus, $\Gamma(F)$ is suitable for comparing variables that are expected to be correlated, such as temperature and elevation.

Here, I calculated the connectivity function of LST through the high-phase analysis. To assess the degree of similarity between the patterns of observed and simulated LST, I used the root-mean-squared error (RMSE) between the two connectivity functions, as in (Koch et al., 2016b), as well as the correlation coefficient (CC). In addition, to study he physical controls on the discrepancies between simulated and remotely-sensed LST patterns, I computed the connectivity function of elevation and vegetation fraction (VF), a time-varying parameter. Since these variables are inversely related to LST (i.e., LST tends to be lower at higher elevation or for higher VF), the low-phase analysis was applied for these two variables.

3) EOF analysis

The EOF analysis is a multivariate statistical method used to analyze large spatiotemporal datasets (Hannachi et al., 2007). It has been mainly applied to interpret climate and atmospheric variables and, more recently, for the outputs of land surface models (Koch et al., 2016b) and hyperresolution hydrologic simulations (MA15; Fang et

al., 2015; Koch et al., 2017). This analysis identifies a set of orthogonal spatial patterns (EOFs), which are time-invariant, and a set of coefficients (principal components, PCs), which control the contribution of each EOF in time. Here, the EOF analysis was applied to the spatial anomalies of LST and VF. Details on the methodology adopted to compute EOFs and PCs are provided in Appendix A of MA15.

Results and Discussion

Performance of Model Simulations of SM at the Stations

Model performances were first evaluated for the simulations of surface (0-5 cm) SM conducted individually at each of the twenty stations listed in Table 3.3. Figure 3.4 shows results for two and four representative stations in the calibration and validation periods, respectively. The quantitative assessment of the model performance is presented in Table 3.5 (individual simulations), which reports bias (B), mean absolute error (MAE), and correlation coefficient (CC) for all stations. During the calibration periods, the model is able to simulate quite well the observed soil moisture response to precipitation and the subsequent recession phases during dry periods [Figure 3.4(a) and (b)]. This is quantified by the low B (-0.004 ± 0.041 m^3/m^3), relatively low MAE (0.040 ± 0.023 m^3/m^3) and high CC (0.81 ± 0.08). During the validation period, model performance is fairly good at most stations [Figure 3.4(c)-(e)] and degrades at a limited number of sites [see Figure 3.4(f)]. This is reflected in a slightly larger magnitude of B (-0.027 ± 0.036 m^3/m^3) and similar MAE (0.039 ± 0.031 m^3/m^3) and CC (0.83 ± 0.07). When considering the simulations at the basin scale, the metrics reported in Table 3.5 (basin simulations) show

that performances are comparable to those of the individual simulations, in terms of B and MAE, but are characterized by lower CC, whose mean decreases to 0.65 and 0.74 in the calibration and validation periods, respectively. These discrepancies can be explained by the fact that, in basin simulations, precipitation forcings from NLDAS are used instead of local observations, and the effect of lateral transport of moisture in the vadose zone is taken into account. Overall, the metrics listed in Table 3.5 indicate similar or better performances than those reported by MA15 for tRIBS and Xiang et al. (2017) for WRF-Hydro in a subset of stations and time periods.

Table 3.5. Metrics quantifying model performances in the simulation of SM at the stations for individual and basin simulations in calibration and validation years. For each metric, italic font is used to indicate the worst performing station. Stations marked by an asterisk were used in MA15 for the corresponding years. Metrics are bias, B = $\bar{S} - \bar{O}$, mean absolute error, MAE = $\frac{1}{N}\sum_{j=1}^{N}|S_j - O_j|$, and correlation coefficient, CC = $\{\sum_{j=1}^{N}(O_j - \bar{O})(S_j - \bar{S})\}/\left\{\left[\sum_{j=1}^{N}(O_j - \bar{O})^2\right]^{0.5}\left[\sum_{j=1}^{N}(S_j - \bar{S})^2\right]^{0.5}\right\}$, where S_j and O_j are the simulated and observed variables for j = 1, …, N time steps, and $\bar{S}$ and $\bar{O}$ are their temporal means.

Calibration		Individual simulations			Basin simulations		
Station	**Year**	**B [m^3/m^3]**	**MAE [m^3/m^3]**	**CC [-]**	**B [m^3/m^3]**	**MAE [m^3/m^3]**	**CC [-]**
130*	2004	-0.001	0.013	0.87	0.012	0.021	0.74
133*	2005	0.015	0.028	0.75	0.044	0.049	0.70
134	2004	0.042	0.046	0.89	-0.057	0.060	0.81
137	2004	-0.022	0.049	*0.62*	-0.002	0.050	*0.42*
138	2004	-0.038	0.040	0.90	0.006	0.060	*0.46*
139	2004	0.047	0.048	0.82	-0.020	0.032	*0.56*
140	2004	*-0.100*	*0.100*	0.83	*-0.100*	*0.100*	0.73
143*	2004	-0.004	0.029	0.82	0.004	0.048	*0.52*
146*	2004	0.025	0.032	0.84	0.027	0.037	0.75
147*	2004	-0.003	0.018	0.78	0.007	0.019	0.81
	Mean	-0.004	0.040	0.81	-0.008	0.048	0.65
	STD	0.041	0.023	0.08	0.040	0.022	0.14
Validation		**Individual simulations**			**Basin simulations**		
Station	**Year**	**B [m^3/m^3]**	**MAE [m^3/m^3]**	**CC [-]**	**B [m^3/m^3]**	**MAE [m^3/m^3]**	**CC [-]**
130*	2007	-0.007	0.020	0.82	0.008	0.023	0.74
130	2013	-0.003	0.011	0.95	0.012	0.015	0.80
131	2013	0.004	0.016	0.88	0.011	0.015	0.80
132*	2004	0.003	0.020	0.86	0.016	0.025	0.76
132*	2007	-0.013	0.019	0.88	0.043	0.049	0.70
132	2013	-0.020	0.021	0.89	-0.023	0.026	0.75
133*	2007	-0.008	0.019	0.89	0.039	0.042	0.89
143*	2005	-0.029	0.029	0.82	0.032	0.042	0.78
144*	2005	-0.013	0.031	0.80	-0.007	0.031	0.82
144*	2007	0.006	0.023	0.88	0.032	0.042	0.78
146*	2007	-0.017	0.021	0.93	0.008	0.029	0.84
146	2013	-0.073	0.074	0.78	-0.087	0.091	*0.54*
147*	2006	0.016	0.018	0.76	0.001	0.035	*0.48*
150	2004	-0.044	0.055	0.82	-0.042	0.060	0.78
151*	2007	-0.022	0.038	0.81	0.017	0.042	0.81
151*	2009	-0.016	0.029	0.78	0.003	0.034	0.79
151	2013	-0.096	0.096	0.89	-0.063	0.069	0.81
154*	2007	0.003	0.024	0.83	0.004	0.029	0.81
154*	2008	0.003	0.028	0.82	0.005	0.029	0.83
156	2013	*-0.136*	*0.146*	*0.68*	*-0.121*	*0.122*	*0.52*
158	2013	-0.068	0.068	*0.66*	0.018	0.040	*0.57*
161	2013	-0.040	0.040	0.89	-0.013	0.030	0.71
165	2013	-0.046	0.046	0.82	-0.021	0.029	0.80
	Mean	-0.027	0.039	0.83	-0.006	0.041	0.74
	STD	0.036	0.031	0.07	0.039	0.024	0.11

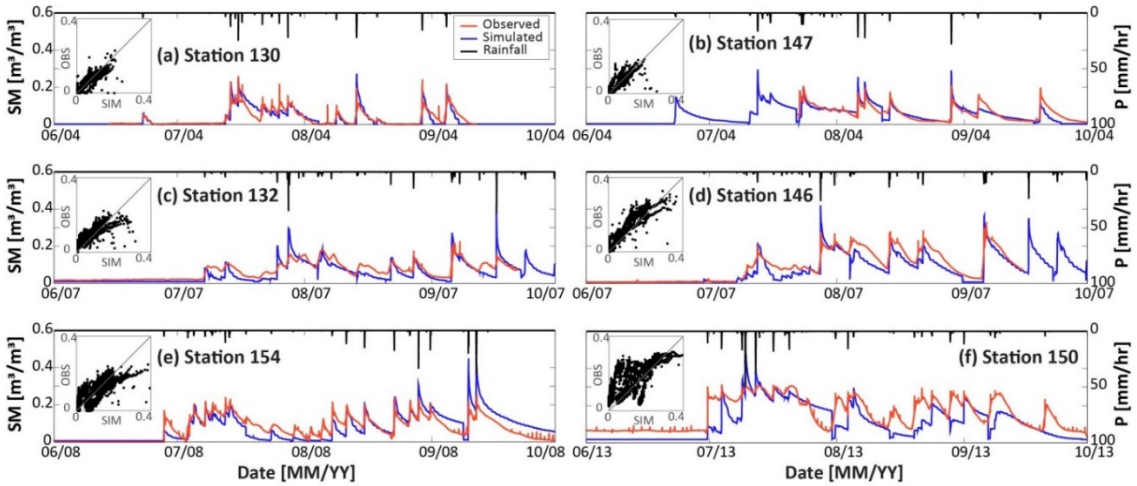

Figure 3.4. Comparison of observed and simulated surface (5-cm depth) SM at representative stations for (a-b) calibration and (c-f) validation periods. Insets show scatterplots of observed and simulated surface SM along with the 1:1 line. Data gaps are due to missing observations or poor data quality.

Performance of Model Simulations of LST at the Stations

Prior to verifying the model ability to simulate the spatial patterns of LST, it is crucial to assess whether the model accurately reproduces LST that is (i) observed at the ground and (ii) estimated from the MODIS remote sensors. I selected nine stations with more than two years of consecutive observations, apart from some short gaps. For each station, the observed and simulated LST values were extracted at MODIS overpass times, while the LST values estimated by MODIS were obtained in the co-located 1-km pixel. Figure 3.5 shows time series and scatterplots for four representative stations, while Table 3.6 presents the metrics for all stations. The discrepancies between ground and remotely-sensed LST are larger than those computed between ground and simulated LST, thus indicating higher skills of hyperresolution simulations to capture the local variability of LST measured by ground sensors. For example, the mean B and MAE between ground and remotely-sensed (ground and simulated) LST are +2.5 ºC and 5.9 ºC (-1.4 ºC and

4.9ºC), respectively. The simulated signal is always highly correlated with both ground and MODIS LST (mean CC ≥ 0.85) and is consistently negatively biased with the MODIS estimates (mean B of -4.0 ºC), especially in winter (mean B of -5.6 ºC; not shown in Table 3.6). Xiang et al. (2017) also found the basin-averaged LST simulated in the RSB by WRF-Hydro to be negatively biased with MODIS products. These findings reveal that the model is (i) able to simulate quite well the LST measured by ground sensors, and (ii) exhibits high correlation and negative bias with LST estimated by MODIS at distributed locations. As a result, since the connectivity and EOF analyses are not affected by the presence of bias, these tools can be used to effectively assess the model ability to simulate the spatial patterns of LST.

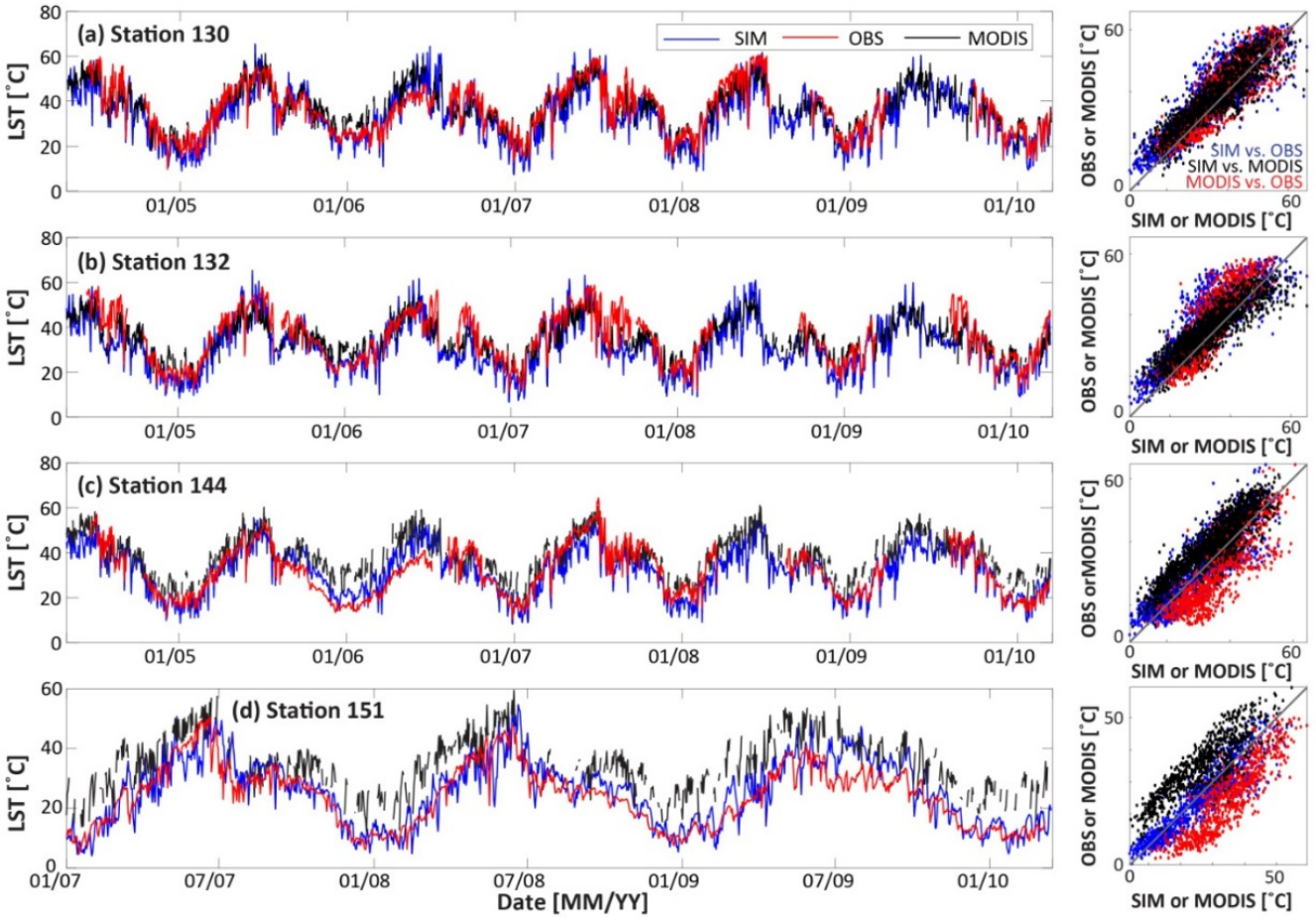

Figure 3.5. Comparison of LST simulated by tRIBS (SIM), observed at the ground (OBS), and estimated by MODIS at four representative stations. Right panels show the corresponding scatterplots. The availability of observed LST data differs at each station.

Table 3.6. Metrics quantifying the comparison between LST simulated by tRIBS (SIM), observed at the ground (OBS), and estimated by MODIS. When comparing SIM with MODIS, negative B means that SIM underestimate MODIS.

Station	B [ºC]			MAE [ºC]			CC [-]		
	SIM vs. OBS	SIM vs. MODIS	MODIS vs. OBS	SIM vs. OBS	SIM vs. MODIS	MODIS vs. OBS	SIM vs. OBS	SIM vs. MODIS	MODIS vs. OBS
130	-3.4	-2.3	-1.5	5.1	4.7	4.1	0.89	0.88	0.88
132	-5.1	-2.0	-3.5	6.0	4.4	5.4	0.87	0.88	0.85
133	2.7	-3.2	6.3	4.7	4.7	7.4	0.88	0.89	0.86
143	2.0	-5.3	7.6	6.6	6.6	9.8	0.79	0.85	0.75
144	-1.2	-5.4	4.3	4.3	6.0	6.1	0.88	0.91	0.82
146	-2.5	-4.4	1.5	5.1	5.4	4.3	0.80	0.86	0.81
147	-2.8	-2.2	-1.5	4.4	4.3	3.3	0.92	0.89	0.90
151	1.6	-8.3	9.7	3.5	8.8	9.9	0.92	0.88	0.88
154	-3.5	-2.7	-1.0	4.5	4.6	3.3	0.92	0.89	0.90
Mean	-1.4	-4.0	2.5	4.9	5.5	5.9	0.87	0.88	0.85
STD	2.6	2.0	4.4	0.9	1.4	2.4	0.04	0.02	0.05

Comparison of Simulated and Remotely-sensed Spatial Patterns of LST

1) Taylor diagram

The standardized Taylor diagram was first used to provide a concise overview of the correspondence between spatial patterns of LST simulated by tRIBS and estimated by MODIS. Figure 3.6(a) shows results considering all available days (total of 1763 where MODIS had <10% of missing data). In most cases, the simulated spatial maps (i) are well correlated with MODIS, with CC mainly included between 0.5 and 0.8; (ii) overestimate the spatial variability of MODIS up to two times; (iii) and are characterized by RMSE ranging from 0.5 to 1.5 times the MODIS spatial STD. Different colors are used to indicate days in the wet summer (July-September) and dry winter (November-March) seasons. There are no marked differences between the two seasons, except for a slightly larger scatter of the points in the summer as compared to winter, especially in terms of CC and nSTD. No clear impact due to the occurrence of precipitation could be detected in either season. Figure 3.6(b) presents a subset of days selected as the best and worst performing cases based on the connectivity analysis (discussed in the result section). Figure 3.6(c) shows the Taylor diagram for the maps of the mean annual, winter and summer LST for each of the 10 simulated years. When aggregation in time is performed, the agreement between simulated and MODIS patterns increases significantly, with $0.8 < CC < 0.9$, $1 < nSTD < 1.5$, and $0.5 < nRMSE < 1$. Clustering of points indicates that the model performances are similar across the years.

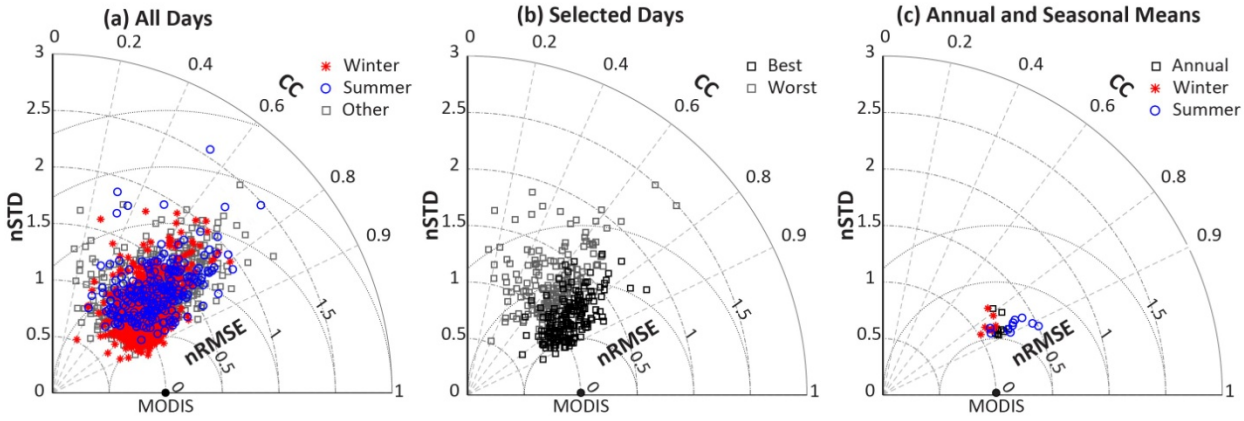

Figure 3.6. Taylor diagrams of LST spatial patterns in (a) all days (total of 1763), (b) days with best and worst performance selected through the connectivity analysis (see Section 4.3.4 for details), and (c) annual, winter, and summer means in each of the 10 years.

2) Connectivity analysis

Since the connectivity analysis is a relatively new tool, I first present an example that helps interpreting the connectivity function, $\Gamma(F)$. Figure 3.7(a) shows $\Gamma(F)$ calculated on simulated and MODIS spatial maps of LST averaged over winter and summer months of all years. $\Gamma(F)$ is calculated for the high-phase analysis, implying that the binary image identifies areas where LST is larger than the threshold x_F. For $F = 0$, the binary image includes the entire basin and $\Gamma(F) = 1$. As F increases, there are less areas in the basin that have LST larger than the associated threshold and $\Gamma(F)$ is lower than 1. Depending on the distinct features of the variables that affect LST, the size and number of connected clusters vary and $\Gamma(F)$ can decrease or increase with F. For $F = 1$, no area of the basin is part of the binary image and $\Gamma(F)$ is again equal to 1.

As reported in Figure 3.7(a), in winter, the model captures the connectivity structure of MODIS LST very well, with CC between the two functions of 0.96 and RMSE of 0.07. Performances slightly degrade in summer (CC of 0.73 and RMSE of

0.16), especially for higher F. To investigate the reasons of these discrepancies and provide further insights on this function, Figure 3.7(b)-(g) report the binary images, $\mathbf{X}_F$, of observed and MODIS patterns for F = 0.8, 0.85 and 0.9. In all cases, the clusters are located in the southwestern part of the basin where elevation is lower and LST is higher. The simulated patterns are always characterized by a larger number of small clusters, $N(\mathbf{X}_F)$. This suggests that hyperresolution simulations are able to capture the small-scale effect of land surface properties that are not represented in MODIS products at 1-km resolution. Despite differences in $N(\mathbf{X}_F)$, the values of $\Gamma(F)$ for F = 0.8 for the two products are not markedly different [Figure 3.7 (b) and (e)]. For F = 0.85 and 0.9, the values of $\Gamma(F)$ differ because the simulations always include a single large cluster, while, in MODIS, this cluster separates into two and three main parts, respectively, leading to a lower $\Gamma(F)$. Overall, Figure 3.7(b)-(g) indicate that the size of the largest cluster significantly controls the value of $\Gamma(F)$.

