Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 112 (2008) 326–336
Remote Sensing of Environment
www.elsevier.com/locate/rse

Spatial sampling uncertainty in SMEX04 soil moisture fields: A data-based resampling experiment

Mekonnen Gebremichael *, Enrique R. Vivoni

Department of Earth and Environmental Science, New Mexico Institute of Mining and Technology, Socorro, NM 87801, USA

Received 6 April 2006; received in revised form 21 October 2006; accepted 8 December 2006

Abstract

A data-based resampling experiment is performed to estimate sampling errors of area-averaged soil moisture estimates due to spatial sampling by ground-based sensors. The data consists of high-resolution soil moisture images derived from the Polarimetric Scanning Radiometer (PSR/CX) sensor flown on an aircraft as part of the summer field experiment (SMEX04 — Soil Moisture Experiment 2004) in the monsoon region of Sonora, Mexico. The sampling characteristics are investigated by accounting for random networks and evenly spaced networks. For random network designs, we develop a simple model that can be used to estimate the sampling uncertainty (expressed as standard deviation of sampling error as a percentage of the areal mean soil moisture) as a function of the number of sensors, mean soil moisture content and averaging area. This model is valid for five or more sensors. The model should prove useful to those wishing to assess the area-averaged performance of a soil moisture network. Furthermore, the method of analysis is applicable to other study regions (Oklahoma, Iowa, Alabama, Georgia, and Arizona) where soil moisture fields have been mapped at high resolution using airborne passive microwave remote sensors.

Keywords: Soil moisture; Remote sensing; SMEX04; Sampling uncertainty

1. Introduction

Daily global soil moisture products available from existing earth observing satellites will enhance the characterization of the near-surface soil moisture state (Jacobs et al., 2004). A common goal of satellite soil moisture estimation techniques is to produce grid-averaged soil moisture values that are as close to the truth as possible. Two quantitative aspects of satellite soil moisture could be distinguished: estimation and validation. Estimation deals with developing retrieval algorithms, while validation deals with quantifying the error in the soil moisture estimates by comparison with the "ground truth" (i.e. independent and more accurate soil moisture estimates). Reliable error estimates are crucial to improve the retrieval algorithm, to attach an appropriate degree of confidence for users of the product and to allow optimal data assimilation techniques.

The conventional "ground truth" data for validating satellite soil moisture products comes from networks of ground-based observational sensors (Cosh et al., 2004; Jacobs et al., 2004; Yoo, 2001). It is well known that the ground-based sensors have their own errors, but the assumption is that these errors (averaged over the satellite footprint) are significantly smaller than the errors in satellite-based products. The major problem arising in this approach is that of a mismatch in scale between satellite output grids (10–50 km) and ground-based sensor points (~ 5 cm). In order to compare the satellite outputs and the ground-sensor observations, the latter is usually transformed to the scale of the satellite footprint. The transformation could be done by simply averaging the ground-based sensor observations within a satellite grid or by using geostatistical interpolation techniques. This transformation introduces spatial sampling error due to the inability of existing ground-based networks to capture the sub-grid scale variability.

The objective of this paper is to quantify the spatial sampling error in area-averaged near-surface soil moisture estimates obtained from a network of ground-based sensors having different distributions in space. The objective is accomplished

* Corresponding author. Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, U-37, Storrs, CT 06269-037, USA. Tel.: +1 860 486 2771; fax: +1 860 486 2298.
E-mail address: mekonnen@engr.uconn.edu (M. Gebremichael).

0034-4257/$ - see front matter © 2007 Published by Elsevier Inc.
doi:10.1016/j.rse.2006.12.021

by simulating sampling experiments over high-resolution soil moisture fields to estimate the sampling errors. The high-resolution (800 m × 800 m) soil moisture images were derived from the Polarimetric Scanning Radiometer (PSR/CX) sensor flown on an aircraft as part of the summer field experiment (SMEX04 — Soil Moisture Experiment 2004) in Sonora, Mexico. The study site is a semiarid region characterized by complex terrain and highly heterogeneous vegetation cover, which exhibits dramatic and fast response to rainfall forcing at and after the onset of the North American monsoon. There were six complete PSR/CX images covering an area of 50 km × 90 km taken in early August during the North American monsoon period over Sonora, but we are using only a sub-domain of 50 km × 75 km in our study.