Model performance of the connectivity function is evaluated in Figure 3.8, which shows CC and RMSE between simulated and MODIS $\Gamma(F)$ at the daily scale along with monthly means to help visualize the overall trend. CC is high in most cases, with a mean of 0.77 (STD of 0.14), while the average RMSE is 0.20 (STD of 0.08). Koch et al. (2016b) reported smaller RMSE of ~0.08, but compared coarser (12-km) LST simulations at a monthly scale. Overall, I conclude that model and remotely-sensed estimates of LST have very similar connectivity functions.

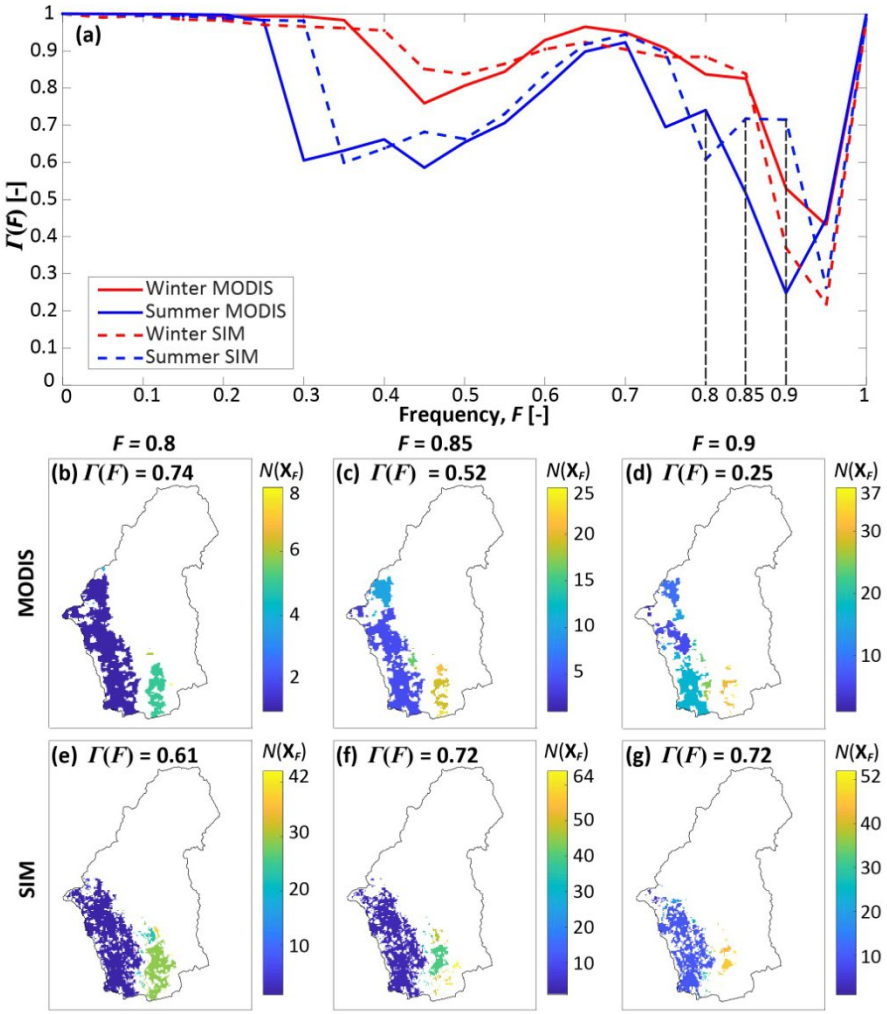

Figure 3.7. (a) Connectivity functions $\Gamma(F)$ of MODIS and simulated (SIM) LST averaged over winter and summer months of all years. (b-g) Binary images $\mathbf{X}_F$ of (b-d) MODIS and (e-g) simulated patterns for F = 0.8, 0.85 and 0.9 ($\Gamma(F)$ identified in panel (a) with dashed lines), with indication of the connected clusters.

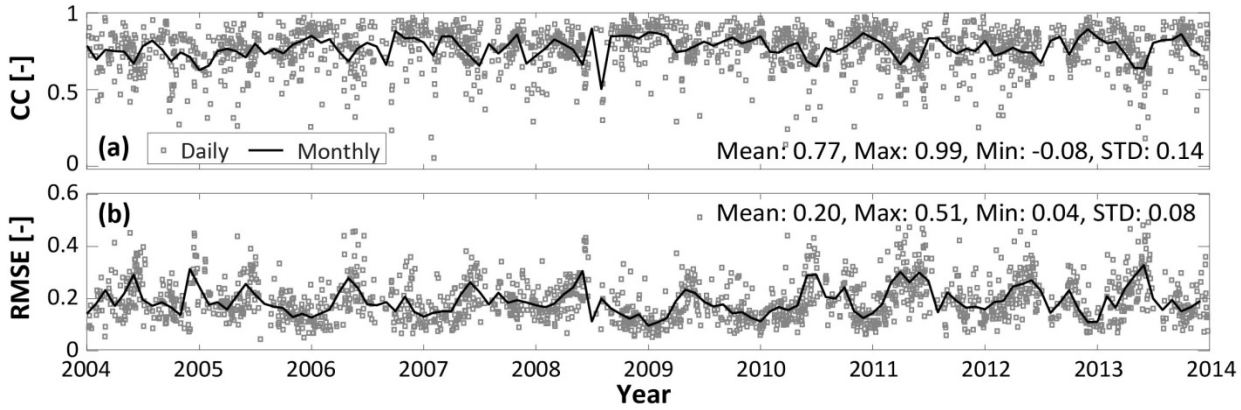

Figure 3.8. (a) CC and (b) RMSE between the connectivity functions of simulated and MODIS LST at daily (markers) and monthly (line) scales. The mean (Mean), maximum (Max), minimum (Min) and standard deviation (STD) of CC and RMSE across the simulation period are also reported.

3) EOF analysis

The EOF analysis was applied to the combined dataset of simulated and MODIS LST patterns of all days, as in Koch et al. (Koch et al., 2017). In such a way: (i) the EOFs are the same for both datasets; and (ii) in a given day, if the maps of simulated and MODIS LST are exactly the same, the associated PCs will be equal, otherwise they will vary. Figure 3.9 (a) and (b) show the first two EOFs, which together explain almost 82% of the total variance of both datasets. To provide a physical interpretation of the EOFs of LST, I also applied the EOF analysis to VF. The first EOF (EOF1) of LST is largely controlled by elevation (CC of -0.94) and, to a smaller extent, by vegetation properties (CC of -0.62 with EOF1 of VF). The second EOF (EOF2) of LST is instead mainly affected by vegetation (CC of -0.56 with EOF1 of VF). Figure 3.9(c) and (d) report the PCs of the observed and simulated LST of each day along with the time series of the monthly means (left panels), as well as the monthly averages over the simulation (right

panels). The contribution of EOF1, quantified through PC1, is constant during the simulation and tends to be, on average, slightly larger for the simulated LST, especially from August to October. Differences between the two products emerge when comparing PC2. MODIS LST leads to a PC2 that is always positive and exhibits a bell-shaped pattern with a peak in July. In contrast, PC2 of simulated LST is negative in all months except the summer monsoon, when it shows a steep bell-shaped pattern with a peak in August. These findings reveal that the vegetation control on simulated LST (via EOF2) is significantly different in summer as compared to the other months, because of the change of sign of PC2. In contrast, the seasonality of vegetation impacts MODIS LST less significantly. Despite these differences, since EOF2 controls only 8.1% of the variability, I can conclude that the hyperresolution simulations of LST are quite similar to the MODIS products.

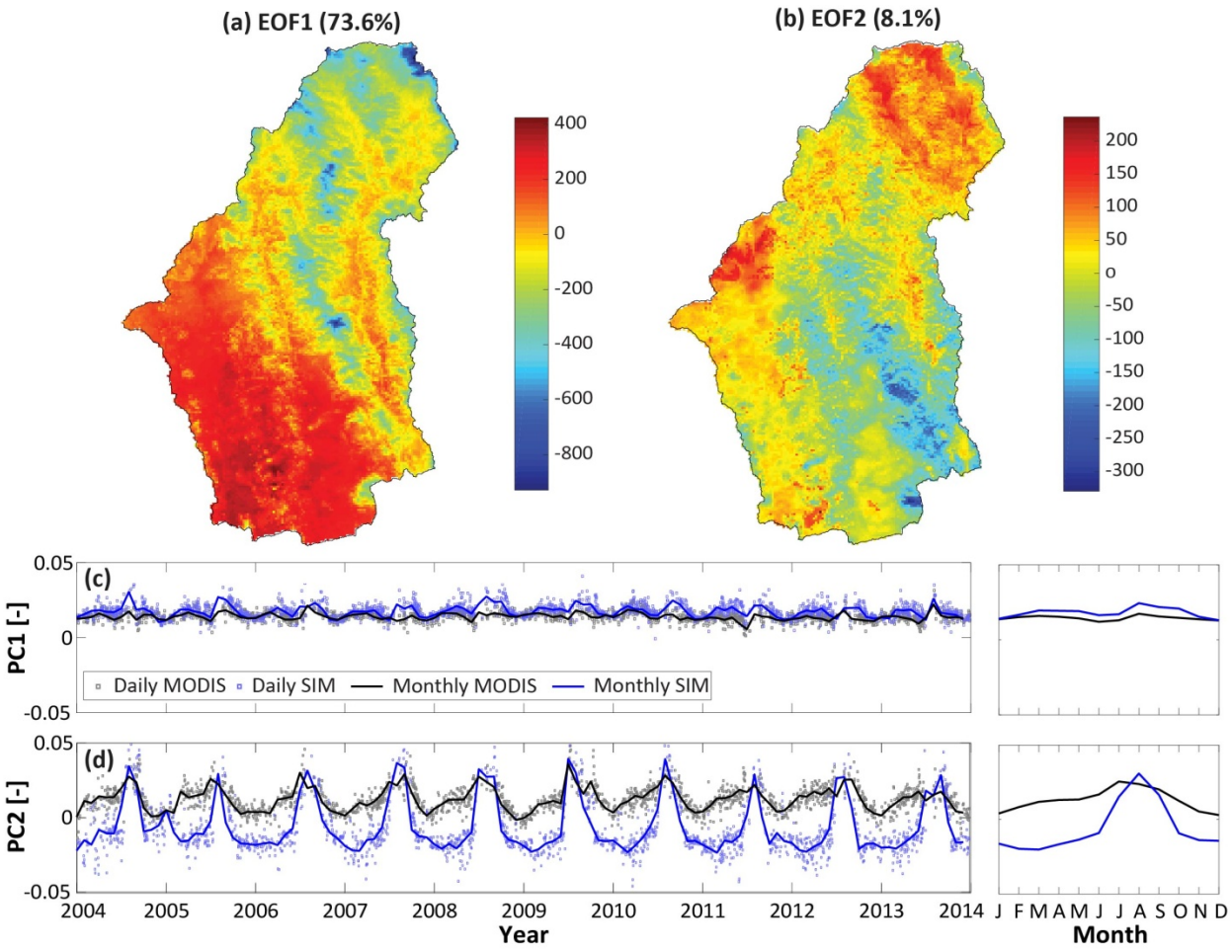

Figure 3.9. (a-b) First two EOFs of simulated and MODIS LST daily spatial patterns with the percentage of the explained variance. (c-d) Left panels: time series of daily and monthly PCs of the two products. Right panels: monthly PCs of the two products averaged across the simulation period.

Physical Controls on Discrepancies between Simulated and MODIS LST Patterns

As a final step, I investigated the physical factors that affect the differences between simulated and MODIS LST patterns. For this goal, I selected 10% of the days with lowest and highest RMSE between the connectivity functions, each including 176 days. The first subset of best performing days mostly occurs in winter months, while the second ones of worst performance are concentrated from April to June, when the basin is very dry. Figure 3.7(b) reports the Taylor diagram for these days, which shows that the worst performing cases have higher nRMSE, lower CC and similar nSTD as compared to the best performing ones. To identify the factors explaining these differences, I compared $\Gamma(F)$ of LST with $\Gamma(F)$ of elevation and VF. Note that VF varies in time and so does its connectivity function. Figures 3.10 and 3.11 show results for one of the best and worst performing days, respectively. For this specific basin, the low-phase $\Gamma(F)$ of elevation is always close to 1. When the model performs well, $\Gamma(F)$ of simulated and MODIS LST are very similar and included between $\Gamma(F)$ of elevation and VF, reflecting the mixed control of these two basin properties on LST. In the worst performing cases, $\Gamma(F)$ of MODIS LST resembles closely that of VF, while $\Gamma(F)$ of simulated LST is always quite close to $\Gamma(F)$ of elevation. Results presented in these two representative days have been found throughout all best and worst performing cases, as quantified by the mean metrics reported in Table 3.7.

To complement and support these findings, I compared the PCs of the best and worst performing cases. PC1 of both LST products and PC2 of simulated LST do not vary significantly across the two subsets of days (not shown). In contrast, PC2 of MODIS

LST increases from a mean of 0.006 across the best performing cases to an average of 0.016 in the worst performing days, implying that the discrepancies between model-derived and remotely-sensed LST emerge when MODIS products are mainly controlled by vegetation properties, as found with the connectivity analysis. A major reason explaining these discrepancies is the fact that the downscaling algorithm used to generate air temperature forcings for tRIBS is based on lapse rates that vary each month (values reported in Liston & Elder, 2006). During days of worst performance, mainly occurring from April to June, the lapse rates are higher (~8 ºC/km) and the effect elevation on simulated LST significantly exceeds that of vegetation. In contrast, during the best performing days, which are concentrated in winter, the lapse rates are lower (~5 ºC/km) and both elevation and vegetation properties affect the simulated LST. To improve hyperresolution hydrologic simulations, the adoption of regionally calibrated monthly lapse rates is then recommended. Overall, these analyses indicate that Taylor diagram, connectivity analysis and EOF analysis should be used in conjunction to quantify complementary aspects of model performances and diagnose potential causes of model errors.

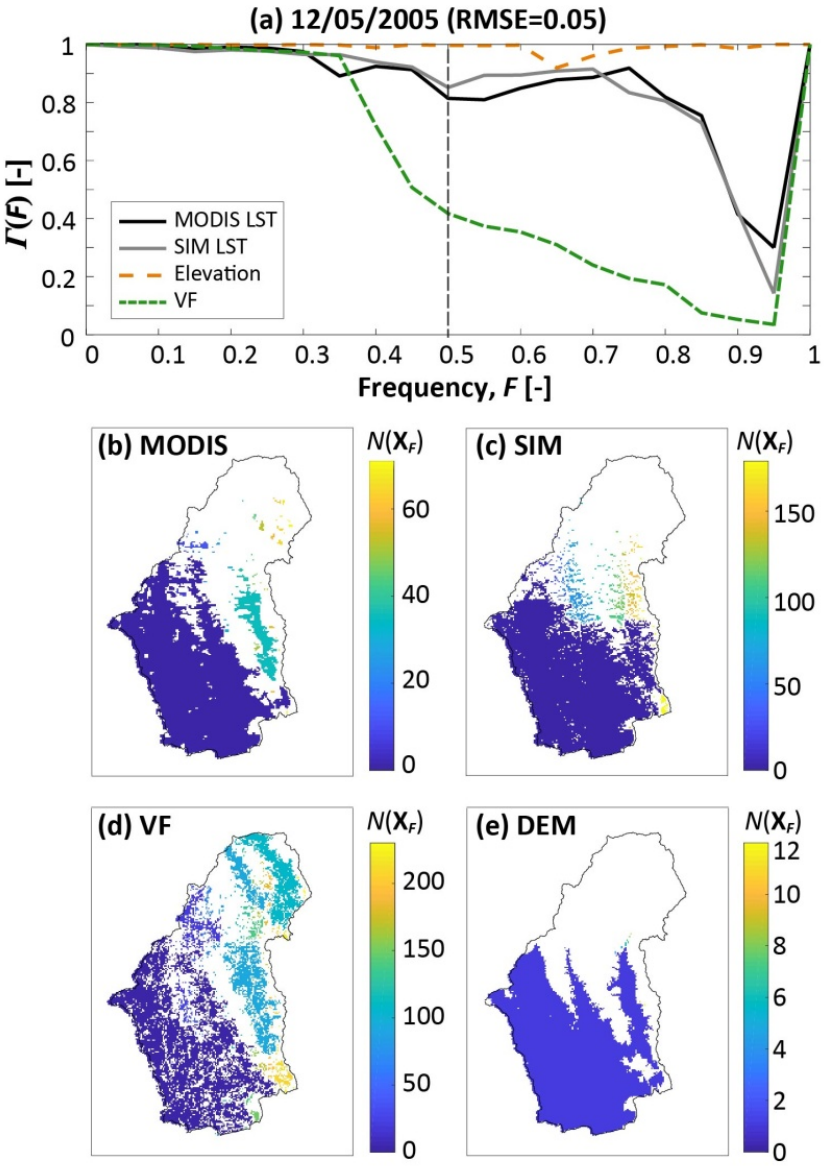

Figure 3.10. (a) Comparison between the connectivity functions of simulated and MODIS LST (high-phase) with those of VF and elevation (low-phase) for 12/5/2005, which is one of the best performing days (CC and RMSE between simulated and MODIS LST of 0.97 and 0.05, respectively). (b-e) Binary images $\mathbf{X}_F$ of all analyzed variables for $F = 0.5$, with indication of the connected clusters.

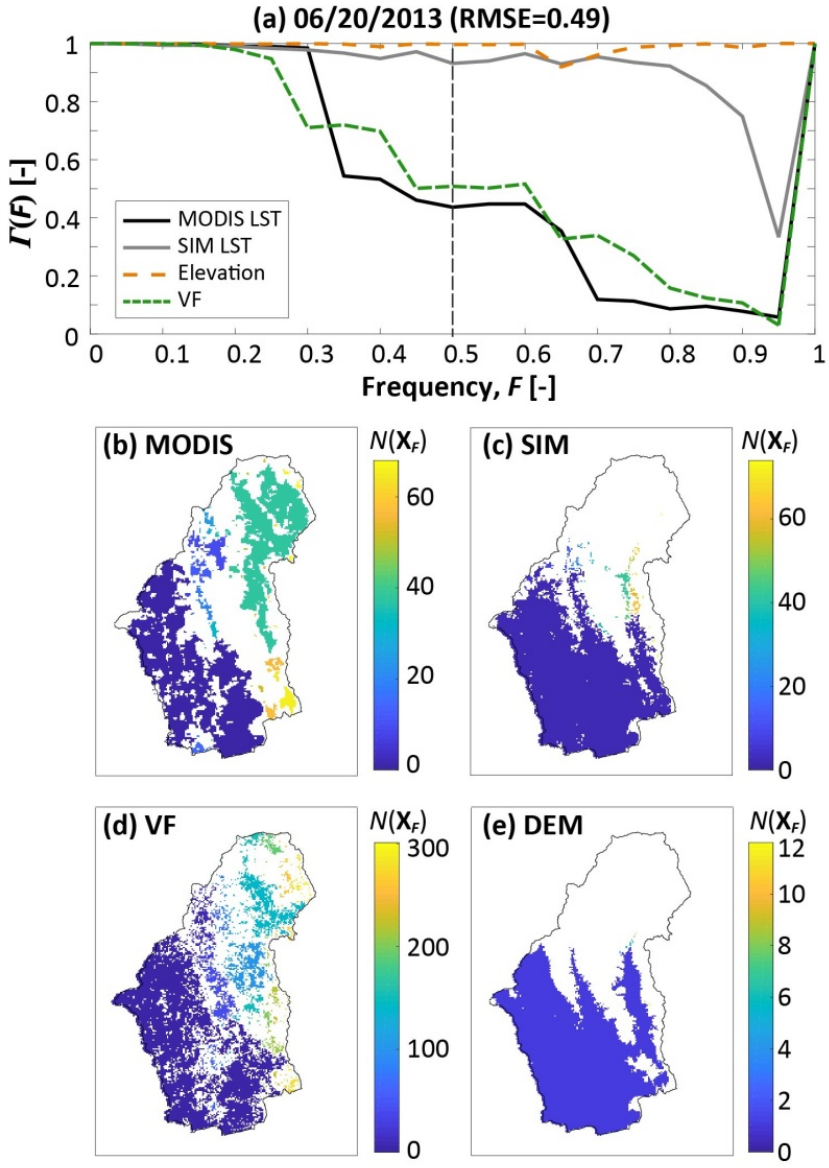

Figure 3.11. Same as Figure 3.10, but for 6/10/2013, which is part of the worst performing days (CC and RMSE between simulated and MODIS LST of 0.54 and 0.49, respectively).

Table 3.7. Mean RMSE and CC across all, worst, and best performing days (see text for details) between the connectivity functions of simulated (SIM) LST, MODIS LST, elevation, and VF.

	RMSE (CC)		
	All	**Worst**	**Best**
MODIS LST-Elevation	0.33 (0.07)	0.45 (0.22)	0.25 (0.03)
MODIS LST-VF	0.27 (0.79)	0.21 (0.85)	0.36 (0.74)
SIM LST- Elevation	0.26 (0.02)	0.25 (0.00)	0.25 (0.02)
SIM LST-VF	0.33 (0.74)	0.36 (0.69)	0.36 (0.74)

Conclusions

Despite recent progress, a number of critical challenges still prevents the achievement of hyperresolution hydrologic simulations. This is especially true for distributed hydrologic models whose application in regional watersheds has been limited by computational, data, and validation challenges. In this study, I proposed strategies to improve the accuracy of hyperresolution hydrologic simulations with a physics-based DHM, by focusing on a regional basin of 21,237 km^2 in northwest Mexico. Since in this study region, like many areas of the world, observations of hydrometeorological variables are sparse and limited, I applied techniques to integrate ground and reanalysis products for the generation of accurate high-resolution (1-km) datasets. These involve bias correcting reanalysis products with ground observations and applying downscaling routines that use terrain information at high resolution, which is available globally. Next, I presented a procedure to generate high-resolution grids of soil parameters that integrates model calibrations conducted at distributed locations, a coarse-resolution soil map based on the FAO dataset, and grids of soil properties at 250-m resolution recently released by ISRIC-WISE at global scale. Finally, I showed that long-term (10 years) hyperresolution

hydrologic simulations obtained with the downscaled hydrometeorological variables and soil parameters were fairly accurate, as they compared well with time series of SM and LST observed at the ground and spatial maps of LST estimated by remote sensors. In particular, I demonstrated the utility of the combined use of Taylor diagram, connectivity analysis, and EOF analysis to quantify complementary aspects of the correspondence between simulated and remotely-sensed spatial patterns, as well as to identify the physical factors that are responsible for the differences between the two products. This is a crucial capability to diagnose model deficiencies and strategize improvement of their physics.

CHAPTER 4

STATISTICAL AND SCALING PROPERTIES OF SIMULATED SOIL MOISTURE

Introduction

As mentioned in Chapters 1 and 2, the investigation of the spatial variability of soil moisture (θ) across different scales requires the availability of high-resolution θ estimates. These have been provided by ground sensors at limited locations and at ~1 km resolution by airborne sensors flown during intensive a limited number of field campaigns, as in the case of the NAFE campaigns analyzed in Chapter 2. As a result, while extremely valuable, the conclusions drawn by studies that have investigated the statistical properties of θ with observed datasets are related to specific locations and times. An alternative strategy to generate high-resolution θ estimates over different regions and times is through the application of hydrologic models, as done in Chapter 3.

Somewhat surprisingly, a very limited number of studies have used θ fields simulated by hydrologic models to investigate the statistical properties of this hydrologic variable. For example, Vivoni et al. (2010) calibrated a physically-based distributed hydrologic model in a small basin in northwest Mexico and used the model outputs to investigate the spatiotemporal variability of soil moisture, including the effect of seasonality of climate and vegetation and the feedback with evapotranspiration fluxes. Manfreda et al. (2007) and Gebremichael et al. (2009) used θ fields simulated by a land surface and a hydrologic model, respectively, to explore the presence of scale-invariant properties. These studies led to important results that helped assessing the utility of hydrologic simulations to analyze the statistical properties of θ. However, these studies

have limitations: Vivoni et al. (2010) focused on a small study site; Manfreda et al. (2007) used model outputs at coarse resolution (~125 km); and Gebremichael et al. (2009) adopted a limited number of metrics to test the scaling properties. Further efforts are then needed to answer the following research question: *do the simulated θ fields exhibit the same statistical properties found by analyzing ground and remotely-sensed observations?* Investigating this question is important because, if simulated θ fields prove to be valuable, they can be used to (i) quantify how different land surface properties and wetness conditions that have not been observed by remote sensors affect the θ statistical properties; and (ii) investigate the existence of scale invariance and multifractality across a larger range of scales that those observed by remote sensors.

Several studies have investigated the statistical properties of soil moisture using observations from remote- and ground-based sensors. A group of these studies has focused on the relation between the spatial mean soil moisture ($<\theta>$) and its variability, measured through the spatial standard deviation (σ_θ) or the coefficient of variation ($CV_\theta = \sigma_\theta / <\theta>$) (e.g., Charpentier and Groffman, 1992; Crow and Wood, 1999; Albertson and Montaldo, 2003; Brocca et al., 2007; Choi and Jacobs, 2007; Lawrence and Hornberger, 2007; Vereecken et al., 2007; Vivoni et al., 2008a; Famiglietti et al., 1999; Mascaro et al., 2011; Mascaro and Vivoni, 2012). These efforts have analyzed data with different supports (the characteristic size of the θ measurement) and extents (the size of the entire area where θ measurements were acquired) collected in a wide range of climatic regions, ranging from arid to temperate and humid. Famiglietti et al. (2008) provide a comprehensive summary of the results across different regions and scales. In general, it

has been found that the relation σ_θ vs. $<\theta>$ is characterized by a convex-upward pattern, where σ_θ increases linearly with $<\theta>$, reaches a maximum value, and decreases as $<\theta>$ approaches porosity. Depending on climate, some or all of these portions of the pattern are observed. Lawrence and Hornberger (2007) attempted to identify the physical controls of this relation, suggesting that: (i) the increasing part of the relation, which is mostly found in arid and semiarid regions, is controlled by wilting point; (ii) the part where σ_θ peaks, which is mainly visible in temperate regions, is affected by soil hydraulic conductivity; and (iii) the decreasing part, which appears in humid sites, is controlled by porosity.

Another group of studies have investigated and found the existence of scale invariance and multifractal properties in spatially-distributed θ estimates retrieved from airborne sensors (Rodriguez-Iturbe et al., 1995; Hu et al., 1997; Kim and Barros, 2002a; Oldak et al., 2002; Das and Mohanty, 2008; Mascaro and Vivoni, 2010). The work presented in Chapter 2 is a further confirmation that θ exhibits this statistical property in two new study regions. Driven by the goal of developing downscaling routines that reproduce multifractality, these studies have analyzed a scale range from ~1 km (the aircraft footprint) to ~30-50 km (the satellite footprint). Given the limited number of field campaigns, further evidences are needed to (i) explore the existence of scale invariance in θ fields at different locations, times, and scale ranges; and (ii) identify the control exerted by terrain, vegetation and soil properties.

Chapter 4 has the goal of filling this research gap by using the hyperresolution simulations of θ obtained in Chapter 3. This extensive dataset is available in a large study

region, the Río Sonora basin (RSB), characterized by highly heterogeneous land surface and terrain features that significantly affect the θ variability, and spans a decade characterized by different wetness states. As a result, the high-resolution θ fields simulated in the RSB offer a unique opportunity to investigate the research question outlined above. For this aim, the simulated θ fields are first aggregated at 1-km resolution, which mimics the aircraft footprint. Next, to sample a large range of terrain, soil, and vegetation characteristics, a total of sixteen square domains of 32-km size (mimicking the satellite footprint) are selected within the RSB. In each domain, the σ_θ vs. $<\theta>$ and CV_θ vs. $<\theta>$ relations are computed and compared with those found with observed datasets. A multilinear regression between the parameters controlling the σ_θ vs. $<\theta>$ relation and the domain terrain, vegetation, and soil properties is used to explore how these properties affect the θ variability. As a next step, the presence of scale invariance is analyzed in each domain following the approach of Mascaro et al. (2010), which is based on Deidda (2000). When multiple scaling regimes are found, the potential reasons explaining the presence of these regimes are investigated focusing on the limitations of simulated θ fields. After eliminating the days where scale invariance is not present, the multifractal downscaling model used by Mascaro et al. (2011a, 2010) and adopted in Chapter 2 is applied to each domain. Following Mascaro et al. (2011b), the model is then used to explore the control exerted by the domain land surface features on θ intermittency and multifractality.

Study Region

The Río Sonora Basin

The study site selected to explore the statistical properties of simulated θ across a range of wetness and land surface conditions is the RSB (drainage area of 21,237 km^2) located in northwest Mexico [Figure 4.1(a)]. Chapter 3 provides a larger description of the basin that is briefly summarized here. Climate in the RSB is arid and semiarid (Peel et al., 2007), with mean annual precipitation (temperature) varying from 350 to 750 mm (11 to 29 ºC). The precipitation regime is characterized by marked seasonality due to the occurrence of the North American Monsoon (NAM) from July to September (Robles-Morua et al., 2012; Vivoni et al., 2010). During the NAM, approximately 40-80% of the total annual precipitation falls in forms of localized thunderstorms with a strong diurnal cycle (Mascaro et al., 2014). Winters are instead relatively dry, apart from a few widespread precipitation events.

Topography in the RSB is quite complex, with elevation (slope) ranging from 200 m to 2618 m (from 0° to 77°) [Figure 4.1(b)]. Soil texture and vegetation in the RSB are also characterized by high spatial heterogeneity. Precipitation seasonality has a significant impact on vegetation, which dramatically greens during the summer season (Forzieri et al., 2011). Figure 4.1(c) shows the map of the time-averaged vegetation fraction (VF) in the RSB derived in 1-km pixels of Moderate Resolution Imaging Spectroradiometer (MODIS, 250-m) imagery by applying the empirical relationships presented in Méndez-Barroso et al. (2014) and Xiang et al. (2014). Figure 4.2(b) shows instead the saturated hydraulic conductivity (K_s) generated by integrating the information

of the soil map provided by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) with the high-resolution (250-m) soil dataset, named SoilGrids250m, recently released by the International Soil Reference Information Centre (ISRIC)-World Soil Information (Hengl et al., 2017). Overall, the strong seasonal variability of climate and vegetation affect the soil moisture spatiotemporal variability within each year (Vivoni et al., 2008a).

Selection of Domains for Statistical Analyses

To conduct the statistical analyses across different land surface features, sixteen square domains of size 32 km by 32 km were selected within the RSB. The domain size has benne selected so that it mimics the resolution of satellite products. The domains cover most of the RSB and do not overlap. This choice allows conducting the analyses under the assumption of statistical independence. Table 4.1 lists the spatial mean and standard deviation of some terrain (elevation and slope), vegetation (VF), and soil (K_s and θ_R; symbols defined in the caption) properties of the domains. To visualize the differences across the sixteen domains, Figure 4.2(a)-(c) show the spatial mean ± standard deviation of elevation, time-averaged VF, and K_s. Domains 1-5 in the upper portion of the basin are characterized by mean elevation >1000 m, relatively higher VF and lower K_s. Domains 7, 8, 11, and 14 have also a relatively high elevation (700-1000 m) and VF, but K_s is higher (~30 mm/h, on average). The other domains are located at lower elevations (< ~700 m), with domains 6, 9, 10 and 12 characterized by the lowest mean VF and relatively high K_s and domains 13, 15, and 16 by slightly higher VF and low K_s. When considering the spatial variability of the properties within the domains, it is

worth noting that: (i) domains 1-5 exhibit lower spatial variability for all properties as compared to the other domains; (ii) the variability of VF is rather constant across all domains; and (iii) domains 6, 13, and 16 have the highest spatial variability of K_s.

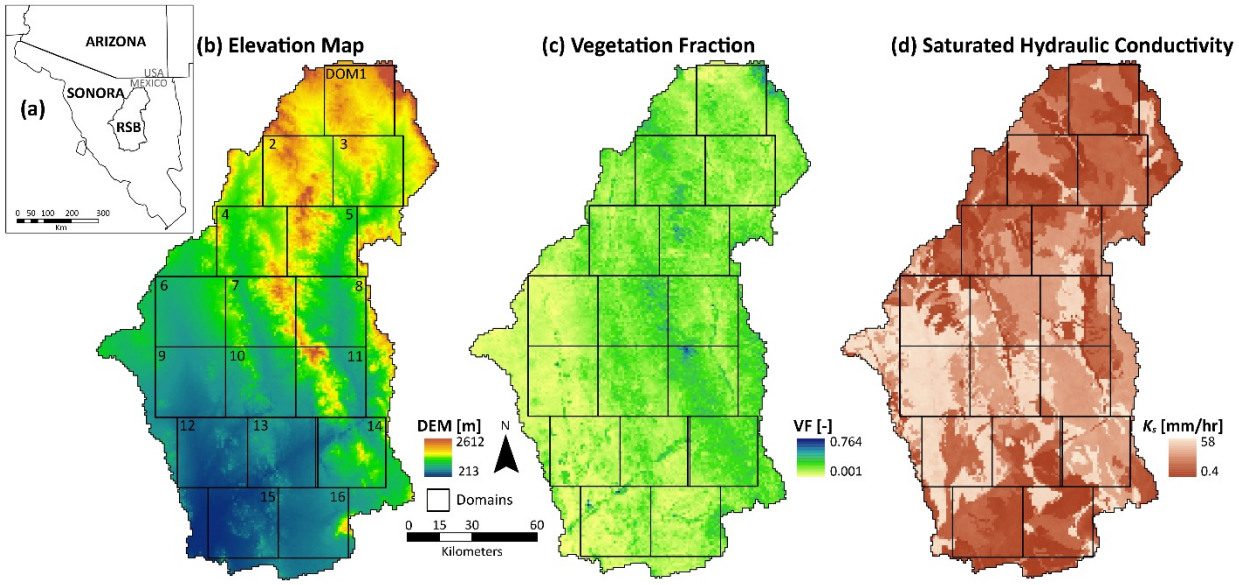

Figure 4.1. (a) Location of the RSB in the state of Sonora, Mexico. (b) Elevation map with the selected sixteen square domains of size 32 km by 32 km. (c) Time-averaged vegetation fraction (VF). (d) Saturated hydraulic conductivity (K_s).

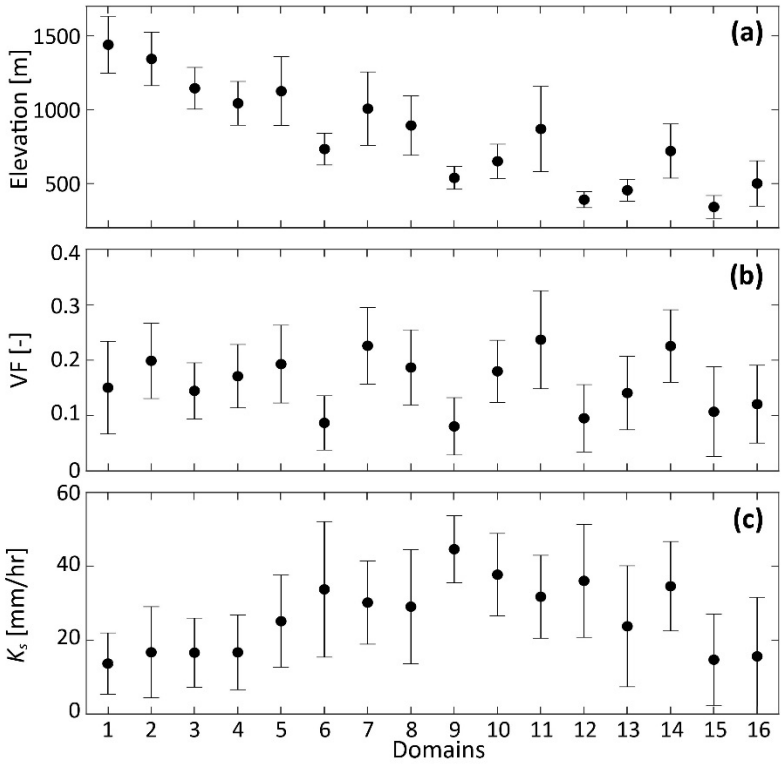

Figure 4.2. Spatial mean ± standard deviation of (a) elevation, (b) time-averaged VF, and (c) K_s in the 16 domains. Basin properties for each domain are described in Table 4.1.

Table 4.1. Basin-averaged properties for the RSB and 16 domains (Dom#): Spatial mean (<->) and standard deviation (σ) of elevation (H), slope (SL), K_s, time-averaged VF, and residual soil moisture $\langle\theta_R\rangle$, as well as the annual precipitation (P).

	<H> [m]	**σ_H [m]**	**<SL> [º]**	**σ_{SL} [º]**	**<VF> [-]**	**σ_{VF} [º]**	**$\langle K_s\rangle$ [mm/hr]**	**σ_{K_s} [mm/hr]**	**$\langle\theta_R\rangle$ [%]**	**σ_{θ_R} [%]**	**P [mm]**
RSB	862.8	418.7	4.38	4.11	0.157	0.084	21.4	16.2	1.23	1.41	416.7
Dom 1	1439.5	192.5	4.55	3.69	0.150	0.084	13.7	8.31	1.99	1.26	464.0
Dom 2	1343.4	180.8	5.78	4.11	0.199	0.068	16.7	12.35	1.16	0.60	424.2
Dom 3	1144.6	141.2	4.31	2.66	0.145	0.051	16.6	9.37	1.03	1.08	450.3
Dom 4	1042.9	147.7	4.85	3.17	0.171	0.057	16.7	10.12	0.92	0.93	380.7
Dom 5	1125.1	233.2	6.45	3.91	0.193	0.070	25.1	12.50	0.95	0.62	428.2
Dom 6	733.2	107.8	1.65	0.93	0.087	0.049	33.8	18.35	0.97	1.34	357.9
Dom 7	1006.5	248.4	6.68	4.03	0.226	0.069	30.2	11.28	0.92	1.05	404.5
Dom 8	892.7	201.5	5.76	4.79	0.187	0.068	29.1	15.46	1.30	1.83	459.5
Dom 9	538.1	77.0	1.53	1.21	0.080	0.052	44.6	9.08	0.56	0.95	359.4
Dom 10	650.6	116.1	3.57	2.53	0.180	0.056	37.7	11.15	1.05	1.43	410.8
Dom 11	869.3	288.3	7.40	5.70	0.237	0.089	31.8	11.24	1.18	1.44	465.3
Dom 12	390.6	54.1	1.63	1.20	0.095	0.061	36.1	15.29	0.41	0.49	369.9
Dom 13	454.4	74.6	2.04	1.67	0.141	0.066	23.8	16.38	0.63	0.48	418.0
Dom 14	719.9	183.2	5.75	4.02	0.225	0.066	34.6	12.11	1.00	1.26	472.8
Dom 15	341.4	77.2	1.90	1.73	0.107	0.081	14.7	12.36	1.28	1.76	378.3
Dom 16	500.2	153.6	2.25	4.07	0.121	0.070	15.6	15.99	0.71	0.63	441.5

Hydrologic Simulations

As described in the Introduction of this Chapter, θ outputs of the hyperresolution hydrologic simulations of Chapter 3 were used here to conduct the statistical analyses. In the following, I provide a brief overview of the setup of the hydrologic model, the datasets used, and the numerical simulations. I refer the reader to Chapter 3 for additional details. The TIN-based Real-time Integrated Basin Simulator (tRIBS, Ivanov et al., 2004a, 2004b) was used to perform the hyper-resolution simulations in the RSB. tRIBS is a fully distributed, physically-based hydrological model that can reproduce the coupled water-energy balance in a continuous fashion. tRIBS is a computational efficient model suitable for hyperresolution simulations, since (i) it is based on an irregular domain, which allows capturing the details of complex topography with minimal loss of information as compared to grid-based models (Vivoni et al., 2005); (ii) it uses approximate solutions in the infiltration and routing schemes (Cabral et al., 1992); and (iii) its code has been parallelized to be run in high performance computing clusters (Vivoni et al., 2011).

To setup tRIBS in the RSB, a TIN domain of 2,612,662 nodes was generated from a 30-m digital elevation model, corresponding to an average cell size of ~88 m and 11.4% of the original DEM pixels. Hydrometeorological forcings were generated by combining the NLDAS reanalysis products at 12-km resolution with ground observations from different networks of rain gauges and weather stations. Meteorological data, including atmospheric pressure, relative humidity, wind speed, and incoming solar radiation, were downscaled at 1-km resolution using the approach of Liston and Elder

(2006) that accounts for variability of terrain properties. Time-varying grids of vegetation parameters at 250- to 1-km resolution were obtained from remotely-sensed imagery of MODIS. The spatial variability of soil properties was characterized by combining the coarse CONABIO map with the SoilGrids250m datasets. Specifically, grids at 250-m resolution of K_s, residual moisture content (θ_r), saturated moisture content (θ_s) and porosity (n) were generated by incorporating the SoilGrids250m maps into the CONABIO soil map. The other model parameters were assumed constant within the polygons of the CONABIO map. The hydrologic model was calibrated against soil moisture observations at distributed locations that cover most of the soil classes of CONABIO. The model was then validated against independent observations, including time series of soil moisture and land surface temperature (LST) recorded at a network of stations, and spatial patterns of LST estimated by MODIS. Simulations were conducted from 1 January 2004 to 31 December 2013. The validation effort included the comparison of observed and simulated time series of θ and LST, as well as the use of multiple techniques to test the correspondence of simulated and observed hyperresolution spatial patterns of LST. Overall, the validation analyses provided confidence in the model's ability to capture the hydrologic response of the complex, regional watershed. The dataset of soil moisture spatial outputs used here consists of 3653 daily maps. Each map was resampled from the irregular mesh to a 1-km regular grid.

Methods

Three main types of statistical analyses were conducted on the simulated soil moisture fields, which are described in the following subsections.

Relation between Spatial Mean and Variability of Soil Moisture

The spatial variability of θ was first analyzed through the relations σ_θ vs. $<\theta>$ and CV_θ vs. $<\theta>$, which were computed over the entire basin and at each domain. σ_θ represents the standard deviation of the soil moisture values in all 1-km pixels included within the 32-km coarse domains. CV_θ is simply equal to $\sigma_\theta/<\theta>$, To interpret these relations, the empirical functions proposed by Famiglietti et al. (2008) were utilized. The relation CV_θ vs. $<\theta>$ can be interpreted through the negative exponential equation:

$$CV_\theta = a \cdot \exp(-b < \theta >), \tag{4.1}$$

which can be rearranged to provide the curve interpreting the σ_θ vs. $<\theta>$ relationship:

$$\sigma_\theta = a < \theta > \cdot \exp(-b < \theta >), \tag{4.2}$$

where a and b are parameters that were estimated at each domain. To analyze the control on these relations of land surface properties, a principal component analysis was performed on terrain (slope), soil (K_s and θ_R), and vegetation (VF) properties of all domains (hereafter, ancillary factors). Next, parameters a and b were linked to the dominant principal components through a multilinear regression.