First, we quantify the sampling error for a variety of number of sensors and soil moisture spatial statistics. To make the assessment valid for random networks (independent of specific networks), we use a resampling experiment to create an ensemble of 500 networks for each number of sensors. We develop a model that can be used to estimate the sampling uncertainty as a function of averaging area, number of sensors, and areal mean soil moisture. Second, we repeat the same analysis for evenly spaced sensors, as a function of the number of sensors and soil moisture condition.

The PSR/CX-derived soil moisture maps are subject to errors arising from different sources. The PSR/CX soil moisture maps have been compared against field measurements in both Vivoni et al. (2008-this issue) and Bindlish et al. (2008-this issue). Vivoni et al. (2008-this issue) found that ground and remotely-sensed soil moisture estimates exhibit similar spatial variations with changes in the mean water content, but there are clear differences in actual soil moisture values from the two estimates. Even though the PSR/CX maps may not represent the true soil moisture fields, we need only to assume that they represent realistic or plausible space-time patterns characterized by realistic probability distribution. Because sampling errors depend on relative differences rather than absolute values, we argue that the resampling approach we describe herein offers a new insight into the statistical structure of error distribution. This work differs from that of Yoo (2001) in that his analysis assumes that the soil moisture fields are (weakly) second-order stationary. Our results of second order moment statistics reveal that such assumptions may not be valid for this region.

2. Study area and its characteristics

The SMEX04 experiment was a large-scale soil moisture experiment conducted in cooperation with the National Aeronautic and Space Administration (NASA), the U.S. Department of Agriculture—Agricultural Research Service (USDA-ARS), and other federal agencies and universities. SMEX04 was conducted in two regional study sites (50 km × 75 km) established in Sonora (Mexico) and Arizona (USA). Our study focuses on the Sonora site. A comprehensive description of the experiment is available at http://hydrolab.arsusda.gov/smex04.

Fig. 1 presents our study area, a 50-km by 75-km box, and its topographic characteristics derived from a 90-m digital elevation model (DEM). The area is bounded by 30.50°N to the north, 29.83°N to the south, 110.75°W to the west, and 110.23°W to the east. Note the north–south trending mountain ranges and river valleys in the study area which form part of the Sierra Madre Occidental. Two major ephemeral (seasonal) rivers flow north–south through the region: Río San Miguel (west) and Río Sonora (east), with the former draining into the latter south of the domain. The topographic distribution is characterized by a high mean elevation and a large elevation range, which are primarily due to the effects of channel incision (Coblentz & Riitters 2004).

We partitioned the study area into six regions, each 25.6 km by 25.6 km, which is approximately equivalent to the spatial scale of the soil moisture products currently available from the AMSR-E

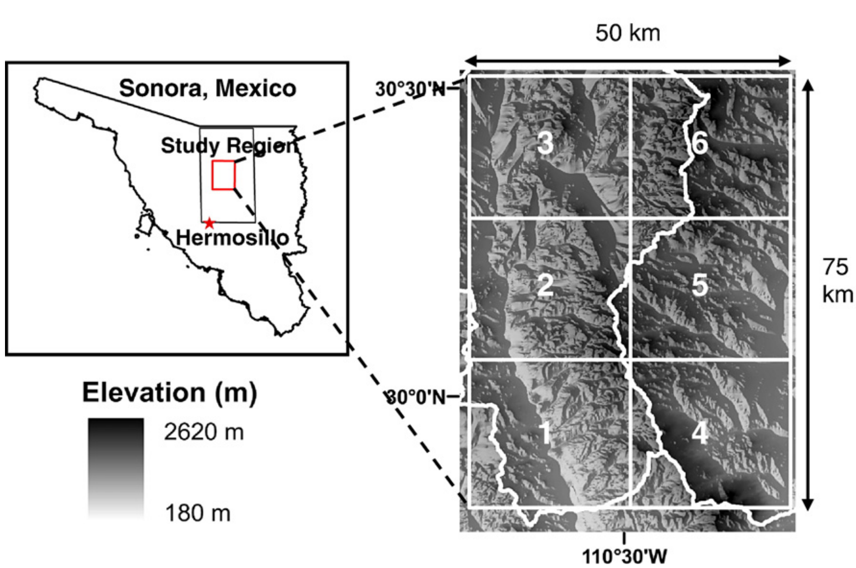

Fig. 1. The study area, a 50-km by 75-km box in northern Sonora, and its topographic characteristics derived from a 90-m digital elevation model (DEM). We partitioned the area into six regions (identified by the numbers), each 25.6 km by 25.6 km.