Scale Invariance and Multifractality

The second type of statistical analyses was the investigation of scale invariance and multifractality according to the approach of Mascaro et al. (2010). Methodological details of these analyses are provided in Chapter 2 and, for completeness, they are

summarized here. The presence of scale invariance was searched from the coarse scale of L = 32 km to the fine scale l = 1 km. Note that a binary approach was adopted, where $L = l \cdot 2^{N_{lev}}$, where N_{lev} = 5 is the number of the downscaling levels. The presence of scale invariance is identified by verifying the linearity of log-log relation

$$\log S_q(\lambda) \sim K(q) \cdot \log \lambda, \tag{4.3}$$

where $S_q(\lambda)$ is the structure function computed from the soil moisture field aggregated at scale at scale $l \leq \lambda \leq L$ for a given moment q (usually ranging from 1.5 to 4):

$$S_q(\lambda) = \frac{1}{N(\lambda)^2} \sum_{i=1}^{N(\lambda)} \sum_{i=1}^{N(\lambda)} \left[\theta_{i,j}(\lambda)\right]^q \tag{4.4}$$

where $\theta_{i,j}(\lambda)$ is the mean value of θ on a grid cell λ x λ in the location (i, j) within the domain and $N(\lambda)^2 = (L/\lambda)^2$ is the number of λ x λ cells embedded in the coarse domain.

If equation (4.3) is verified, then a scaling regime is found from l to L and the slope $K(q)$ of (4.3) can be estimated through linear regression. The multifractal behavior is investigated by inspecting the relation linking $K(q)$ with q: if it is linear, the field is fractal; otherwise the field is multifractal (non-linear) and $K(q)$ is called multifractal exponents. $K(q)$ controls the intermittency of the field, with larger (smaller) values corresponding to more intermittent (smoother) fields. Note that, if this analysis is applied to two fields, where one is obtained from the other by adding a bias, results are different, and the differences become larger as the bias increases.

The Multifractal Downscaling Model

For the simulated θ fields exhibiting the presence of scale invariance and multifractality, the downscaling model proposed by Deidda (2000) was used to explore

how terrain, soil, and vegetation affect the multifractal properties of θ. The model is based on a log-Poisson stochastic generator of homogeneous, random binary cascades that reproduce the observed scale-invariance and multifractal properties. The generator depends on two parameters, c and β, that control the theoretical expectation for the multifractal exponents (Deidda et al., 1999; Deidda, 2000):

$$K(q) = c \cdot \frac{q(1-\beta) - (1-\beta^q)}{\ln(2)}. \tag{4.5}$$

Parameters c and β are estimated by fitting (4.5) to the observed multifractal exponents $K(q)$. Then, empirical calibration relations are identified between c and β and coarse-scale predictors. Previously, Mascaro et al. (2011a, 2010) and the work in Chapter 2 (Ko et al., 2016) found β to be fairly constant, while c was linked to the coarse-scale $<\theta>$ according to an exponentially decreasing relation:

$$c = c_\infty + \alpha \cdot e^{-\gamma \cdot <\theta>} \tag{4.6}$$

with parameters c_∞, α and γ. As done for parameters a and b of equation (4.2), the regression parameters of (4.6) were linked to the dominant principal components of the ancillary factors via multilinear regression. This relation will be used to explore the impact of land surface properties on the multifractal properties of simulated θ fields.

Results and Discussion

Interannual and Seasonal Variability of Soil Moisture

Prior to exploring the statistical properties of simulated θ, an overview of the interannual and interannual variability of $<\theta>$ is given in Figure 4.3, which shows the time series of $<\theta>$ simulated during the period from 2004 to 2013 in the RSB and in

selected domains. The panels on the right reports the corresponding long-term monthly means of $<\theta>$, along with the spatial standard deviation (bars). The distributed tRIBS simulations were able to (i) capture the response of θ to individual storms and the recession behavior during the interstorm periods; (ii) represent well the seasonality due to the NAM, as shown by the high values of $<\theta>$ and σ_θ during summer; and (iii) reproduce the effect of spatially-variable forcings and land-surface properties, as shown by the differences of $<\theta>$ across the domains.

Spatial Variability of Simulated Soil Moisture

Figure 4.4 shows the σ_θ vs. $<\theta>$ and CV_θ vs. $<\theta>$ relations obtained with daily data of all years in the entire RSB. Different markers are used for summer (July-September), Winter (November-March), and other seasons (October, April, May). The solid lines are equations (4.1) and (4.2) fitted to all points. The equations were also separately fitted to winter and summer points and no statistically significant differences at the 0.05% level (i.e., using the 95% confidence intervals) were found. Note that the availability of a long-term record of θ allowed analyzing a similar range of $<\theta>$ in both seasons, although winter is generally drier. This result indicates that the distinct hydrologic processes occurring in winter and summer do not significantly affect how, on average, σ_θ is related to $<\theta>$. Considering now the solid lines, Figure 4.4(a) shows that σ_θ increases with $<\theta>$ until it reaches around 15%, and, then, decreases beyond that value, when the field becomes more spatially uniform, likely due to the impact of precipitation covering large areas. Figure 4.4(b) shows instead that CV_θ exponentially

decreases with increasing $<\theta>$, implying the relative variability decreases with higher moisture contents.

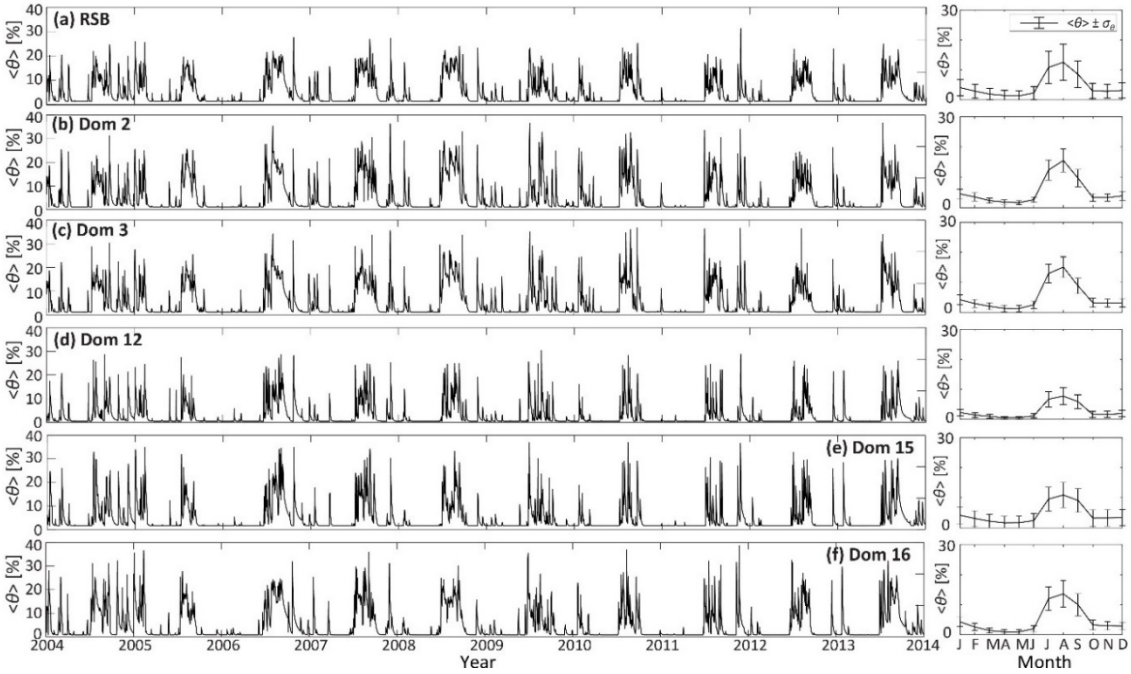

Figure 4.3. (a) Basin-averaged simulated soil moisture ($<\theta>$) from 2004 to 2013 for the RSB. (b-f) Same as (a) but for representative domains. Right panels: monthly $<\theta>$ with the monthly averaged standard deviation of θ (σ_θ).

Results presented in Figure 4.4 were compared with findings of Famiglietti et al. (2008) and Mascaro and Vivoni (2010), who applied the same analysis in different study sites on domains of ~50 km by ~50 km, representing the largest extent that was analyzed. Famiglietti et al. (2008) computed the relations (4.2) and (4.3) using ground observations collected in humid regions of USA during the Soil Moisture Experiment 2002 and 2003 (SMEX02 and SMEX03). Mascaro and Vivoni (2010) performed the analysis in a portion of the RSB, using aircraft θ estimates from the Polarimetric Scanning Radiometer (PSR)

sensor at 800-m resolution collected during the SMEX04 experiment in summer 2004. As shown in Figure 4.4, the relations estimated on the simulated θ fields are very similar to those found on both observed datasets when considering the increasing portion. The relations tend to differ as $<\theta>$ reaches ~15%, especially considering SMEX04. This can be explained considering that: (i) the SMEX04 dataset includes only 11 θ maps and that, during this experiment, $<\theta>$ exceeded 15% only once, and (ii) the extent of the analyses is different. Parameters a and b of the lines are reported in Table 4.2, along with and the root mean square error (RMSE) between each relation and the daily simulated θ. Despite differences between the σ_θ vs. $<\theta>$ lines emerge for $<\theta> >$ ~15%, the RMSE is quite similar across all lines.

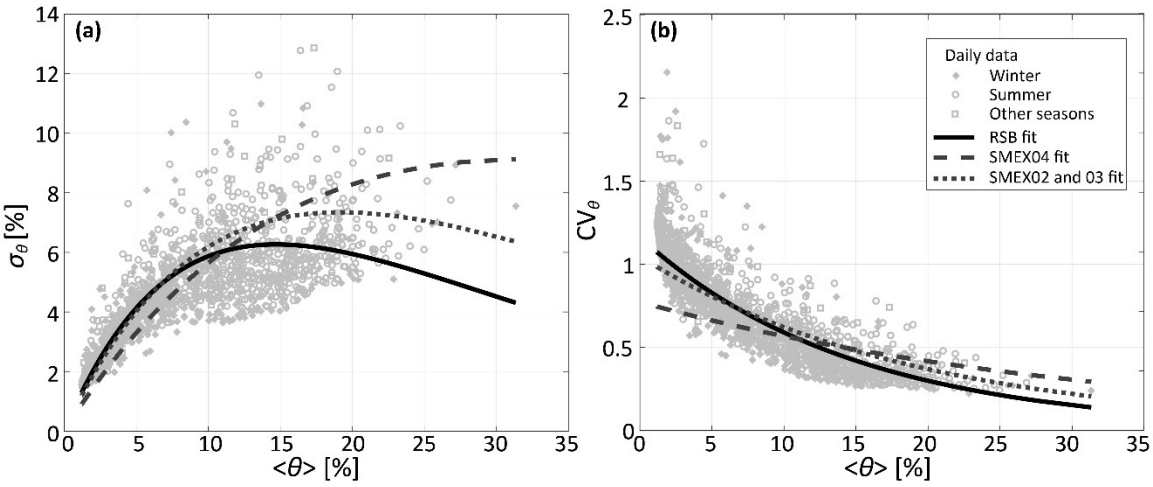

Figure 4.4. (a) Relation between the daily spatial standard deviation σ_θ [%] and spatial mean $<\theta>$ [%] in the RSB along with the empirical fit $\sigma_\theta = <\theta> \cdot a \cdot e^{-b<\theta>}$ estimated for all seasons. Makers indicate seasons: winter (November-March) and summer (July-September). (b) Same as (a) but for CV_θ, with the line $CV_\theta = a \cdot e^{-b<\theta>}$. In each panel, the empirical curves obtained with two observed aircraft-derived datasets of (i) SMEX04 in Sonora (Mascaro and Vivoni, 2010) and (ii) SMEX02 and SMEX03 (Famiglietti et al., 2008) are also shown.

Table 4.2. Regression parameters of equations (4.2) and (4.3) for the RSB, SMEX04 (Mascaro and Vivoni, 2010), and SMEX02 and SMEX03 (Famiglietti et al., 2008), along with corresponding RMSE between points and curve.

	a	***b***	**RMSE (σ_θ)**	**RMSE (CV_θ)**
RSB	0.804	0.070	0.844	0.107
SMEX04	0.770	0.031	1.030	0.383
SMEX02-03	0.884	5.807	0.947	0.325

To explore the reason of the differences between simulated and SMEX04-PSR, I obtained the SMEX04 gridded datasets estimated using not only PSR but also 2D Synthetic Aperture Radiometer (2D-STAR). These are microwave sensors measuring near-surface (0-5 cm) brightness temperature at different wavelengths. A total of sixteen images (11 from PSR and 5 from 2D-STAR) with an extent of 96 km × 52.8 km at 800-m were available. Simulated θ fields were obtained at the time of the day close to when PSR and 2D-STAR were flown and, then, converted into 800-m gridded maps. The flight dates and times, and the corresponding time of the selected simulated θ fields are listed in Table 4.3 (Ryu et al., 2010; Bindlish et al., 2008). Note that the flight time of PSR and 2D-STAR is the same. When σ_θ and $<\theta>$ were calculated, the pixels with no data in PSR and 2D-STAR datasets were excluded from the simulated θ fields. For this reason, the simulated values of σ_θ and $<\theta>$ for each day were slightly different for each case.

As shown in Figure 4.5, the σ_θ vs. $<\theta>$ and CV_θ vs. $<\theta>$ relations estimated on PSR and 2D-STAR datasets are significantly different, due to the diverse ability of the sensors and the retrieval algorithm. For both PSR and 2D-STAR, the σ_θ vs. $<\theta>$ relations are increasing. However, the values of σ_θ of 2D-STAR and the range of $<\theta>$ are much lower than those derived from PSR. The relation obtained from the simulations is instead

convex-upward with σ_θ increasing with increasing $<\theta>$ until it reaches ~10 % and, then, decreasing up to ~15%. In general, the simulated values underestimate σ_θ and overestimate $<\theta>$ of both PSR and 2D-STAR. These differences can be caused by: (1) errors and coarse resolution of in the precipitation products used to force the hydrologic simulations; and (2) time difference between the observations and simulations.

Despite this difference, the simulated relations are closed to those obtained with the 2D-STAR soil moisture estimates. 2D-STAR, which uses dual-polarized L-band brightness temperature, has been considered able to produce more accurate θ estimates than C-band or X-band, that PSR uses, because the influence of vegetation, soil roughness, and radio interference can be minimized at L-band (Calvet et al., 2011; Cui et al., 2017). Overall, the analyses at basin scale based on ten years of simulations (see Figure 4.4) and those focused on the days and extent of the SMEX04 datasets (see Figure 4.5) showed that the model is able to simulate reasonably well the patterns estimated from observed datasets.

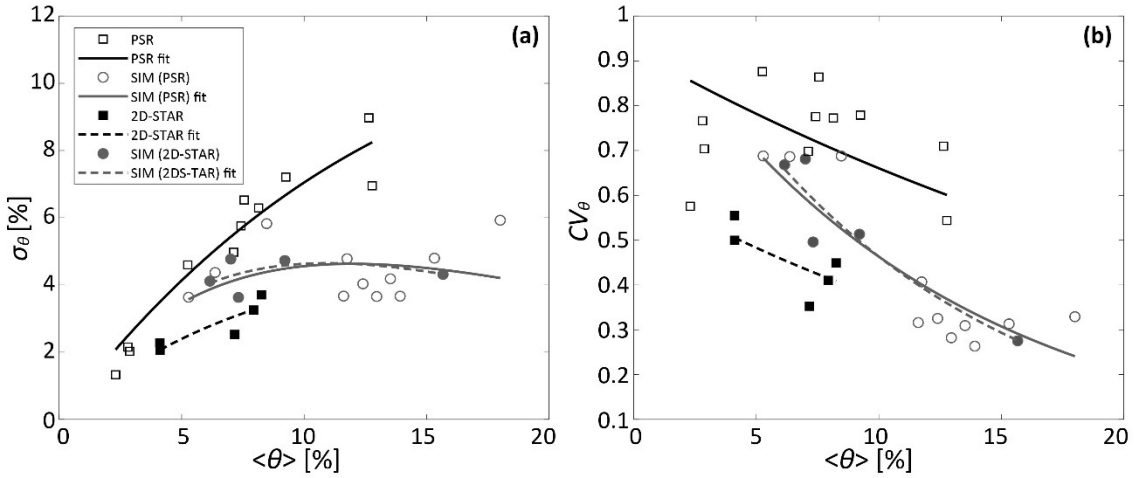

Figure 4.5. Same as Figure 4.4 but with SMEX04 datasets obtained from PSR and 2D-STAR sensors and simulations at the same date and close time when both sensors were flown.

Table 4.3. SMEX04 PSR and 2D-STAR local flight times over the Sonora site and corresponding time of the simulated θ.

Date in 2004	Time	PSR	2D-STAR	SIM
5	15:14–16:01	x		16:00
7	11:35–12:40	x	x	12:00
8	9:38–10:14	x	x	10:00
9	9:23–10:11	x		10:00
10	9:19–10:06	x		10:00
12	9:36–10:26	x		10:00
13	9:37–10:24	x		10:00
14	9:36–10:25	x		10:00
24	9:46–10:32	x	x	10:00
25	13:11–14:00	x	x	14:00
26	9:36–11:17	x	x	11:00

The σ_θ vs. $<\theta>$ relation was then analyzed at each domain. Results are shown in Figure 4.6, while regression parameters of the fitted line and RMSE are reported in Table 4.4. Equation (4.2) was first estimated locally (LOC, gray lines), obtaining a set of a and b parameters at each domain. Next, to investigate how the σ_θ vs. $<\theta>$ relation is affected by terrain, soil, and vegetation properties, a principal component analysis (PCA) was

applied to a total of eight ancillary factors computed for each domain (Table 4.1), including topographic (<SL> and σ_{SL}), vegetation (<VF> and σ_{VF}), and soil variables (<K_s>, σ_{K_s}, <θ_R>, and σ_{θ_R}). The PCA was applied to reduce the dimensionality of the dataset, while, at the same time, capturing the essential features of each domain. Note that elevation was not used because it did not appear to be related to *a* and *b*. By an orthogonal rotation of the analyzed dataset, the PCA identifies a minimal number of principal components (PCs) that explain the largest variability of the original dataset. As reported in Table 4.5, the first three PCs explain 99% of the variability. The scores of the PCs (PCS_j, j = 1, 2, 3) were then linked to *a* and *b* with a multilinear regression as:

$$\begin{cases} a = m_0 + m_1 PCS_1 + m_2 PCS_2 + m_3 PCS_3 \\ b = n_0 + n_1 PCS_1 + n_2 PCS_2 + n_3 PCS_3 \end{cases} \tag{4.7}$$

with parameters m_j and n_j. Table 4.5 also reports the coefficients of the first three principal components (PCs) and the correlation coefficients (CCs) between the scores of the PCs, which support the physical interpretation of the PCs. The first PC (86% of the variance) captures the variability of K_s, as revealed by the high CC of 0.9998, followed by <θ_R> with CC of -0.47 and <VF> with CC of -0.39. As a result, soil texture and, to a minor extent, vegetation play an important role on the spatial variability of soil moisture. The second and third PCs account for different aspects of topography, soil and vegetation.

Estimates of the regional parameters m_j and n_j are reported in Table 4.6. These values can be used to derive *a* and *b* starting from the domain ancillary factors. Results of this regression (labeled ANC) are plotted in Figure 4.6 (black lines). The parameters of *a* and *b* estimated by (4.7) and the RMSE are reported in Table 4.4. To compare ANC and

LOC performances, I calculated the percent difference of RMSE (α_{RMSE}) with the metrics obtained in LOC. The α_{RMSE} was calculated as

$$\alpha_{\mathrm{RMSE}} = \frac{-(\mathrm{RMSE}_{\mathrm{ANC}} - \mathrm{RMSE}_{\mathrm{LOC}})}{\mathrm{RMSE}_{\mathrm{LOC}}} \cdot 100, \tag{4.8}$$

where $\mathrm{RMSE}_{\mathrm{ANC}}$ and $\mathrm{RMSE}_{\mathrm{LOC}}$ are the metrics computed in ANC and LOC approaches. The negative values of α_{RMSE} indicate a degradation in the fit derived from ANC as compared to LOC. Even if ANC has lower performances than LOC, the results show the ANC calibration method captures the local behavior at most of domains, with lower skills in domains 14 and 15 (see α_{RMSE} reported in Table 4.4). These two domains exhibit the drastic linear change in the variation with small changes in $\langle\theta\rangle$ for dry conditions and the maximum θ variance is observed early than other domains.

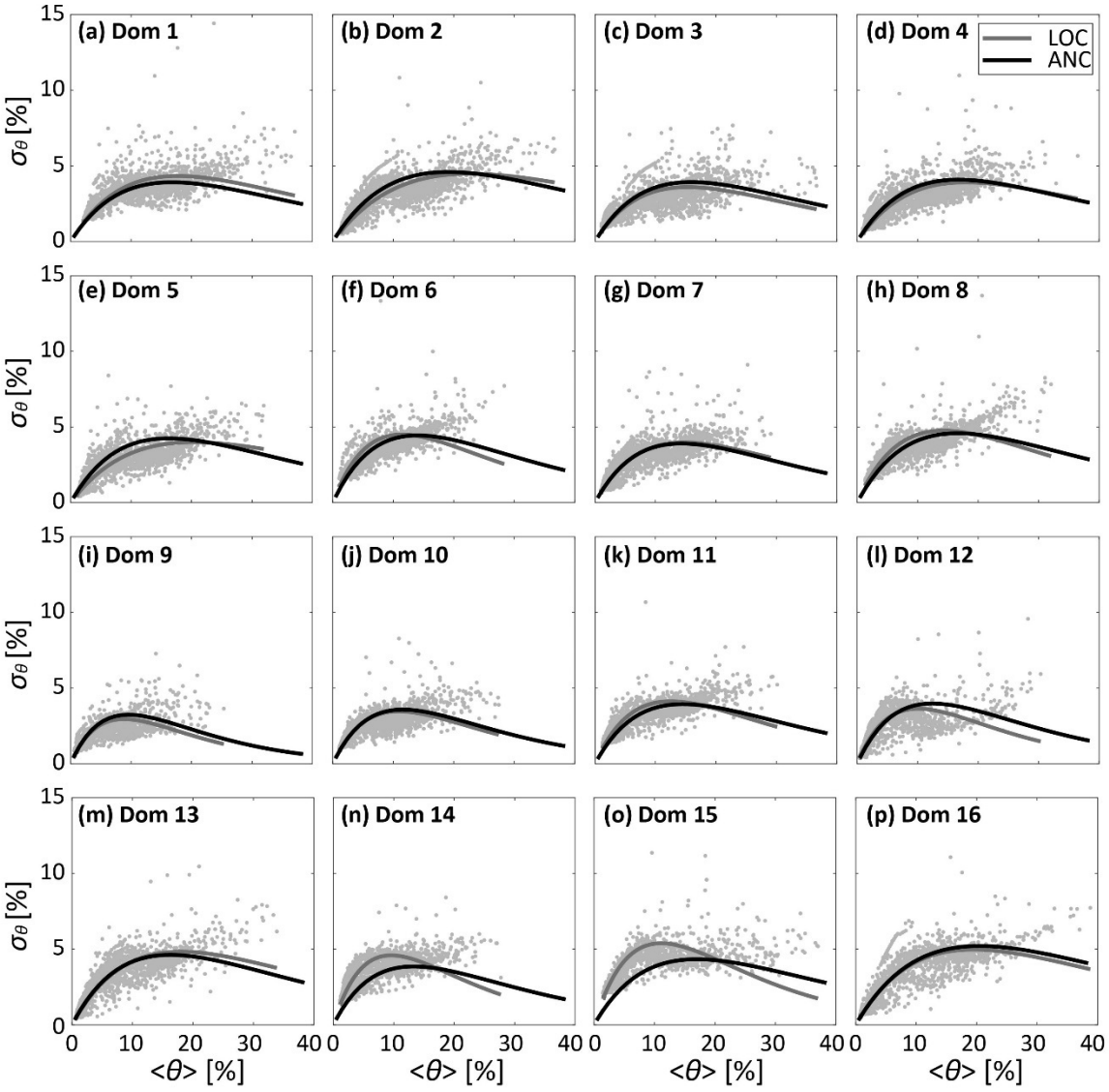

Figure 4.6. Relations between σ_θ and $<\theta>$ [%] at each domain along with the empirical fit derived from LOC (gray lines) and ANC (black lines) regressions.

Table 4.4. Regression statistics for σ_θ vs. <θ> for each domain. RMSE estimated between the data and the empirical regression fits derived from LOC and ANC approaches and α_{RMSE}. The average, standard deviation (STD), max and min were calculated from domains.

	LOC Regression			**ANC Regression**			
	a	*b*	RMSE [%]	*a*	*b*	RMSE [%]	α_{RMSE} [%]
Dom 1	0.665	0.056	0.630	0.636	0.060	0.661	-4.9
Dom 2	0.538	0.044	0.600	0.655	0.052	0.639	-6.5
Dom 3	0.640	0.065	0.620	0.671	0.063	0.645	-4.0
Dom 4	0.593	0.056	0.570	0.665	0.060	0.592	-3.9
Dom 5	0.535	0.049	0.550	0.713	0.062	0.631	-14.7
Dom 6	1.069	0.087	0.670	0.862	0.071	0.718	-7.2
Dom 7	0.730	0.068	0.520	0.745	0.070	0.518	0.4
Dom 8	0.908	0.070	0.660	0.751	0.060	0.708	-7.3
Dom 9	0.909	0.113	0.500	0.920	0.105	0.514	-2.8
Dom 10	1.128	0.121	0.510	0.843	0.087	0.509	0.2
Dom 11	0.895	0.080	0.510	0.737	0.069	0.558	-9.4
Dom 12	0.980	0.099	0.560	0.871	0.080	0.583	-4.1
Dom 13	0.732	0.056	0.560	0.773	0.062	0.570	-1.8
Dom 14	1.303	0.104	0.550	0.790	0.075	0.857	-55.8
Dom 15	1.320	0.090	0.640	0.695	0.059	1.203	-88.0
Dom 16	0.695	0.051	0.580	0.692	0.049	0.586	-1.0
Average	0.853	0.076	0.577	0.751	0.068	0.656	-13.2
STD	0.246	0.023	0.053	0.083	0.013	0.166	23.2
Max	1.320	0.121	0.670	0.920	0.105	1.203	0.4
Min	0.535	0.044	0.500	0.636	0.049	0.509	-88.0

Table 4.5. Ancillary factors selected for the PCA. PCs with percent of variance explained in parentheses are reported along with CCs with *PCS*s. <-> and σ are used to indicate the spatial mean and standard deviation, respectively. SL is slope.

	PC1 (86%)	PC2 (9%)	PC3 (4%)	CC with PCS_1	CC with PCS_2	CC with PCS_3
<SL>	-0.0256	-0.4416	0.6969	-0.1216	**-0.6688**	**0.7183**
σ_{SL}	-0.0373	-0.2296	0.5157	-0.2559	-0.5013	**0.7663**
<VF>	-0.0003	-0.0096	0.0176	-0.0517	**-0.5841**	**0.7315**
σ_{VF}	-0.0005	-0.0009	0.0020	**-0.3879**	-0.2374	0.3707
$<K_s>$	0.9981	-0.0533	0.0187	**0.9998**	-0.0170	0.0041
σ_{K_s}	0.0377	0.8638	0.4967	0.1267	**0.9237**	0.3615
$<\theta_R>$	-0.0175	-0.0483	0.0239	**-0.4702**	-0.4141	0.1396
σ_{θ_R}	0.0040	-0.0289	0.0191	0.0902	-0.2084	0.0941

Table 4.6. Parameters m_j and n_j of ANC calibration relation.

Parameter	Value
m_0	0.751
m_1	0.008
m_2	0.008
m_3	-0.009
n_0	0.068
n_1	0.001
n_2	-0.001
n_3	-0.003

To analyze the impact of geophysical controls on the spatial variability of θ fields, a simple comparison has been performed between domains with contrasting terrain, vegetation, and soil properties. Figure 4.7(a) compares domains with contrasting terrain properties: domain 3 with higher <H> and <SL> has smaller σ_θ across $< \theta >$ than domain 16, possibly due to the lower evaporation rate at higher elevation areas with steep slopes. Figure 4.7(b) presents the comparison between two domains with contrasting vegetation properties: domain 2 with higher <VF> shows higher spatial variability with the slow decreasing rate as $<\theta>$ increases when $<\theta>$ is larger than ~10%. This can be due to both different evapotranspiration rates and sampling variability. Finally, Figure 4.7(c) shows the comparison between two domains with contrasting soil properties: domain 12 with higher K_s shows lower spatial variability than domain 15, implying that faster draining is associated with less spatial heterogeneity of θ. Additional analyses are needed to investigate more systematically all these aspects that have been here discussed just for a few representative cases.

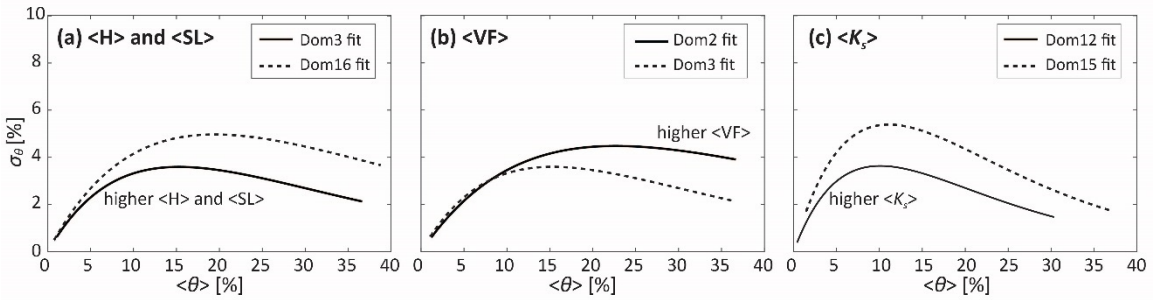

Figure 4.7. Relation between σ_θ and $<\theta>$ for (a) domain 3 and 16 having a large contrast in terrain properties (<H> and <SL>), (b) domain 2 and 3 having a large contrast in vegetation property (<VF>), and (c) domain 12 and 15 having a large contrast in soil property ($<K_s>$).

Scale invariance and multifractality of the simulated soil moisture fields

Scale invariance analysis was conducted for the moments q = 1.5, 2, ...,4 in the simulated θ fields in each domain. Following Mascaro et al., (2010), the presence of scale invariance was identified if the correlation coefficient (CC) of the regression for q = 4 was higher than 0.9. Additionally, I used the root mean square error (RMSE) of the regression larger than 0.4 to exclude the days having multi-scaling regimes due to the impact of artificial discontinuity of P at 12 km and the impact of a model input for θ_R that is not scale-invariant. Figure 4.8 shows examples of the scale-invariance analysis in wet and dry days in domains 2 and in residual soil moisture in domain 2, 12 and 15. The presence of scale invariance was tested by verifying the linearity of (4.1) through the estimation of RMSE and CC for q = 4. Table 4.7 reports results of the mean RMSE and CC across the scale-invariant θ fields as well as the number of discarded θ fields, the mean of $<\theta>$ of discarded θ fields and θ_R of each domain. Overall, most of θ fields (97%) in majority domains (70%, 11 out of 16) are scale-invariant, except 5 domains having scale-invariant θ fields less than 45% of θ fields. These discarded θ fields in the 5

domains are generally very dry with the values $<\theta>$ close to $<\theta_R>$ that are not scale-invariant in these domains, implying that the model input for the soil parameters (e.g., θ_R and K_s) could break the scale invariance property that natural θ field exhibits. Figure 4.8(e) and (e1) show the impact of soil input clearly. A soil texture class having very high θ_R value than the rest soil classes plays a similar role of irrigation in very dry lands and it brakes the scale-invariance property of $<\theta>$ field. While the discarded θ fields in the other domains have a relatively large mean of $<\theta>$ (e.g., domain 1, 2, and 10), indicating that the unnatural discontinuity of P at 12 km could affect the scale invariance property [see Figure 4.8(c1)].

As I found in Chapter 2, there might be an impact of irrigation on scale invariance properties in the RSB. Scale invariance analyses were already conducted in the Rio Sonora basin in two previous studies, including Mascaro et al. (2011) and Mascaro and Vivoni (2010), using SMEX04 remotely sensed soil moisture fields. Both studies identified the presence of scale invariance from 0.8 km to 25.6 km and 51.2 km, respectively. Mascaro et al. (2011) also conducted preliminary analyses of the effect of irrigated fields. These authors detected pixels that were consistently wetter or drier during the entire duration of the experiment. However, they found that there is no direct linkage between wet pixels and cropland. Similarly, Mascaro and Vivoni (2010) did not find any impact of irrigation. As a result, based on these two studies, I assumed that the influence of irrigated fields in the domain on the scale invariance properties might be negligible or not present.

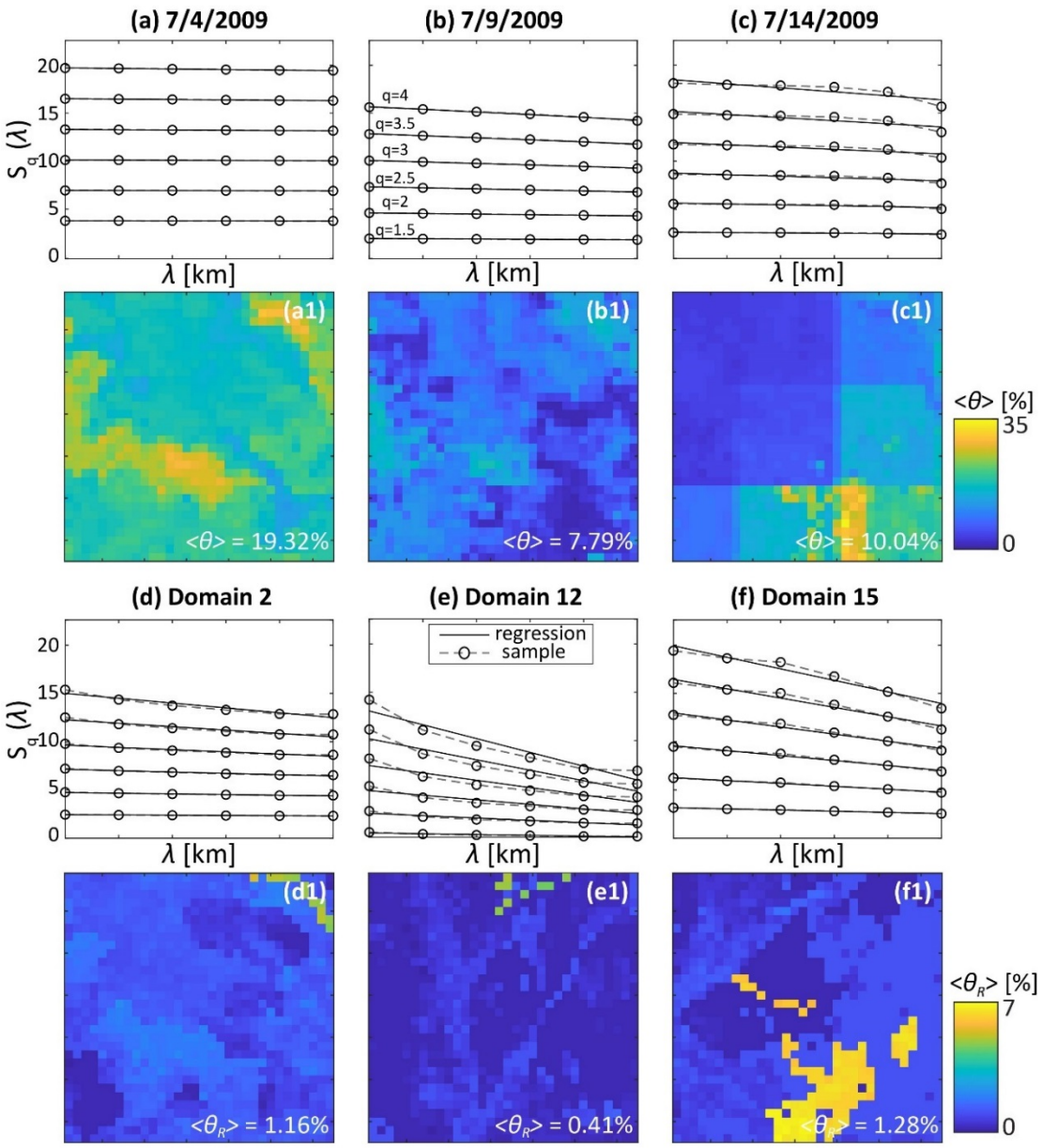

Figure 4.8. Examples of presence of scale invariance in (a) wet day, (b) dry day, and (c) day with the impact of discontinuity of P at 12 km in domains 2 and in residual soil moisture in (d) domain 2, (e) domain 12 and (f) domain 15. (a1-c1) Spatial maps of $<\theta>$. (d1-f1) Spatial maps of $<\theta_R>$.

Table 4.7. Mean across all scale-Invariant θ fields and θ_R of RMSE and CC computed from the scale invariance analysis when $q = 4$ for each domain.

	Number of scale-invariant θ fields	RMSE [%]	CC [-]	Number of Discarded θ fields	$<\theta>$of Discarded θ fields	$<\theta_R>$ [%]	RMSE [%]	CC [-]
Dom 1	3633	0.163	0.981	20	9.494	1.99	0.209	0.980
Dom 2	3622	0.117	0.980	31	8.742	1.16	0.264	0.956
Dom 3	3608	0.172	0.986	45	5.344	1.03	0.308	0.986
Dom 4	1738	0.095	0.981	1915	1.300	0.92	0.763	0.928
Dom 5	3527	0.175	0.945	126	3.197	0.95	0.632	0.854
Dom 6	1487	0.121	0.985	2166	1.233	0.97	0.720	0.949
Dom 7	1825	0.120	0.978	1828	1.220	0.92	0.822	0.927
Dom 8	3634	0.239	0.984	19	7.772	1.30	0.290	0.991
Dom 9	3604	0.218	0.985	49	3.867	0.56	0.343	0.992
Dom 10	3627	0.159	0.988	26	8.511	1.05	0.502	0.972
Dom 11	3630	0.203	0.987	23	6.317	1.18	0.372	0.983
Dom 12	3192	0.292	0.976	461	2.312	0.41	0.765	0.954
Dom 13	3384	0.236	0.971	269	3.242	0.63	0.609	0.929
Dom 14	1268	0.135	0.984	2385	1.383	1.00	0.470	0.979
Dom 15	1374	0.170	0.982	2279	1.727	1.28	0.443	0.977
Dom 16	3565	0.238	0.985	88	3.652	0.71	0.612	0.960

To investigate the impact of P at 12 km and the soil inputs, I studied the time series of RMSE and Figure 4.9 present this for the representative domains. It is evident that to see the impact of the discontinuity of the coarse resolution P from Figure 4.9(a) and (b) with the few spikes. The days having a constant value of RMSE for all domains are the driest days when their $<\theta>$ are close to $<\theta_R>$. A large number of days having RMSE higher than 0.5 was found in domain 15 and these were excluded for multifractal analysis which is designed based on the hypothesis of spatial homogeneity.

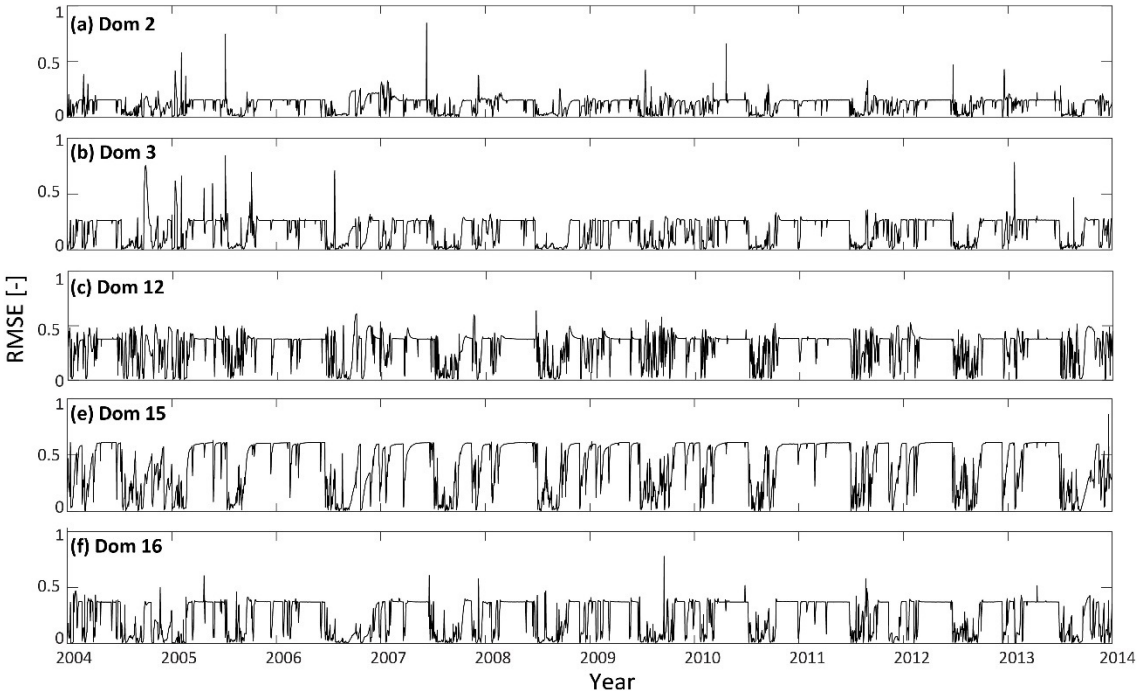

Figure 4.9. Time series of RMSE between samples and the linear regression model in the representative domains.

For each domain, c and β of the multifractal model were estimated by fitting the theoretical expectation for $K(q)$. Figure 4.10 shows results for selected wet and dry days in four pairs of two domains contrasting in terrain, vegetation, soil and annual P. In all cases, the relations between $K(q)$ and q are nonlinear, indicating the presence of multifractality. Figure 4.10 also shows the θ with lower $<\theta>$ has higher $K(q)$ for all the cases even in the different domains and conditions. The $K(q)$ is an indicator to measure the variation of the intermittency in a $<\theta>$ field within the aggregation scales. The higher $K(q)$ will be obtained from a more intermittent $<\theta>$ field since its variability will significantly decrease as the aggregation of the field processes. As a result, our findings reveal that θ values in dry conditions (or domains) are more intermittent while in wet conditions are more uniformly distributed, consistent with the result in Chapter 2 (Ko et al., 2016) and previous studies by Hu et al. (Hu et al., 1997), Oldak et al. (Oldak et al., 2002), Das and Mohanty (Das and Mohanty, 2008), and Mascaro and Vivoni (Mascaro and Vivoni, 2010).