Table 1
Land cover types and their areal coverage (%) in each region (SIUE-IMADES, 1998)

Land cover type	Regions					
	1	2	3	4	5	6
Irrigated agriculture	3	3	3	2	0	0
Oak savanna/forest	9	16	6	30	19	44
Subtropical shrubland	75	8	0	54	58	0
Desert shrubland and mesquite forest	11	61	71	13	15	33
Grassland	2	12	17	0	6	23
Riparian forest	0	0	0	1	2	0
Bare soil	0	0	3	0	0	0

(Advanced Microwave Scanning Radiometer) instrument on board the Aqua satellite (Njoku et al., 2003). All regions except one (region 3) contain watershed divides indicating large relief variability within each region. The smallest relief range (850 m to 1750 m) is observed for region 3, and the largest for region 4 (600 m to 2030 m). Vivoni et al. (2007) showed that there is a strong control exerted by topography on the spatial and temporal variability in soil moisture, with distinct landscape regions experiencing different hydrologic regimes.

Table 1 presents an inventory of the land cover type which shows that plant communities vary considerably and include desert shrub, mesquite forest, subtropical shrub and oak savanna. The downstream regions (1 and 2) are dominated by subtropical shrublands, known as Sinaloan Thornscrub (Brown, 1994). This ecosystem primarily consists of thorny trees and shrubs, such as Palo Verde (*Cercidume sonorae*) and Palo Blanco (*Piscidia mollis*), which leaf-on and become green during rainy periods. Region 6 (characterized by the highest elevations) is dominated by oak savanna; this ecosystem is part of the Madrean evergreen woodland and consists of individual trees, typically Emory Oak (*Quercus emoryi*), interspersed with grasses and cacti (Brown, 1994). The remaining regions are dominated by either subtropical shrublands or desert shrublands and mesquite forest. Soils include a wide range of textures with large regions of both coarse and medium textures. Table 2 gives broad soil texture class characteristics in terms of coarse, medium and fine. The soils are dominantly characterized by either coarse or medium textures, with coarse soils becoming more dominant in the southern portion of the study region.

The study area is a semiarid region, which lies in the periphery of the core of the North American monsoon region, and receives 40–65% of its annual rainfall from the monsoon (Douglas et al., 1993). Gebremichael et al. (2007) investigated in detail the spatial and temporal variability of summer 2004 rainfall in the study area. They reported that (i) the two-month (July–August) summer 2004 rainfall varied from 130 mm to 260 mm depending on the location, (ii) the spatial correlation of rainfall between two locations decays exponentially with increasing the separation distance, and becomes insignificant when the distance reaches ~15 km, and (iii) in contrast to the upstream (northern section of the study area), the downstream section is characterized by strong convective systems that peak late diurnally and have smaller rainfall totals.

3. PSR/CX soil moisture datasets and their statistical properties

The Polarimetric Scanning Radiometer (PSR/CX), an airborne imaging radiometer working at both C- and X-band frequencies operated by the NOAA Environmental Technology Laboratory (Piepmeier & Gasiewski, 2001), was flown aboard the NASA P-3 aircraft for the purpose of obtaining polarimetric microwave emission. At 7300 m above sea level, the flights were mostly operated during the morning hours to avoid development of convective systems, which can hinder an aircraft operation. Typically, 1 h was required to complete the study area. The PSR/CX measurements supplied six complete maps of the region (5, 9, 10, 12, 13, and 14 August 2004). The approximate flight times were roughly 15:15–16:00 local hours on 5 August 2004, and 9:20–10:25 on the other days. Bindlish et al. (2008—this issue) provides details regarding the algorithm used to obtain soil moisture estimates from the PSR/CX 7.32H GHz brightness temperature measurements in a regularly spaced grid at nominal resolution of 800 m. Fig. 2 displays the images of the PSR/CX-derived volumetric soil moisture content (%vsm) over the study area. The driest surface condition has a 2%vsm soil moisture content. The land surfaces are characterized by a wide range of moisture conditions varying from the wettest (5 August 2004) to the driest (14 August 2004). The soil moisture fields generally exhibit spatial variability, the magnitude of which varies regionally and temporally. Below, we offer the statistical descriptions of these fields for each ~25-km by ~25-km regional box.