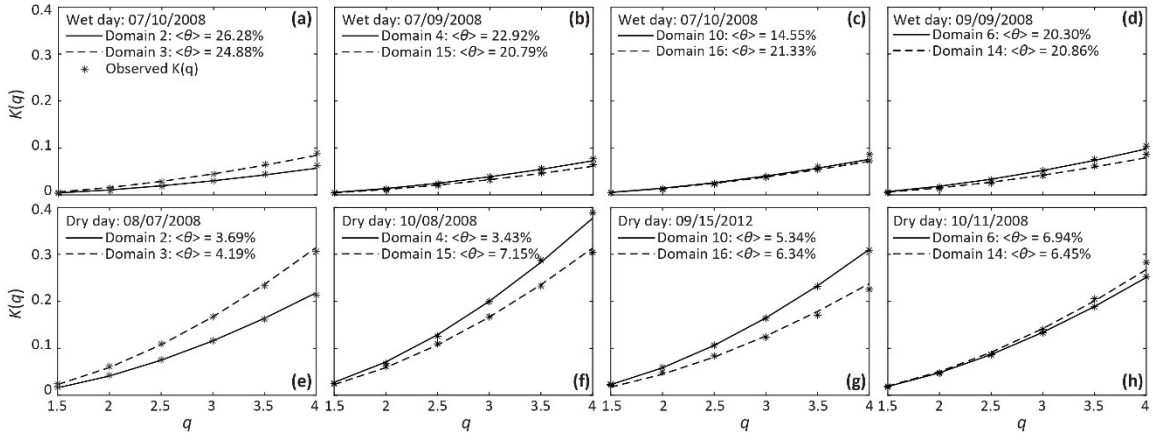

Figure 4.10. Relation between multifractal exponent $K(q)$ and moments q for four pairs consisting of two domains contrasting in (a, e: domains 2 and 3) terrain, (b, f: domain 4

and 15) vegetation, (c, g: domains 10 and 16) soil, and (d, h: domain 6 and 14) annual P on the same (a-d) wet and (e-h) dry days for each group.

The two parameters, c and β, were estimated in all domains. The values of β were found to be fairly constant for all domains with an average value of 0.82±0.15, whereas the values of c are highly scattered. As a next step, I assumed a fixed value of β equal to the average and then estimated c. The resulting values of c ranged from 0.04 to 6.58 with an exponentially decreasing relation with $<\theta>$, similar to the finding by Mascaro et al. (Mascaro et al., 2010, 2011). Mascaro et al. (2010) found a relatively smaller range of c with a smaller β for SMXE04, Sonora but it might be due to its narrow range of θ observations.

Two approaches were tested for calibrating the multifractal model. In the first approach labeled as the local (LOC), the sets of parameters, c_∞, α and γ, in (4.4) were estimated through the nonlinear fitting for each domain. The value of c_∞ that controls the intermittency of θ in larger $<\theta>$ was fixed to 0.2 due to the small variation of c_∞ over the domains then reestimated α and γ. The values of α and γ, the RSME, and R^2 are reported in Table 4.8. The second approach labeled as ancillary (ANC) was tested by applying the PCA to the same eight ancillary factors used for the analysis of the spatial variability of θ (Table 4.1) to interpret the relation between c and $<\theta>$. To minimize the number of parameters involved in the ANC approach, I tested the possible presence of a relation between α and γ for all domains and linear regression relation was found as $\gamma = 0.0611\alpha - 0.0341$ with RMSE of 0.087 and R^2 of 0.65. In this way, (4.4) depends only on α and this parameter can be linked to the PCS_j (j = 1,2,3) with a multilinear regression as:

$$\alpha = k_0 + k_1\mathrm{PCS}_1 + k_2\mathrm{PCS}_2 + k_3\mathrm{PCS}_3 \qquad (4.8)$$

with parameters k_0, k_1, k_2, and k_3 reported in Table 4.8 and the estimates of α and γ, RMSE and R^2 are summarized in Table 4.9. Results of the ANC and LOC approaches in the sixteen domains are plotted in Figure 4.11.

Table 4.8. Parameters k_0, k_1, k_2, and k_3

Parameter	Value
k_0	4.1920
k_1	0.0706
k_2	-0.0551
k_3	-0.1645

Table 4.9. Average of parameters between c and $<\theta>$ estimated from LOC and ANC approaches. RMSE and R^2 are reported.

	LOC Regression			**ANC Regression**			
	α	γ	RMSE	α	γ	RMSE	α_{RMSE} [%]
Dom 1	3.888	0.202	0.243	3.760	0.203	0.245	-0.8
Dom 2	1.636	0.093	0.425	3.320	0.066	1.064	-150.4
Dom 3	7.270	0.578	0.301	3.934	0.410	0.631	-109.6
Dom 4	3.199	0.238	0.492	3.758	0.161	0.623	-26.6
Dom 5	1.415	0.095	0.461	3.846	0.052	1.524	-230.6
Dom 6	4.433	0.190	0.364	4.377	0.237	0.456	-25.3
Dom 7	3.514	0.209	0.466	4.326	0.181	0.546	-17.2
Dom 8	6.765	0.289	0.333	3.744	0.379	1.320	-296.4
Dom 9	8.378	0.589	0.423	6.307	0.478	0.721	-70.4
Dom 10	5.321	0.295	0.285	5.271	0.291	0.285	0.0
Dom 11	6.061	0.269	0.285	4.249	0.336	0.891	-212.6
Dom 12	4.162	0.203	0.457	4.910	0.220	0.647	-41.6
Dom 13	2.817	0.127	0.429	3.834	0.138	0.734	-71.1
Dom 14	5.953	0.178	0.450	4.617	0.330	0.910	-102.2
Dom 15	5.816	0.160	0.364	3.711	0.321	1.111	-205.2
Dom 16	6.190	0.434	0.507	3.111	0.344	1.316	-159.6
Average	4.801	0.259	0.393	4.192	0.259	0.814	-107.5
STD	1.935	0.147	0.080	0.768	0.118	0.362	89.3
Max	8.378	0.589	0.507	6.307	0.478	1.524	0.0
Min	1.415	0.093	0.243	3.111	0.052	0.245	-296.4

The multifractal model performances for the ANC approach are, in general, less accurate than LOC but 63% of the domains (10 domains out of 16; domain 1, 4, 6, 7, 9-14) shows relatively similar performances to LOC with RMSE of 0.607. Figure 4.11(b) and (e) show the difference of LOC and ANC approaches in the variation of c in the dry conditions (~5% of $<\theta>$) in domain 2 and 5. These domains have a large range of variation of c in very dry conditions as well as majority of the pairs of c estimates and $<\theta>$ were clustered where $<\theta>$ = ~1.23% and ~1.0% for domain 2 and 5 respectively which are the extremely dry conditions that $<\theta>$ is close to $<\theta_R>$. This is obviously caused by the impact of the model input for the soil parameters and affected the empirical fits derived from LOC approach. ANC approach couldn't capture the drastic changes in the dry conditions (less intermittency up to ~10% of $<\theta>$) in domain 3 and 16 [see Figure 4.11(c) and (p)] by underestimating both α and γ, and it underestimates the intermittency of $<\theta>$ in domain 14 and 15 [see Figure 4.11(n) and (o)] by underestimate α and overestimate γ. It might be due to the simple linear relation of α and γ I used. The results were also compared with findings of Mascaro and Vivoni (2010), who applied the same analysis using SMEX04 datasets conducted in the same region. Figure 4.12 shows the comparison in the two domain 4 and 7 located within the coverage of SMEX04. The behaviors of the relations are relatively similar, but the smaller variation range of c was found in SMEX04. Possible reasons are; 1) a small number of θ fields used, 2) smaller range of $<\theta>$, and 3) the smaller value of the fixed β (0.71 for SMEX04 whereas 0.82 for the RSB) that can lead higher values of c.

To understand the geophysical control factors on the multifractal model, I also conducted the same analysis of comparing two domains contrasting in land surface properties as I previously did for spatial variability of θ. As shown in Figure 4.13(a), there is no significant impact of terrain properties on the multifractal properties even I found its impact on the spatial variability of θ [see Figure 4.7(a)]. Figure 4.13(b) shows the comparison between two domains with contrasting <VF>. Two domains have significantly different behaviors in the relation. Domain 3 with higher vegetation fraction tends to be more intermittent, implying different evapotranspiration rates play a major role on the heterogeneity in the field. Lastly, Figure 4.13(c) presents the comparison between two domains with contrasting $<K_s>$. Domain 12 with higher Ks shows less intermittency in the field, implying that faster draining is associated with less spatial heterogeneity of θ. Overall, the vegetation and soil properties significantly affect both spatial variability and multifractality of θ but soil properties have a larger impact on the spatial variability while vegetation properties have a significant impact on the multifractal properties of θ. The higher impact of terrain properties was found on the spatial variability, but it was minimal on multifractality.

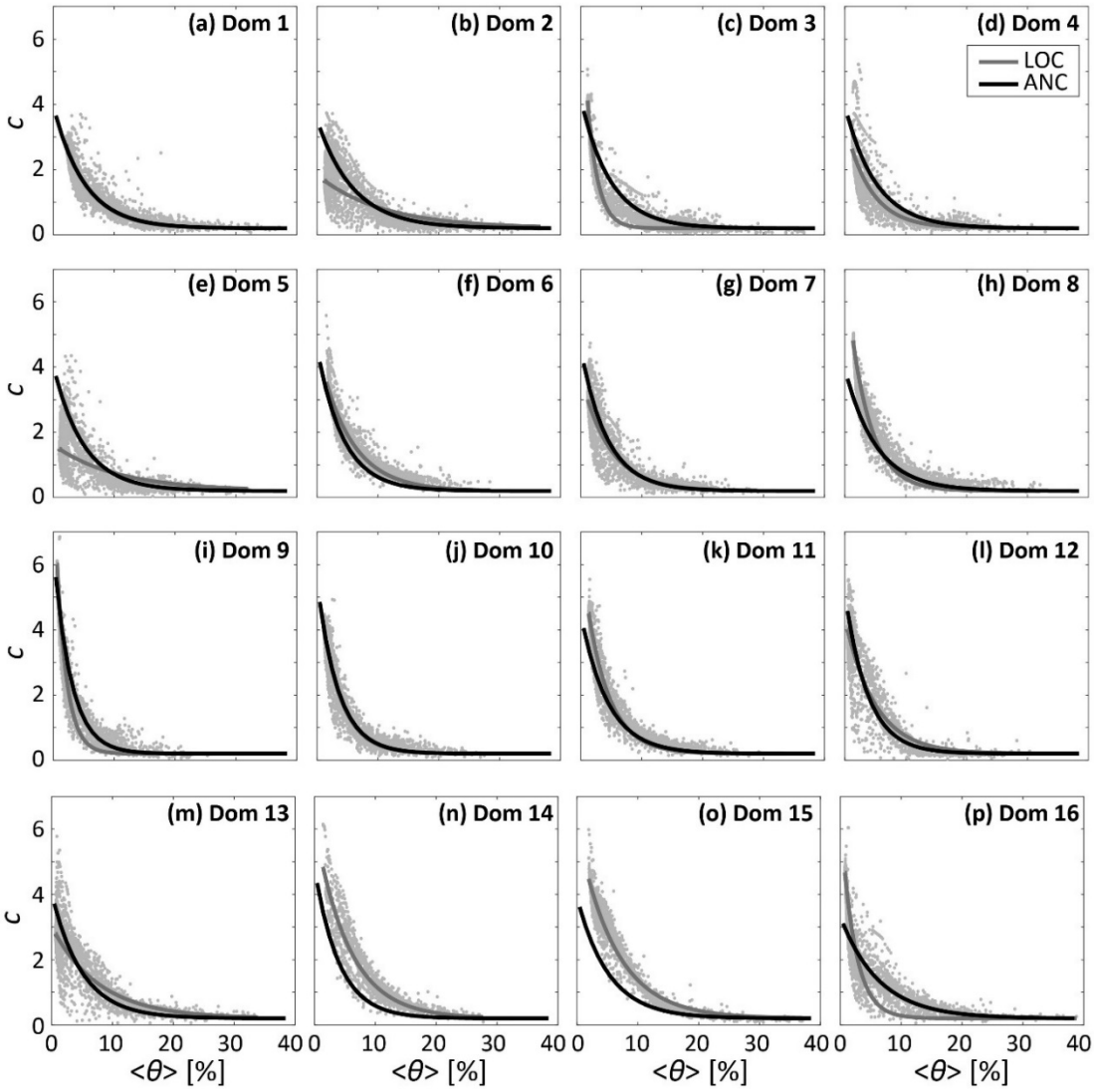

Figure 4.11. Relation between the *c* and <θ> [%] of each domain along with the empirical fit derived from LOC (gray lines) and ANC (black lines) regression.

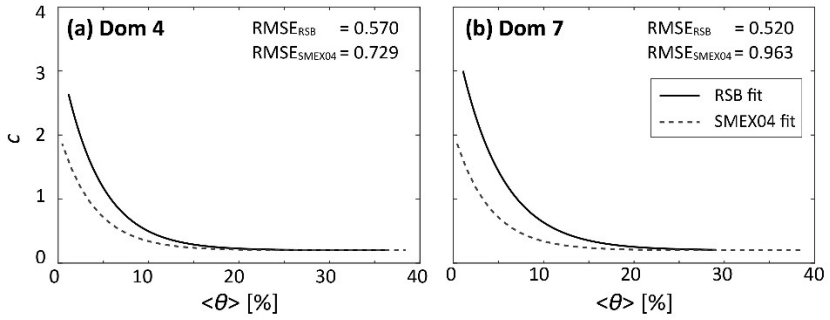

Figure 4.12. Relation between c vs. $<\theta>$ [%] in (a) domain 4 and (c) domain 7 which are located within the SMEX04 coverage. The empirical curve obtained with observed aircraft-derived datasets of SMEX04 in Sonora (Mascaro and Vivoni, 2010) are shown. RMSE between the daily points and each curve of RSB and SMEX04 are also presented in the panels.

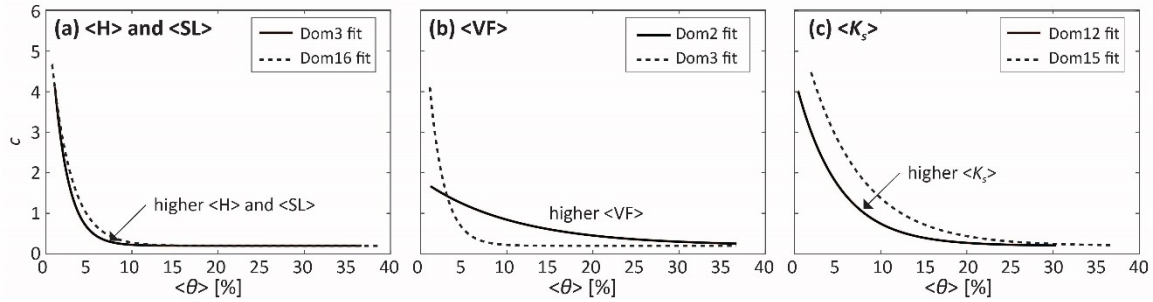

Figure 4.13. Relation between c vs. $<\theta>$. (a) domain 3 and 16 having a large contrast in terrain properties (<H> and <SL>), (b) domain 2 and 3 having a large contrast in vegetation property (<VF>), and (c) domain 12 and 15 having a large contrast in soil property ($<K_s>$).

Conclusions

Relatively few ground or remote sensing observation of soil moisture are available in the semiarid regions and only a few studies have used simulated θ to analyze the statistical and scaling properties of soil moisture. In this chapter, I used an extensive dataset of simulated θ in the RSB, characterized by complex terrain features with heterogeneous land surface to investigate; i) the statistical and scaling properties of θ fields, ii) if the simulated θ fields exhibit the similar statistical properties found by analyzing ground and remotely-sensed observations, and iii) how they vary under different soil, vegetation and terrain conditions. The results from this chapter reveal the following main conclusions;

1) The distributed tRBS simulates were able to capture the interannual and seasonal variability of θ and reproduce the effect of spatially-variable forcings and land surface properties.
2) Variability of θ (σ_θ) shows a convex-upward behavior with increasing $<\theta>$ and relative variability (CV_θ) decreases exponentially with higher moisture contents. The relations estimated on the simulated θ fields were very similar to the findings of previous studies, Famiglietti et al. (2008) using SMEX02 and 03 and Mascaro and Vivoni 2010. The behaviors of relations at the high $<\theta>$ can be explained by the differences in size of the datasets, range of $<\theta>$ values, and the extent scale of the analysis.

3) Scale invariance properties can be affected by either the coarse resolution of meteorological forcings potentially introducing the unnatural discontinuity on θ fields or soil parameter inputs that are initially not scale-invariant.
4) The relations linking θ intermittency preparties (quantified by c) and $<\theta>$ show that the θ spatial distribution in drier (wetter) domains tends to ve more intermittent (smoother), consistent with previous studies. They were also relatively similar to the result that Mascaro and Vivoni found in SMEX04.
5) In all the domains, the vegetation and soil properties significantly affect both spatial variability and multifractality of θ. The higher impact of terrain properties is found on the spatial variability, but it was minimal on multifractality.

Overall, this chapter suggests that the high-resolution θ fields can be used to quantify the impact of land surface properties and wetness conditions that have not been observed by ground and remote sensors on the statistical properties of θ and investigate the existence of scale invariance and multifractality across a wide range of scales. Future studies should be conducted to test the presence of scale invariance on the simulated fields at ~100 m, which is the highest spatial resolution of the model domain. The limitations of hyperresolution simulations that this chapter found can be addressed by: (i) applying sophisticated downscaling strategies in advance to increase the resolution of rainfall fields, in order to limit the impact of discontinuity of input data at coarse resolution and (ii) testing the technique proposed to generate high-resolution soil parameters in areas where there is sufficient information on soil texture.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

General Conclusions

Soil moisture (θ) is a fundamental variable controlling the exchange of water and energy at the land surface. As a result, the characterization and investigation of the spatial distribution of θ across multiple scales are essential for many applications including flood prediction, drought monitoring, and weather forecasting. Empirical evidences have demonstrated that θ fields exhibit scale invariance and multifractal properties across a range of scales. However, these evidences have been limited to a few locations and times using high-resolution (~1 km) θ images collected from airborne sensors during intensive field campaigns. This dissertation advances the characterization of the scaling invariance and multifractal properties of θ by (1) analyzing the role of irrigation in observed θ products collected during two field campaigns in Australia (Chapter 2); and (2) investigating the variation in time and across different landscape conditions of these properties using θ outputs of a distributed hydrologic model (Chapters 3 and 4).

Chapter 2 of my dissertation had the main goal of investigating how the presence of irrigated fields modifies the spatial distribution of θ and the associated scale invariance properties, which have been so far observed only in natural landscapes. I used the aircraft-based θ datasets from the NAFE05 and NAFE06 field campaigns carried out in two semiarid sites in Australia characterized by the presence of agricultural districts of different extent. A framework has been proposed to: 1) quantitatively analyze and compare the scale invariance and multifractal properties of θ in the presence of irrigation;

and 2) filter out the effect of irrigated croplands in the application of a multifractal downscaling algorithm based on the hypothesis of spatial homogeneity. Results showed that irrigation affects the scale-invariance properties in a large high-density agricultural district in the semi-arid NAFE06 site, although it does not have a significant impact on the sparser agricultural districts of the temperate NAFE05 region. I calibrated the multifractal downscaling model using the θ fields with attenuated effect of irrigation by removed the irrigated areas and interpolated from surrounding values from the natural landscape, thus mimicking natural settings. Performances tested against aircraft and, for the first time, ground-based θ observations, were adequate in most cases. Some deficiencies were found for drier conditions in regions with a higher percentage of irrigated fields, suggesting the need to further refine the techniques for detecting irrigated croplands. Overall, the findings of this work reveal that the impact of irrigation on the soil moisture statistical variability and downscaling is larger in drier regions or conditions, where irrigation creates a drastic contrast with the surrounding areas. The results of these analyses increase the utility of coarse θ satellite products by improvement of the calibration of multifractal downscaling models in agricultural regions. Furthermore, the framework proposed in Chapter 2 can be applied or easily adapted to investigate the effect of different sources of spatial heterogeneity, such as the presence of built areas and human activities.

An approach that allows advancing the characterization of scale invariance and multifractal properties of soil moisture to regions and times that have not monitored during intensive field campaigns, is through the use of distributed hydrologic models run

at very high spatial resolutions (hyperresolution). In Chapter 3, I aimed at generating reliable high-resolution θ field over a large basin using a physically-based hydrologic model by addressing important challenges associated with hyperresolution hydrologic simulations. At the same time, I evaluated and quantify the model ability to simulate spatial patterns of daily land surface temperature (LST) by comparing the outputs against the LST products at 1-km resolution from the Moderate Resolution Imaging Spectroradiometer (MODIS). I used the TIN-based Real-time Integrated Basin Simulator (tRIBS) DHM at a nominal resolution of ~88 m in a regional basin of ~21,000 km^2 in northwest Mexico. It was applied from 2004 to 2013 and model performances were extensively tested by comparing simulated outputs against independent observations, including time series of soil moisture recorded at a network of stations and spatial patterns of daily land surface temperature (LST) estimated by remote sensors. The model was able to simulate quite well the observed soil moisture response to precipitation and the subsequent recession phases during dry periods during the calibration periods as well as validation periods. The spatial variability of LST is captured well by the hyperresolution simulated patterns, and vegetation properties are the major physical factors explaining the discrepancies between simulated and remotely-sensed products. The strategies and results presented here are based on global datasets and robust statistical methods that can be utilized in different settings with other DHMs to diagnose model deficiencies and strategize improvement of their physics.

The hyperresolution simulations conducted in Chapter 3 generated a database of reliable soil moisture maps at high spatial and temporal resolution that span a large basin

of ~21000 km^2 and more than 10 years. In chapter 4, I investigated; (i) whether simulated θ data exhibits the scaling properties including scale invariance and multifractality found by analyzing ground and remotely-sensed observations, and (ii) how these properties vary under different soil, vegetation, and terrain conditions. The simulated θ fields capture the spatial patterns and scaling properties estimated from observed datasets very well. The statistical properties reproduced by simulated θ were found to be significantly controlled by soil and vegetation properties and a slight of terrain properties. These results suggest that the high-resolution θ fields can be used to quantify the impact of land surface properties and wetness conditions that have not been observed by ground and remote sensors on the statistical properties of θ and investigate the existence of scale invariance and multifractality across a wide range of scales.