Fig. 3 shows the relationship between the coefficient of variation (CV) and the mean of the volumetric soil moisture content, for each region, derived from the six soil moisture images shown in Fig. 2. The regional soil moisture content ranges from 2% to 20%vsm. Region 1 (a downstream region) registers the highest regional mean (20%vsm), whereas region 3 (an upstream region) registers the lowest (2%vsm). This suggests that satellite soil moisture algorithms need to have the capability to resolve soil moisture variability ranging from 2%vsm to 20%vsm, so that they could be useful in semiarid monsoon regions. Fig. 3 also reveals that there is a relationship between CV and areal mean soil moisture content: CV decreases with increasing mean soil moisture for values >2.7%vsm. The decay of relative variability with increasing wetness is consistent with the findings of earlier studies (Bell et al., 1980; Charpentier & Groffman, 1992; Famiglietti et al., 1999; Owe et al., 1982). The implication is that large rainfall events leading to soil moisture increases result in smaller sub-grid scale variability. For land surfaces with areal mean soil moisture exceeding 10%

Table 2
Soil texture classes and their areal coverage (%) in each region (INIFAP, 2001)

Soil texture class	Regions					
	1	2	3	4	5	6
Coarse	69	52	10	86	85	33
Medium	31	46	77	14	13	66
Fine	0	2	13	0	2	1

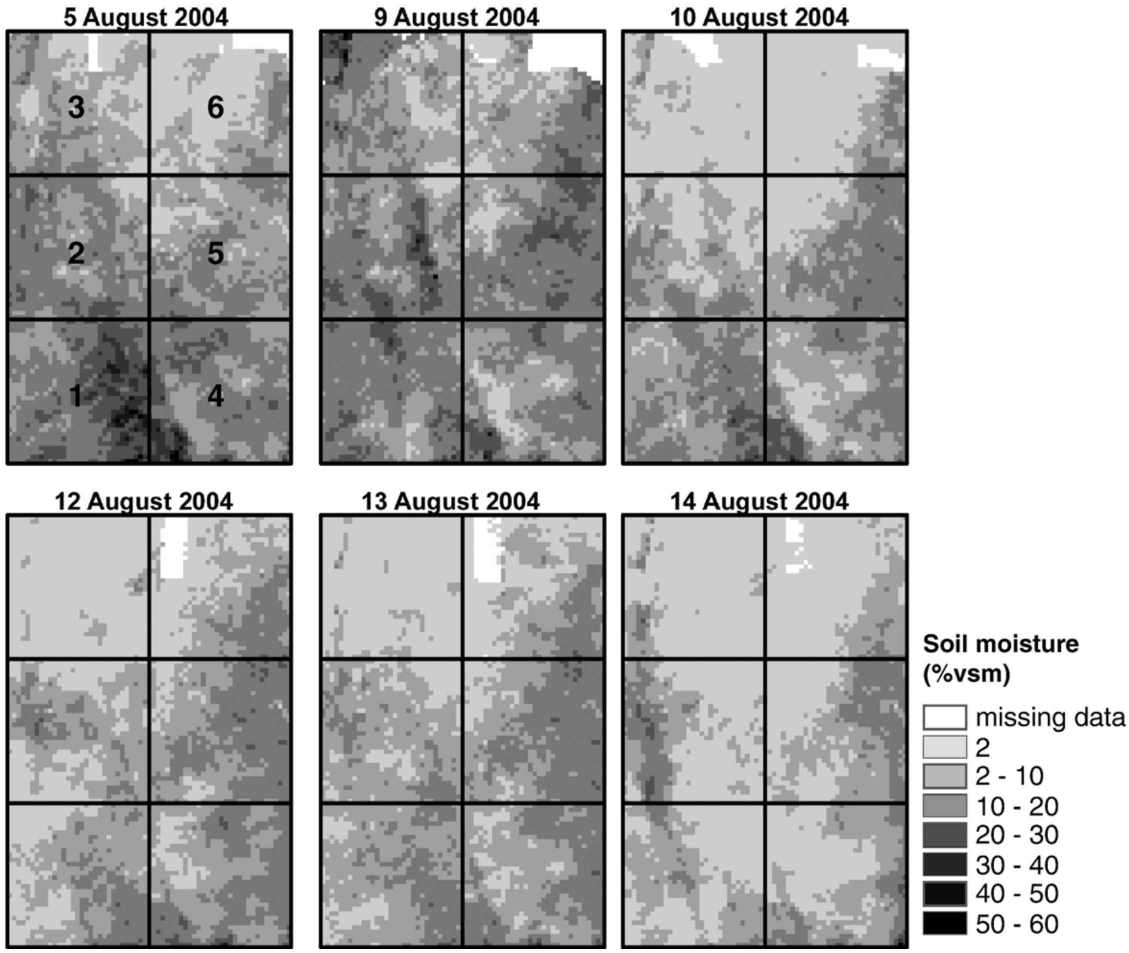

Fig. 2. PSR/CX-derived volumetric near-surface soil moisture content (%vsm) over the study area. The area is divided into six regions numbered as shown on the left top panel.

vsm, the CV is smaller than 55%, while it can reach up to 100% when the areal mean soil moisture is less than 10%vsm.

In the above analysis, we used the normalized standard deviation statistic to measure the sub-grid (or inter-footprint) scale variability; that is the variability among 800-m PSR pixels within a 25-km by 25-km regional box. The standard deviation measures the variability of the majority (for example 68% of the population clustered around the mean for normally distributed random variables) of the PSR/CX footprints. We used another statistic called the mean relative difference to examine how each of the PSR/CX footprints (0.8 km × 0.8 km) differs from the regional (25.6 km × 25.6 km) mean. The mean relative difference ($\overline{\eta}_{i,j}$) is defined as

$$\overline{\eta}_{i,j} = \frac{1}{n_t}\sum_{t=1}^{n_t} \frac{\theta_{i,j,t} - \overline{\theta}_{j,t}}{\overline{\theta}_{j,t}}, \tag{1}$$

where

$$\overline{\theta}_{j,t} = \frac{1}{n_{j,t}}\sum_{i=1}^{n_{j,t}} \theta_{i,j,t}, \tag{2}$$

where, $t = 1, 2, \ldots, n_t$ (number of dates), $j = 1, 2, \ldots, n_c$ (number of regions), and $i = 1, 2, \ldots, n_{j,t}$ (number of footprints within region j at time t). The mean relative difference at a footprint identifies whether that location is wetter (i.e. $\overline{\eta}_{i,j} > 0$) or drier (i.e. $\overline{\eta}_{i,j} < 0$) than the region on average. Fig. 4 depicts the spatial map of the mean relative difference ($\overline{\eta}_{i,j}$) in percentages. The minimum value of the mean relative difference is −80%, and the maximum value reaches 370%, indicating that the spatial

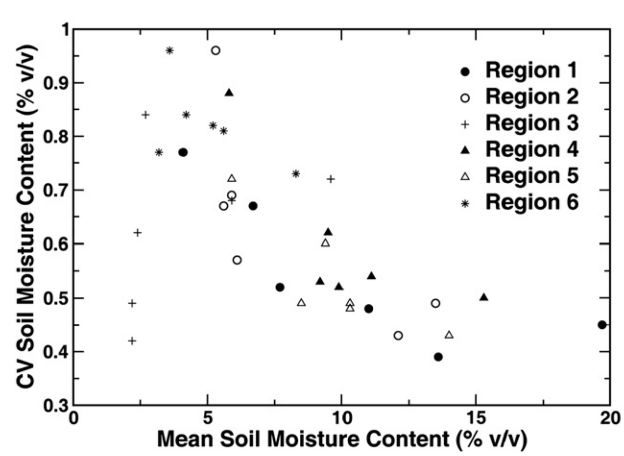

Fig. 3. The coefficient of variation (CV) as a function of the mean of volumetric soil moisture content during the SMEX04 experiment, derived from six PSR/CX images shown in Fig. 2.

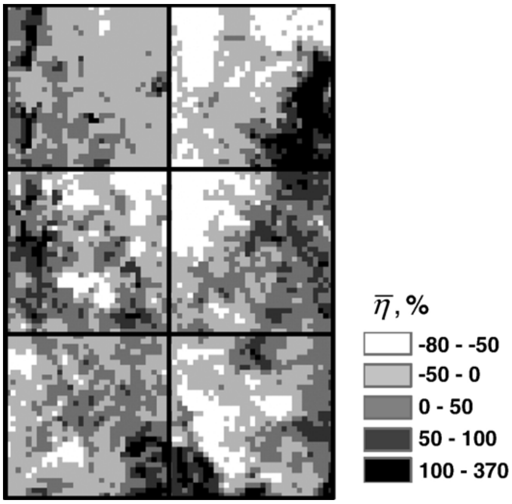

Fig. 4. (a) Mean relative difference (%) of volumetric soil moisture content, computed from six PSR/CX images.

variability of soil moisture is skewed to the right. This suggests that the regional mean strongly underestimates the wettest areas than it overestimates the driest areas, consistent with the findings of Mohanty and Skaggs (2001) and Jacobs et al. (2004) for regions of the Little Washita watershed in Oklahoma and the Walnut Creek watershed in Iowa, respectively.