Overall, the knowledge acquired through this dissertation will support the refinement of the calibration of multifractal downscaling models. Specifically, calibrations relations of the multifractal downscaling model proposed by Deidda et al. (2000) have been identified in the two regions of Australia analyzed in Chapter 2 and in the Rio Sonora basin, analyzed in Chapter 3 and 4. These relations can be used in the future in these study sites to simulate the small-scale spatial variability of soil moisture starting from coarse satellite estimates. Ultimately, this will support the decision makers and practitioners, including weather forecasters, farmers, and water managers who can apply the information to increase the accuracy of weather forecasts, the ability to monitor water availability and manage water resources, as well as plan irrigation schedules.

Future Work

There are many ways that my dissertation work can be expanded upon. In Chapter 2, I have shown that irrigation is a major factor influencing the spatial variability of θ fields and inclusion of irrigation in downscaling schemes is crucial in drier regions or conditions. The multifractal downscaling scheme can be refined by introducing techniques for postprocessing the disaggregated fields generated in a spatially homogeneous framework to include the spatial heterogeneity due to irrigated conditions. To do so, identification of irrigated croplands correctly is highly required. The framework proposed in Chapter 2 for detecting the irrigated croplands via remote sensed imagery has worked nicely in most cases, but it may have not been fully effective in some conditions. In future study, the framework should be refined to increase its accuracy by considering additional remote sensed imageries or indices. In addition, agricultural practices such as tillage and fallow system of farming can bring the spatial heterogeneity in soil moisture fields, thus necessitating inclusion of identification techniques to detect them into the framework.

Distributed hydrologic models applied at high resolution over large basins can generate high-resolution soil moisture data to improve the calibration relations for any type of downscaling method at a new location. In addition, in regions with the presence of irrigation, it is necessary to account for this source of spatial heterogeneity. Developing irrigation module for a hydrological model is highly required to improve its applicability for sites with presence of irrigation. The refined framework detecting the

potential factors bring the spatial heterogeneity via remote sensed imagery and attenuating their effect can be adapted into the module.

I have recognized that limitations of hydrological simulations for studying statistical and scaling properties of soil moisture. The use of meteorological forcings, especially precipitation, at coarse resolution can bring unnatural discontinuities in soil moisture fields and they would affect the scale invariance properties of soil moisture. To overcome, sophisticated downscaling strategies should be applied in advance to increase the resolution of rainfall fields, in order to limit the impact of discontinuity of input data at coarse resolution. Furthermore, additional efforts to test the technique proposed to generate high-resolution soil parameters are needed in areas where there is sufficient information on soil texture. This will allow better testing the utility of the global datasets from ISCRI-WISE.

The downscaling model was applied up to 31.25 m which is a scale representative of the ground measurements in Chapter 2, but it was not tested because gridded remotely sensed fields at that scale were not available. On the other hand, investigation of statistical and scaling properties of the simulated soil moisture has been carried out within the range from 32-km to 1-km. Future studies should be conducted to test the presence of scale invariance on the simulated fields at ~100 m, which is the highest spatial resolution of the model domain.

REFERENCES

Ahmadalipour, A., H. Moradkhani and A. Rana. 2018. Accounting for downscaling and model uncertainty in fine-resolution seasonal climate projections over the Columbia River Basin. *Climate Dynamics* 50(1–2): 717–733.

Albertson, J.D., and N. Montaldo. 2003. Temporal dynamics of soil moisture variability: 1. Theoretical basis. *Water Resources Research* 39(10).

Baker, I.T., P.J. Sellers, A.S. Denning, I. Medina, P. Kraus, K.D. Haynes and S.C. Biraud. 2017. Closing the scale gap between land surface parameterizations and GCMs with a new scheme, SiB3-Bins. *Journal of Advances in Modeling Earth Systems* 9(1): 691–711.

Barnes, S.L. 1994. Applications of the Barnes objective analysis scheme. Part I: Effects of undersampling, wave position, and station randomness. *Journal of Atmospheric and Oceanic Technology* 11(6): 1433–1448.

Bertoldi, G., S. Della Chiesa, C. Notarnicola, L. Pasolli, G. Niedrist and U. Tappeiner. 2014. Estimation of soil moisture patterns in mountain grasslands by means of SAR RADARSAT2 images andhydrological modeling. *Journal of Hydrology* 516: 245–257. http://dx.doi.org/10.1016/j.jhydrol.2014.02.018.

Beven, K.J., and H.L. Cloke. 2012. Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al. *Water Resources Research* 48(1).

Bierkens, M.F.P., V.A. Bell, P. Burek, N.W. Chaney, L.E. Condon, C.H. David, A. de Roo, P. Döll, N. Drost and J.S. Famiglietti. 2015. Hyper-resolution global hydrological modelling: What is next? Everywhere and locally relevant. *Hydrological Processes* 29(2): 310–320.

Bindlish, R., T.J. Jackson, A. Gasiewski, B. Stankov, M. Klein, M.H. Cosh, I. Mladenova, et al. 2008. Aircraft based soil moisture retrievals under mixed vegetation and topographic conditions. *Remote Sensing of Environment* 112(2): 375–390.

Brito-Castillo, L., A. V Douglas, A. Leyva-Contreras and D. Lluch-Belda. 2003. The effect of large-scale circulation on precipitation and streamflow in the Gulf of California continental watershed. *International Journal of Climatology* 23(7): 751–768.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis and S. Hasenauer. 2010. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences* 14(10): 1881–1893.

Brocca, L., R. Morbidelli, F. Melone and T. Moramarco. 2007. Soil moisture spatial variability in experimental areas of central Italy. *Journal of Hydrology* 333(2–4): 356–373.

Brocca, L., T. Tullo, F. Melone, T. Moramarco and R. Morbidelli. 2012. Catchment scale soil moisture spatial-temporal variability. *Journal of Hydrology* 422–423: 63–75. http://dx.doi.org/10.1016/j.jhydrol.2011.12.039.

Cabral, M.C., L. Garrote, R.L. Bras and D. Entekhabi. 1992. A kinematic model of infiltration and runoff generation in layered and sloped soils. *Advances in Water Resources* 15(5): 311–324.

Cai, X., M. Pan, N.W. Chaney, A. Colliander, S. Misra, M.H. Cosh, W.T. Crow, T.J. Jackson and E.F. Wood. 2017. Validation of SMAP soil moisture for the SMAPVEX15 field campaign using a hyper-resolution model. *Water Resources Research* 53(4): 3013–3028.

Calvet, J.C., J.P. Wigneron, J. Walker, F. Karbou, A. Chanzy and C. Albergel. 2011. Sensitivity of Passive Microwave Observations to Soil Moisture and Vegetation Water Content: L-Band to W-Band. *IEEE Transactions on Geoscience and Remote Sensing* 49(4): 1190–1199.

Camporese, M., C. Paniconi, M. Putti and S. Orlandini. 2010. Surface-subsurface fCamporese, M., C.Fatichi, S., ECamporese, M., C. Paniconi, M. Putti and S. Orlandini. 2010. Surface-subsurface fCamporese, M., C.Fatichi, S., E.R. Vivoni, F.L. Ogden, V.Y. Ivanov, B. Mirus, D. Gochis, C.W. Downer, M. Camporese, J.H. Da. *Water Resources Research* 46(2).

Charpentier, M.A., and P.M. Groffman. 1992. Soil moisture variability within remote sensing pixels. *Journal of Geophysical Research* 97(D17): 18987.

Choi, M., and J.M. Jacobs. 2007. Soil moisture variability of root zone profiles within SMEX02 remote sensing footprints. *Advances in Water Resources* 30(4): 883–896.

Cosh, M.H., and W. Brutsaert. 1999. Aspects of soil moisture variability in the Washita ’92 study region. *Journal of Geophysical Research* 104(D16): 749.

Crow, W.T., and E.F. Wood. 1999. Multi-scale dynamics of soil moisture variability observed during SGP’97. *Geophysical Research Letters* 26(23): 3485–3488.

Crow, W.T., and E.F. Wood. 2002. The value of coarse-scale soil moisture observations for regional surface energy balance modeling. *Journal of Hydrometeorology* 3(4): 467–482.

Cui, C., J. Xu, J. Zeng, K.-S. Chen, X. Bai, H. Lu, Q. Chen, et al. 2017. Soil moisture mapping from satellites: An intercomparison of SMAP, SMOS, FY3B, AMSR2, and

ESA CCI over two dense network regions at different spatial scales. *Remote Sensing* 10(2): 33.

Das, N.N., and B.P. Mohanty. 2008. Temporal dynamics of PSR-based soil moisture across spatial scales in an agricultural landscape during SMEX02: A wavelet approach. *Remote Sensing of Environment* 112(2): 522–534.

Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo and P. Bauer. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137(656): 553–597.

Deidda, R. 2000. Rainfall downscaling in a space-time multifractal framework. *Water Resources Research* 36(7): 1779–1794.

Deidda, R., R. Benzi and F. Siccardi. 1999. Multifractal modeling of anomalous scaling laws in rainfall. *Water Resources Research* 35(6): 1853–1867.

dell' Arciprete, D., R. Bersezio, F. Felletti, M. Giudici, A. Comunian and P. Renard. 2012. Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. *Hydrogeology Journal* 20(2): 299–311.

Do, H.X., L. Gudmundsson, M. Leonard and S. Westra. 2018. The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data* 10(2): 765.

Entekhabi, D. 1995. Recent advances in land-atmosphere interaction research. *Reviews of Geophysics* 33(S2): 995–1003.

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 and J. Johnson. 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE* 98(5): 704–716.

Famiglietti, J.S., J.A. Devereaux, C.A. Laymon, T. Tsegaye, P.R. Houser, T.J. Jackson, S.T. Graham, M. Rodell and P.J. Van Oevelen. 1999. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resources Research* 35(6): 1839–1851.

Famiglietti, J.S., D. Ryu, A.A. Berg, M. Rodell and T.J. Jackson. 2008. Field observations of soil moisture variability across scales. *Water Resources Research* 44(1): 1–16.

Fang, Z., H. Bogena, S. Kollet, J. Koch and H. Vereecken. 2015. Spatio-temporal validation of long-term 3D hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis. *Journal of*

Hydrology 529: 1754–1767.

Fatichi, S., E.R. Vivoni, F.L. Ogden, V.Y. Ivanov, B. Mirus, D.J. Gochis, C.W. Downer, M. Camporese, J.H. Davison and B. Ebel. 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology* 537: 45–60.

Fensholt, R., I. Sandholt and M.S. Rasmussen. 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environment* 91(3–4): 490–507.

Forzieri, G., F. Castelli and E.R. Vivoni. 2011. Vegetation Dynamics within the North American Monsoon Region. *Journal of Climate* 24(6): 1763–1783. http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3847.1.

Franz, T.E., T.D. Loecke, A.J. Burgin, Y. Zhou, T. Le and D. Moscicki. 2017. Spatio-temporal predictions of soil properties and states in variably saturated landscapes. *Journal of Geophysical Research: Biogeosciences*.

Gautam, J., and G. Mascaro. 2018. Evaluation of Coupled Model Intercomparison Project Phase 5 historical simulations in the Colorado River basin. *International Journal of Climatology*: 1–17.

Gebremichael, M., R. Rigon, G. Bertoldi and T.M. Over. 2009. On the scaling characteristics of observed and simulated spatial soil moisture fields. *Nonlinear Processes in Geophysics* 16(1): 141–150.

Gebremichael, M., E.R. Vivoni, C.J. Watts and J.C. Rodríguez. 2007. Submesoscale spatiotemporal variability of North American monsoon rainfall over complex terrain. *Journal of Climate* 20(9): 1751–1773.

Gibson, J., T.E. Franz, T. Wang, J. Gates, P. Grassini, H. Yang and D. Eisenhauer. 2017. A case study of field-scale maize irrigation patterns in western Nebraska: implications for water managers and recommendations for hyper-resolution land surface modeling. *Hydrology and Earth System Sciences* 21(2): 1051.

Gochis, D.J., L. Brito-Castillo and W.J. Shuttleworth. 2006. Hydroclimatology of the North American Monsoon region in northwest Mexico. *Journal of Hydrology* 316(1–4): 53–70.

Grayson, R.B., G. Blöschl, A.W. Western and T.A. McMahon. 2002. Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources* 25(8–12): 1313–1334.

Grayson, R.B., A.W. Western, F.H.S. Chiew and G. Blöschl. 1997. Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resources*

Research 33(12): 2897–2908.

Hannachi, A., I.T. Jolliffe and D.B. Stephenson. 2007. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology* 27(9): 1119–1152.

Harris, I., P.D. Jones, T.J. Osborn and D.H. Lister. 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology* 34(3): 623–642.

Heathman, G.C., M.H. Cosh, E. Han, T.J. Jackson, L. McKee and S. McAfee. 2012. Field scale spatiotemporal analysis of surface soil moisture for evaluating point-scale in situ networks. *Geoderma* 170: 195–205.

Hengl, T., J.M. de Jesus, G.B.M. Heuvelink, M.R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M.N. Wright, X. Geng and B. Bauer-Marschallinger. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* 12(2): e0169748.

Hengl, T., J.M. de Jesus, R.A. MacMillan, N.H. Batjes, G.B.M. Heuvelink, E. Ribeiro, A. Samuel-Rosa, B. Kempen, J.G.B. Leenaars and M.G. Walsh. 2014. SoilGrids1km—global soil information based on automated mapping. *PLoS ONE* 9(8): e105992.

Hu, Z., Y. Chen and S. Islam. 1998. Multiscaling properties of soil moisture images and decomposition of large-and small-scale features using wavelet transforms. *International Journal of Remote Sensing* 19(13): 2451–2467.

Hu, Z.L., S. Islam and Y.Z. Cheng. 1997. Statistical characterization of remotely sensed soil moisture images. *Remote Sensing of Environment* 61(2): 310–318.

Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25(3): 295–309.

Ivanov, V.Y., E.R. Vivoni, R.L. Bras and D. Entekhabi. 2004a. Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resources Research* 40(11).

Ivanov, V.Y., E.R. Vivoni, R.L. Bras and D. Entekhabi. 2004b. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. *Journal of Hydrology* 298(1–4): 80–111.

Johnson, R.H., P.E. Ciesielski, B.D. McNoldy, P.J. Rogers and R.K. Taft. 2007. Multiscale variability of the flow during the North American Monsoon Experiment. *Journal of Climate* 20(9): 1628–1648.

Kaheil, Y.H., M.K. Gill, M. McKee, L.A. Bastidas and E. Rosero. 2008. Downscaling

and assimilation of surface soil moisture using ground truth measurements. *IEEE Transactions on Geoscience and Remote Sensing* 46(5): 1375–1384.

Kaplan, S., and S.W. Myint. 2012. Estimating irrigated agricultural water use through landsat TM and a simplified surface energy balance modeling in the semi-arid environments of Arizona. *Photogrammetric Engineering and Remote Sensing* 78(8): 849–859.

Karypis, G., and V. Kumar. 1998. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing* 20(1): 359–392.

Kerr, Y.H., P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font and M. Berger. 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.

Kim, G., and A.P. Barros. 2002a. Space–time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data. *Remote Sensing of Environment* 81(2): 393–403.

Kim, G., and A.P. Barros. 2002b. Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sensing of Environment* 83(3): 400–413.

Ko, A., G. Mascaro and E.R. Vivoni. 2016. Irrigation Impacts on Scaling Properties of Soil Moisture and the Calibration of a Multifractal Downscaling Model. *IEEE Transactions on Geoscience and Remote Sensing* 54(6): 3128–3142.

Koch, J., T. Cornelissen, Z. Fang and H. Bogena. 2016a. Inter-comparison of three distributed hydrological models with respect to seasonal variability of soil moisture patterns at a small forested catchment. *Journal of Hydrology* 533: 234–249.

Koch, J., K.H. Jensen and S. Stisen. 2015. Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. *Water Resources Research* 51(2): 1225–1246.

Koch, J., G. Mendiguren, G. Mariethoz and S. Stisen. 2017. Spatial sensitivity analysis of simulated land surface patterns in a catchment model using a set of innovative spatial performance metrics. *Journal of Hydrometeorology* 18(4): 1121–1142. http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0148.1.

Koch, J., A. Siemann, S. Stisen and J. Sheffield. 2016b. Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. *Journal of Geophysical Research: Atmospheres* 121(10): 5430–5452.

Koch, J., and S. Stisen. 2017. Citizen science: A new perspective to advance spatial pattern evaluation in hydrology. *PloS One* 12(5): e0178165.

Koch, S.E., M. DesJardins and P.J. Kocin. 1983. An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Climate and Applied Meteorology* 22(9): 1487–1503.

Kollet, S.J., and R.M. Maxwell. 2006. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources* 29(7): 945–958.

Kollet, S.J., R.M. Maxwell, C.S. Woodward, S. Smith, J. Vanderborght, H. Vereecken and C. Simmer. 2010. Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research* 46(4).

Kumar, P. 1999. A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture. *IEEE Transactions on Geoscience and Remote Sensing* 37(1): 182–197.

Lawrence, J.E., and G.M. Hornberger. 2007. Soil moisture variability across climate zones. *Geophysical Research Letters* 34(20).

Lehner, B., K. Verdin and A. Jarvis. 2008. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union* 89(10): 93–94.

Lin, D.-S., E.F. Wood, P.A. Troch, M. Mancini and T.J. Jackson. 1994. Comparisons of remotely sensed and model-simulated soil moisture over a heterogeneous watershed. *Remote Sensing of Environment* 48(2): 159–171.

Liston, G.E., and K. Elder. 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *Journal of Hydrometeorology* 7(2): 217–234.

Lymburner, L., P. Tan, N. Mueller, R. Thackway, A. Lewis, M. Thankappan, L. Randall, A. Islam and U. Senarath. 2011. The National Dynamic Land Cover Dataset. *Geoscience Australia, Symonston, Australia* 10.

Mahmood, T.H., and E.R. Vivoni. 2008. Evaluation of distributed soil moisture simulations through field observations during the North American monsoon in Redondo Creek, New Mexico. *Ecohydrology* 1(3): 271–287.

Mahmood, T.H., and E.R. Vivoni. 2011. A climate-induced threshold in hydrologic response in a semiarid ponderosa pine hillslope. *Water Resources Research* 47(9).

Maidment, D.R. 2016. Open water data in space and time. *JAWRA Journal of the American Water Resources Association* 52(4): 816–824.

Maidment, D.R. 2017. Conceptual framework for the national flood interoperability experiment. *JAWRA Journal of the American Water Resources Association* 53(2): 245–257.

Manfreda, S., M.F. McCabe, M. Fiorentino, I. Rodríguez-Iturbe and E.F. Wood. 2007. Scaling characteristics of spatial patterns of soil moisture from distributed modelling. *Advances in Water Resources* 30(10): 2145–2150.

Martínez García, G., Y.A. Pachepsky and H. Vereecken. 2014. Effect of soil hydraulic properties on the relationship between the spatial mean and variability of soil moisture. *Journal of Hydrology* 516: 154–160.

Mascaro, G., R. Deidda and E.R. Vivoni. 2008. A New Verification Method to Ensure Consistent Ensemble Forecasts through Calibrated Precipitation Downscaling Models. *Monthly Weather Review* 136(9): 3374–3391. http://dx.doi.org/10.1175/2008MWR2339.1.

Mascaro, G., and E.R. Vivoni. 2010. Statistical and scaling properties of remotely-sensed soil moisture in two contrasting domains in the North American monsoon region. *Journal of Arid Environments* 74(5): 572–578.

Mascaro, G., and E.R. Vivoni. 2012. Utility of coarse and downscaled soil moisture products at L-band for hydrologic modeling at the catchment scale. *Geophysical Research Letters* 39(10).

Mascaro, G., E.R. Vivoni and R. Deidda. 2010. Downscaling soil moisture in the southern Great Plains through a calibrated multifractal model for land surface modeling applications. *Water Resources Research* 46(8).

Mascaro, G., E.R. Vivoni and R. Deidda. 2011. Soil moisture downscaling across climate regions and its emergent properties. *Journal of Geophysical Research: Atmospheres (1984–2012)* 116(D22).

Mascaro, G., E.R. Vivoni, D.J. Gochis, C.J. Watts and J.C. Rodríguez. 2014. Temporal downscaling and statistical analysis of rainfall across a topographic transect in Northwest Mexico. *Journal of Applied Meteorology and Climatology* 53(4): 910–927.

Mascaro, G., E.R. Vivoni and L.A. Méndez-Barroso. 2015. Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions. *Advances in Water Resources* 83: 190–206.

McKenzie, N.J., D.W. Jacquier, L.J. Ashton and H.P. Cresswell. 2000. Estimation of soil

properties using the Atlas of Australian Soils. *CSIRO Land and Water Technical Report* 11(00): 1–12.

Melsen, L.A., A.J. Teuling, P.J.J.F. Torfs, R. Uijlenhoet, N. Mizukami and M.P. Clark. 2016. HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences* 20(3): 1069–1079.

Méndez-Barroso, L.A., and E.R. Vivoni. 2010. Observed shifts in land surface conditions during the North American Monsoon: Implications for a vegetation–rainfall feedback mechanism. *Journal of Arid Environments* 74(5): 549–555.

Méndez-Barroso, L.A., E.R. Vivoni and G. Mascaro. 2016. Impact of spatially-variable soil thickness and texture on simulated hydrologic conditions in a semiarid watershed in northwest Mexico. *Revista Mexicana de Ciencias Geológicas* 33(3).