Fig. 5 exhibits the spatial soil moisture distribution function in the form of its probabilities of exceedence, for each region and day. The plots are in log-linear scales. Region 1 was the wettest on August 5th, and continued drying with each day. Region 2 became the wettest on August 9th, due to a strong storm event that occurred on August 8th. Then on August 10th, the region dried down, with soil moisture in some spots sharply dropping from 50%vsm to 15%vsm in one day. Wet and dry conditions alternated daily until August 14th. Region 3 shows similar temporal fluctuation with region 1. Region 4 was the wettest on August 5th, got drier on August 9th; and showed no appreciable change until August 14th. Regions 5 and 6 are similar: they became the wettest on August 9th, dried down on August 10th, showed no appreciable change until August 13th, and dried down on August 14th. Overall results show that the soil moisture fields show strong day-to-day variability, and the nature and trend of temporal fluctuations vary from region to region.

To identify any spatial structure in the soil moisture fields, we estimated the spatial correlation function using a two-step procedure. First, we obtained the inter-footprint correlation coefficient estimate as a function of the separation distance, for each realization (region and day). To perform this, we initially grouped several pairs of footprints which are approximately at a prefixed distance, h, apart. We then applied the standard (Pearson's) formula for correlation coefficient estimation. This approach implies that the mean, covariance and variances are estimated from a single realization. This implies that we have assumed the soil moisture process is ergodic, meaning the individual available realization manifests in the domain of definition of the same probability distribution of all the theoretically possible realizations (Baachi & Kottegoda, 1995). Another implicit assumption is isotropy (correlation depends only on the separation distance and not on the direction). In a topographically dominated landscape, the assumption of isotropy might not be realistic. Second, we fitted a parametric model to each set of the inter-footprint correlation coefficients. We chose the following modified-exponential formula for this modeling:

$$\rho_g(h) = c \, \exp\left[-\left(\frac{h}{d_0}\right)^p\right], \tag{3}$$

where h is the separation distance between two footprints, ρ_g is the fitted correlation estimate, and c, d_0, and p are functional parameters. We defined the correlation distance (d) as the

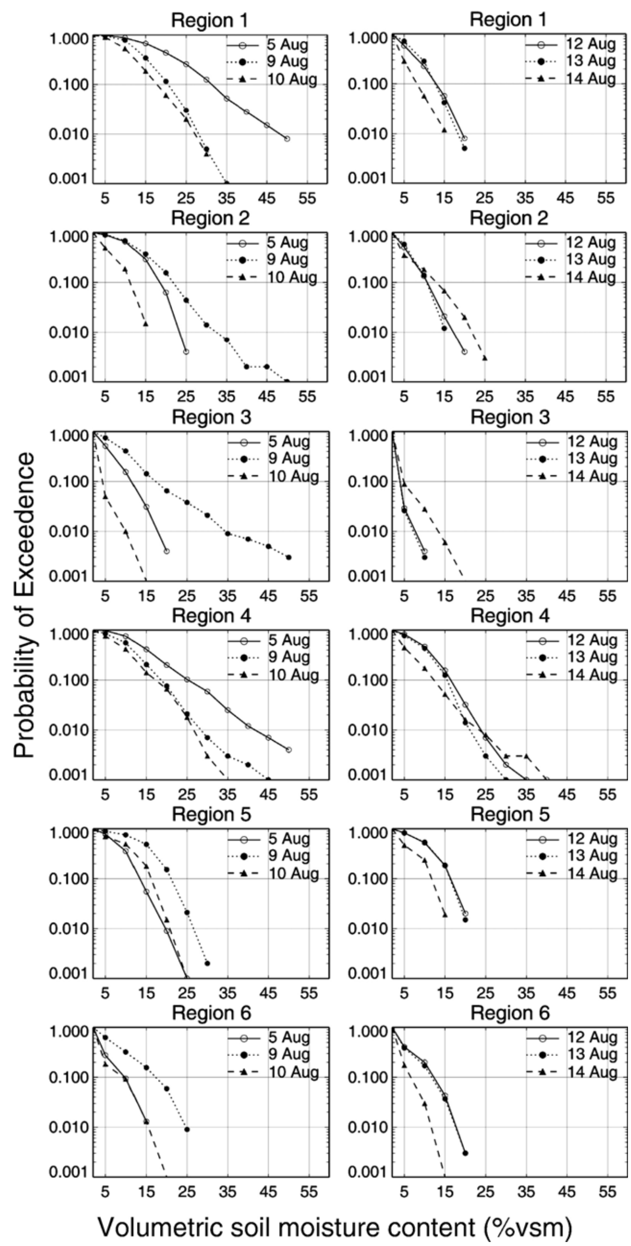

Fig. 5. Probabilities of exceedence of volumetric soil moisture obtained from six spatial PSR/CX images for each region.