Méndez-Barroso, L.A., E.R. Vivoni, A. Robles-Morua, G. Mascaro, E.A. Yépez, J.C. Rodríguez, C.J. Watts, J. Garatuza-Payan and J.A. Saíz-Hernández. 2014. A modeling approach reveals differences in evapotranspiration and its partitioning in two semiarid ecosystems in Northwest Mexico. *Water Resources Research* 50(4): 3229–3252.

Méndez-Barroso, L.A., E.R. Vivoni, C.J. Watts and J.C. Rodríguez. 2009. Seasonal and interannual relations between precipitation, surface soil moisture and vegetation dynamics in the North American monsoon region. *Journal of Hydrology* 377(1–2): 59–70.

Merlin, O., A. Chehbouni, G. Boulet and Y.H. Kerr. 2006a. Assimilation of Disaggregated Microwave Soil Moisture into a Hydrologic Model Using Coarse-Scale Meteorological Data. *Journal of Hydrometeorology* 7(6): 1308–1322. http://dx.doi.org/10.1175/JHM552.1.

Merlin, O., A. Chehbouni, Y.H. Kerr and D.C. Goodrich. 2006b. A downscaling method for distributing surface soil moisture within a microwave pixel: Application to the Monsoon'90 data. *Remote Sensing of Environment* 101(3): 379–389.

Merlin, O., A.G. Chehbouni, Y.H. Kerr, E.G. Njoku and D. Entekhabi. 2005. A combined modeling and multipectral/multiresolution remote sensing approach for disaggregation of surface soil moisture: Application to SMOS configuration. *IEEE Transactions on Geoscience and Remote Sensing* 43(9): 2036–2050.

Merlin, O., J.P. Walker, A. Chehbouni and Y.H. Kerr. 2008a. Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment* 112(10): 3935–3946.

Merlin, O., J.P. Walker, J.D. Kalma, E.J. Kim, J.M. Hacker, R. Panciera, R. Young, G. Summerell, J. Hornbuckle and M. Hafeez. 2008b. The NAFE'06 data set: Towards

soil moisture retrieval at intermediate resolution. *Advances in Water Resources* 31(11): 1444–1455.

Merlin, O., J.P. Walker, R. Panciera, M.J. Escorihuela and T.J. Jackson. 2009. Assessing the SMOS soil moisture retrieval parameters with high-resolution NAFE'06 data. *Geoscience and Remote Sensing Letters, IEEE* 6(4): 635–639.

Mitchell, K.E., D. Lohmann, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, B.A. Cosgrove, J. Sheffield, Q. Duan and L. Luo. 2004. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres* 109(D7).

Mohanty, B.P., J.S. Famiglietti and T.H. Skaggs. 2000. Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resources Research* 36(12): 3675–3686.

Monteith, J.L. 1965. Evaporation and environment. In *Symposia of the Society for Experimental Biology*, 19:205–234.

Njoku, E.G., T.J. Jackson, V. Lakshmi, T.K. Chan and S. V Nghiem. 2003. Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing* 41(2): 215–229.

Oldak, A., Y. Pachepsky, T.J. Jackson and W.J. Rawls. 2002. Statistical properties of soil moisture images revisited. *Journal of Hydrology* 255(1): 12–24.

Orth, R., M. Staudinger, S.I. Seneviratne, J. Seibert and M. Zappa. 2015. Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology* 523: 147–159.

Panciera, R. 2009. *Effect of land surface heterogeneity on satellite near-surface soil moisture observations*. University of Melbourne, Department of Civil and Environmental Engineering.

Panciera, R., J.P. Walker, J.D. Kalma, E.J. Kim, J.M. Hacker, O. Merlin, M. Berger and N. Skou. 2008. The NAFE'05/CoSMOS data set: Toward SMOS soil moisture retrieval, downscaling, and assimilation. *IEEE Transactions on Geoscience and Remote Sensing* 46(3): 736–745.

Panciera, R., J.P. Walker, J.D. Kalma, E.J. Kim, K. Saleh and J.-P. Wigneron. 2009. Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm. *Remote Sensing of Environment* 113(2): 435–444.

Paniconi, C., and M. Putti. 2015. Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research* 51(9): 7090–7129.

Peel, M.C., B.L. Finlayson and T.A. McMahon. 2007. Updated World Map of the Köppen-Geiger Climate Classification. *Hydrol. Earth Systm. Sci. Discuss*. 4: 439–473.

Pellenq, J., J.D. Kalma, G. Boulet, G.-M. Saulnier, S. Wooldridge, Y.H. Kerr and A. Chehbouni. 2003. A disaggregation scheme for soil moisture based on topography and soil depth. *Journal of Hydrology* 276(1–4): 112–127.

Penmen, H.L. 1948. Natural evaporation from open water, hare soil and grass. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 193(1032): 120–145.

Perry, M.A., and J.D. Niemann. 2008. Generation of soil moisture patterns at the catchment scale by EOF interpolation. *Hydrology and Earth System Sciences Discussions* 12(1): 39–53.

Peters-Lidard, C.D., F. Pan and E.F. Wood. 2001. A re-examination of modeled and measured soil moisture spatial variability and its implications for land surface modeling. *Advances in Water Resources* 24(9–10): 1069–1083.

Pierini, N.A., E.R. Vivoni, A. Robles-Morua, R.L. Scott and M.A. Nearing. 2014. Using observations and a distributed hydrologic model to explore runoff thresholds linked with mesquite encroachment in the Sonoran Desert. *Water Resources Research* 50(10): 8191–8215.

Quinn, J.W. 2001. Band Combinations. available: http://Aweb.pdx.edu/~emch/ip1/bandcombinations.html.

Rebel, K.T., R.A.M. De Jeu, P. Ciais, N. Viovy, S.L. Piao, G. Kiely and A.J. Dolman. 2012. A global analysis of soil moisture derived from satellite observations and a land surface model. *Hydrology and Earth System Sciences* 16: 833–847.

Reichle, R.H., R.D. Koster, P. Liu, S.P.P. Mahanama, E.G. Njoku and M. Owe. 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research: Atmospheres* 112(D9).

Renard, P., and D. Allard. 2013. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources* 51: 168–196.

Richards, J.A., and J.A. Richards. 1999. *Remote sensing digital image analysis*. Vol. 3. Springer.

Robles-Morua, A., D. Che, A.S. Mayer and E.R. Vivoni. 2015. Hydrological assessment of proposed reservoirs in the Sonora River Basin, Mexico, under historical and

future climate scenarios. *Hydrological Sciences Journal* 60(1): 50–66. http://www.tandfonline.com/doi/abs/10.1080/02626667.2013.878462.

Robles-Morua, A., E.R. Vivoni and A.S. Mayer. 2012. Distributed Hydrologic Modeling in Northwest Mexico Reveals the Links between Runoff Mechanisms and Evapotranspiration. *Journal of Hydrometeorology* 13(3): 785–807. http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-11-0112.1.

Rodriguez-Iturbe, I., G.K. Vogel, R. Rigon, D. Entekhabi, F. Castelli and A. Rinaldo. 1995. On the spatial organization of soil moisture fields. *Geophysical Research Letters* 22(20): 2757–2760.

Rowe, A.K., S.A. Rutledge, T.J. Lang, P.E. Ciesielski and S.M. Saleeby. 2008. Elevation-dependent trends in precipitation observed during NAME. *Monthly Weather Review* 136(12): 4962–4979.

Ryu, D., T.J. Jackson, R. Bindlish, D.M. Le Vi and M. Haken. 2010. Soil Moisture Retrieval Using a Two-Dimensional L-Band Synthetic Aperture Radiometer in a Semiarid Environment. *IEEE Transactions on Geoscience and Remote Sensing* 48(12): 4273–4284.

Ryu, Y., J. Verfaillie, C. Macfarlane, H. Kobayashi, O. Sonnentag, R. Vargas, S. Ma and D.D. Baldocchi. 2012. Continuous observation of tree leaf area index at ecosystem scale using upward-pointing digital cameras. *Remote Sensing of Environment* 126: 116–125.

Saulnier, G.-M., K. Beven and C. Obled. 1997. Including spatially variable effective soil depths in TOPMODEL. *Journal of Hydrology* 202(1–4): 158–172.

Schaap, M.G., F.J. Leij and M.T. Van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* 251(3–4): 163–176.

Senatore, A., G. Mendicino, D.J. Gochis, W. Yu, D.N. Yates and H. Kunstmann. 2015. Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *Journal of Advances in Modeling Earth Systems* 7(4): 1693–1715.

Shen, X., and E.N. Anagnostou. 2017. A framework to improve hyper-resolution hydrological simulation in snow-affected regions. *Journal of Hydrology* 552: 1–12.

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

Siebert, S., M. Kummu, M. Porkka, P. Döll, N. Ramankutty and B.R. Scanlon. 2015. A

global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences* 19(3): 1521–1545.

Singh, R.S., J.T. Reager, N.L. Miller and J.S. Famiglietti. 2015. Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4. 0 simulations over the Southwestern US. *Water Resources Research* 51(4): 2648–2667.

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki and D. Stooksbury. 2002. The drought monitor. *Bulletin of the American Meteorological Society* 83(8): 1181–1190.

Tao, F., M. Yokozawa, Y. Hayashi and E. Lin. 2003. Changes in agricultural water demands and soil moisture in China over the last half-century and their effects on agricultural production. *Agricultural and Forest Meteorology* 118(3–4): 251–261.

Taylor, K.E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106(D7): 7183–7192.

Teuling, A.J., and P.A. Troch. 2005. Improved understanding of soil moisture variability dynamics. *Geophysical Research Letters* 32(5): L05404.

Vereecken, H., J.A. Huisman, Y. Pachepsky, C. Montzka, J. van der Kruk, H. Bogena, L. Weihermu¨ller, M. Herbst, G. Martinez and J. Vanderborght. 2014. On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology* 516: 76–96.

Vereecken, H., T. Kamai, T. Harter, R. Kasteel, J. Hopmans and J. Vanderborght. 2007. Explaining soil moisture variability as a function of mean soil moisture: A stochastic unsaturated flow perspective. *Geophysical Research Letters* 34(22): L22402.

Vivoni, E.R., D. Entekhabi, R.L. Bras and V.Y. Ivanov. 2007a. Controls on runoff generation and scale-dependence in a distributed hydrologic model. *Hydrology and Earth System Sciences Discussions* 11(5): 1683–1701.

Vivoni, E.R., M. Gebremichael, C.J. Watts, R. Bindlish and T.J. Jackson. 2008a. Comparison of ground-based and remotely-sensed surface soil moisture estimates over complex terrain during SMEX04. *Remote Sensing of Environment* 112(2): 314–325.

Vivoni, E.R., H.A. Gutiérrez-Jurado, C.A. Aragón, L.A. Méndez-Barroso, A.J. Rinehart, R.L. Wyckoff, J.C. Rodríguez, C.J. Watts, J.D. Bolten and V. Lakshmi. 2007b. Variation of hydrometeorological conditions along a topographic transect in northwestern Mexico during the North American monsoon. *Journal of Climate* 20(9): 1792–1809.

Vivoni, E.R., V.Y. Ivanov, R.L. Bras and D. Entekhabi. 2004. Generation of triangulated irregular networks based on hydrological similarity. *Journal of Hydrologic Engineering* 9(4): 288–302.

Vivoni, E.R., V.Y. Ivanov, R.L. Bras and D. Entekhabi. 2005. On the effects of triangulated terrain resolution on distributed hydrologic model response. *Hydrological Processes* 19(11): 2101–2122.

Vivoni, E.R., G. Mascaro, S. Mniszewski, P. Fasel, E.P. Springer, V.Y. Ivanov and R.L. Bras. 2011. Real-world hydrologic assessment of a fully-distributed hydrological model in a parallel computing environment. *Journal of Hydrology* 409(1–2): 483–496.

Vivoni, E.R., A.J. Rinehart, L.A. Méndez-Barroso, C.A. Aragón, G. Bisht, M.B. Cardenas, E. Engle, B.A. Forman, M.D. Frisbee and H.A. Gutiérrez-Jurado. 2008b. Vegetation controls on soil moisture distribution in the Valles Caldera, New Mexico, during the North American monsoon. *Ecohydrology* 1(3): 225–238.

Vivoni, E.R., J.C. Rodríguez and C.J. Watts. 2010. On the spatiotemporal variability of soil moisture and evapotranspiration in a mountainous basin within the North American monsoon region. *Water Resources Research* 46(2).

Vivoni, E.R., K. Tai and D.J. Gochis. 2009. Effects of initial soil moisture on rainfall generation and subsequent hydrologic response during the North American monsoon. *Journal of Hydrometeorology* 10(3): 644–664.

Walker, J., R. Panciera and E. Kim. 2008. High resolution airborne soil moisture mapping. In *Proceedings of the 14th Australasin Remote Sensing and Photogrammetry Conference, Darwin, Australia, 30 Sept-2 Oct*.

Walker, J.P., O. Merlin, R. Panciera, J.D. Kalma, E. Kim and J.M. Hacker. 2006. National Airborne Field Experiments for soil moisture remote sensing. In *Proceedings 30th Hydrology and Water Resources Symposium*.

Western, A.W., G. Blöschl and R.B. Grayson. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research* 37(1): 83–97.

Western, A.W., R.B. Grayson, G. B16schl, G.R. Willgoose and T.A. Mcmahon. 1999. Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resources Research* 35(3): 797–810.

Wood, E.F., J.K. Roundy, T.J. Troy, L.P.H. Van Beek, M.F.P. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek and J.S. Famiglietti. 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research* 47(5).

Xiang, T., E.R. Vivoni and D.J. Gochis. 2014. Seasonal evolution of ecohydrological controls on land surface temperature over complex terrain. *Water Resources Research* 50(5): 3852–3874.

Xiang, T., E.R. Vivoni, D.J. Gochis and G. Mascaro. 2017. On the diurnal cycle of surface energy fluxes in the North American monsoon region using the WRF-Hydro modeling system. *Journal of Geophysical Research: Atmospheres* 122(17): 9024–9049.

Zhou, Q., S. Yang, C. Zhao, M. Cai, H. Lou, Y. Luo and L. Hou. 2016. Development and implementation of a spatial unit non-overlapping water stress index for water scarcity evaluation with a moderate spatial resolution. *Ecological Indicators* 69: 422–433.

Zink, M., J. Mai, M. Cuntz and L. Samaniego. 2018. Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature. *Water Resources Research*.

APPENDIX A

NAFE05 AND NAFE06 DATASETS

This appendix describes the data collected from NAFE05 and NAFE06 as stored in a disital format. This data is organized withing the digital folder (/Seagate Expansion Drive/DIGITAL_APPENDIX/AppendixA). The data is organized as follows:

Folder Name:	Description:
NAFE05	This folder contains data collected from NAFE05 and remote sensing products that I used for data analyses in Chapter 2.
NAFE06	Same as NAFE05 but for NAFE06

The folder "NAFE05" is organized as follows:

Folder Name:	Description:
GIS	This folder contains all GIS data for NAFE05. - Digital elevation model - Flight paths - Boundary of Goulburn watershed - Location of ground stations - Boundary of NAFE05 study area
Landcover	This folder contains land cover map derived from Landsat 5 Thematic Mapper (TM)
Soiltype	This folder contains soil type shape file ("Goulburn_soiltype_poly.shp") and database for it.
GroundObservations_SM	This folder contains all ground-based soil moisture observations at point scale collected during the field campaign.
GroundObservations_ST	This folder contains all ground-based soil temperature observations at point scale collected during the field campaign.
Soil_Moisture_1km	This folder contains aircraft-based soil moisture estimates at 1 km in binary format ("NAFE05_soil_moisture_1km.mat") and a paper of data provider (Rocco Panciera) that used the NAFE05 data ("Panciera2010_published.pdf")
Landsat_USGS	This folder contains three days of Landsat 5TM datasets that I used to calculate the indices to identify irrigated areas in Chapter 2.

The folder “NAFE06” is organized as follows:

Folder Name:	Description:
GIS	This folder contains all GIS data for NAFE06. - Digital elevation model - Flight paths - Boundary of Goulburn watershed - Location of ground stations - Boundary of NAFE06 study area
Landcover	This folder contains land cover map derived from Landsat 5 Thematic Mapper (TM)
Soiltype	This folder contains soil type shape file (“Goulburn_soiltype_poly.shp”) and database for it.
GroundObservations	This folder contains all ground-based soil moisture observations at point scale collected during the field campaign. “Hydraprobe_Point_data _regionalDays” “Hydraprobe_Point_data_transectDays” folder contains
Soil_Moisture_1km	This folder contains 11 days of aircraft-based soil moisture estimates at 1 km in cvs format (“smDOY.csv”). DOY: day of year.
Landsat_USGS	This folder contains three days of Landsat 5TM datasets that I used to calculate the indices to identify irrigated areas in Chapter 2.

APPENDIX B

RIO SONORA BASIN DATASETS

This appendix describes the data collected for the hydrological simulations over the Rio Sonora basin as stored in a disital format. This data is organized withing the digital folder (/Seagate Expansion Drive/DIGITAL_APPENDIX/AppendixB). The data is organized as follows:

Folder Name:	Description:
GIS	This folder contains all GIS datasets: - Digital elevation model (*.tif). - Rio Sonora basin boundary (*.shape). - Location of ground stations (*.shape). - Soil class map (*.tif). - Land cover map (*.tif). - NLDAS 1km boundary map for Rio Sonora basin.
MODIS	This folder contains MODIS products that were used to derive time-varying vegetation parameters and to validate the tRIBS. There are four folders: - NDVI: MOD13Q1 at 250m. - LAI: MOD15A2 at 1 km. - Albedo: MOD43A3 at 500m (white sky short wave). - LST (MOD11A1 and MYD11A1 at 1km.
NLDAS	This folder contains NLDAS reanalysis datasets that are not processed.
Observations_G	This folder contains station observations for soil moisture and land surface temperature.

APPENDIX C

tRIBS MODEL SETUP and OUTPUT

This appendix describes the setup and output of tRIBS simulations over the Rio Sonora basin as stored in a disital format. This data is organized withing the digital folder (/Seagate Expansion Drive/DIGITAL_APPENDIX/AppendixC). The data is organized as follows:

Folder Name:	Description:
tRIBS setup	This folder contains the model setups used in Chapter 3.
tRIBS output	This folder contains the tRIBS outputs of hyperresolution simulations conducted in Chapter 3 and used in Chapter 4.

The folder "tRIBS setup" is organized as follows:

Folder Name:	Description:
Input	This folder contains tabular format of soil parameters and land use parameters, bedrock depth, and initial ground water table obtained from 5-years spin-up simulation. The files are: - CONABIO_ST250_17class (*.sdtt and *.soi) - INEGI_LC250_14class (*.ldtt and *.lan) - RSN_bedrock30m (*.brd) - Groundwater_initial_500m_SpinupFinal (*.iwt)
LandUse	This folder contains vegetation parameters (*.txt) derived from MOIDS products using a set of empirical equations presented in Xiang et al. (2014). There are six subfolders: - AL20042013: Albedo. - CC20042013: Maximum Canopy storage Capacity (CC). - OT20042013: Optical Transmission coefficient (OT). - SR20042013: minimum canopy Stomatal Resistance (SR). - TF20042013: free ThroughFall coefficient (TF). - VF20042013: Vegetation Fraction (VF).
Rain	Hourly rainfall data obtained from NLDAS reanalysis product.
ReachFiles	Stream networks in the basin that tRIBS reads.
Reservoir	It contains information of reservoirs in Rio Sonora basin.
SoilTexture	This folder contains soil parameters (*.txt) generated and used in Chapter 3. The files are: - AS250: Saturated Anisotropy Ratio. - AU250: Unsaturated Anisotropy Ratio. - KS250: Saturated Hydraulic Conductivity. - PsiB250: Air Entry Bubbling Pressure. - SHcsDry250: Soil Heat Capacity in dry condition. - SHcsWet250: Soil Heat Capacity in wet condition.

	- ThetaR250: Residual Soil Moisture. - ThetaS250: Soil Moisture at Saturation. - VHksDry250: Volumetric Heat Conductivity in dry condition. - VHkswet250: Volumetric Heat Conductivity in wet condition. - F501: Decay parameter. - N250: Porosity. - M250: Pore distribution index.
Weather	This folder contains weather forcings (*.txt) obtained from NLDAS and downscaled in Chapter 3. This folder contains 5 subfolders: - IS: Incoming Solar radiation. - PA: Pressure. - RH: Relative humidity. - TA: Air Temperature. - US: Wind Speed.

The folder "tRIBS output" is organized as follows:

Folder Name:	Description:
Output	This folder contains two folders (hyd and voronoi) that store all the time series of simulations (*.mrf, *.pixel, *.qout, and *_Outlet.qout) and spatial outputs (*timestamp_00d and *timestamp_00i).
Restart	This folder contains files that stored simulation states at a certain simulation time.
Binaryoutputs	It contains binary outputs (*.mat) converted from dynamic (00d) and time-integrated (00i) spatial outputs that tRIBS generates. There are four folders: - 00d: all the dynamic spatial outputs - 00i: all the time-integrated spatial outputs - LST: simulated land surface temperature - SM: simulated soil moisture

APPENDIX D

MATLAB SCRIPTS

This appendix describes matlab scripts used to process data within this dissertation. The files are stored in digital format under the folder (Seagate Expansion/Drive/DIGITAL_APPENDIX/AppendixD). This folder is organized as follows:

Folder Name:	Description:
Chapter2	All the scripts used within chapter 2 to process data and build figures. There are two folders: DataProcess: Scripts used for data process. Figures: Scripts used for figure generation.
Chapter3	All the scripts used within chapter 3 to process data and build figures. There are two folders: DataProcess: Scripts used for data process. Figures: Scripts used for figure generation.
Chapter4	All the scripts used within chapter 4 to process data and build figures. There are two folders: DataProcess: Scripts used for data process. Figures: Scripts used for figure generation.