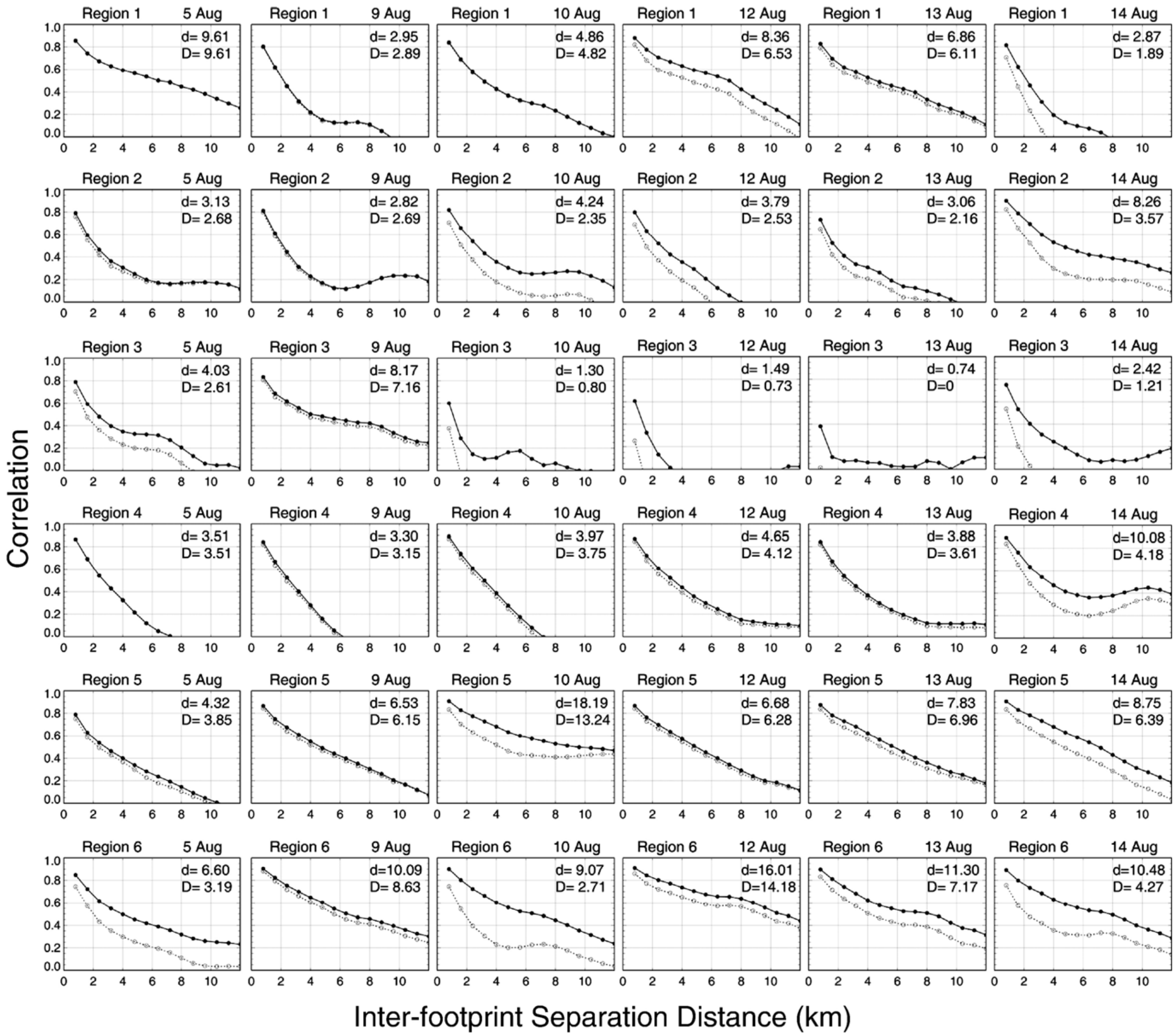

Fig. 6. Spatial correlation functions obtained from the PSR/CX-derived volumetric soil moisture contents for each date and region. Solid lines (with filled circles) represent (unconditional) correlation functions obtained using all available footprints, whereas dotted lines (with open circles) represent conditional correlation functions obtained using footprints that have soil moisture values exceeding 2%vsm. D and d refer to the conditional and unconditional correlation distances, respectively.

distance at which the correlation is $[\exp(1)]^{-1}$ or approximately 0.3679, consistent with the definition of the e-folding distance. We solved for the correlation distance from Eq. (3) as:

$$d = d_0\left[-\ln\left(\frac{1}{c\exp(1)}\right)\right]^{1/p}. \tag{4}$$

In Fig. 6, we present the estimated correlation coefficients between sets of pairs of footprints as a function of the separation distance, as well as the correlation distance d. In all cases, correlations decrease with increasing separation distance, as expected. The correlation distance for each day ranges from 3 to 10 km, 3 to 8 km, 1 to 8 km, 3 to 10 km, 7 to 19 km, and 7 to 16 km, for regions 1 through 6, respectively. This high daily fluctuation in the correlation distance suggests that considering the same correlation function for the entire set of events (as is often done in geostatistical techniques) may be unsatisfactory for this area. The case of the minimum correlation distance (~1 km) occurs for region 3 on August 13th. Inspection of Fig. 5 reveals that the driest soil moisture field occurs for region 3 on August 12th and 13th. Examination of Fig. 2 further shows that region 3 on August 13th consists of small isolated wet areas amidst large regions of dry soil. The case of the largest correlation distance (18 km) is observed for region 5 on August 10th. Inspection of Fig. 4 reveals that this case takes place as the region is drying down uniformly. The second largest correlation distance (16 km) is observed on August 12th for region 6, which is also on a dry-down phase.

As pointed out by Berndtsson (1987), the correlation estimate is affected by the inclusion or exclusion of dry areas in the analysis. We repeated the above exercise but for areas with soil moisture content exceeding 2%vsm (in statistics language "conditioned on areas with greater than 2%vsm"). The resulting conditional correlation coefficients and correlation distances (D) are superimposed in Fig. 6. The conditional correlation coefficients are always equal to or less than the unconditional correlation coefficients, as expected. The difference between the two correlation coefficients varies from 0 to 0.4, higher values suggesting larger areas with soil moisture less than 2%vsm. As a result, the conditional correlation distances are also equal to or less than the unconditional values. The differences between the conditional and unconditional correlation distances vary from 0 to 6.4 km, larger values for areas dominated by moisture values less than 2%vsm. The conditional correlation distances vary from 2 to 10 km, 2 to 4 km, ~0 to 7 km, 3 to 4 km, 4 to 13 km, 4 to 14 km, for regions 1 through 6, respectively. The conditional correlation distances also exhibit high daily fluctuation, suggesting that the assumption of second-order stationarity in time (or considering the same correlation for the entire set of summer events) may not be appropriate for this area.

4. Sampling uncertainty analysis

Our starting point is the PSR/CX soil moisture content maps available at a resolution of 800 m × 800 m covering each region, an area of 25.6 km × 25.6 km. We assume that these maps are possible realizations of the true soil moisture random field. Using all the footprints in a given region j at time t results in the regional mean $\overline{\theta}_{j,t}$ (see Eq. (2)). Sampling a smaller number of footprints leads to a sampling error. Let $\hat{\theta}_{j,t}$ be the sample mean derived from the small number of footprints using

$$\hat{\theta}_{j,t} = \frac{1}{n_{j,t}} \sum_{i=1}^{n_{j,t}} \theta_{i,j,t} \delta_{i,j,t}, \tag{5}$$

where the Kronecker delta function, $\delta(i)$, is one if the footprint is selected and zero otherwise.

The sampling error is defined as

$$\varepsilon_{j,t} = \hat{\theta}_{j,t} - \overline{\theta}_{j,t}. \tag{6}$$

The sampling uncertainty of $\hat{\theta}_{j,t}$ as an estimate of $\overline{\theta}_{j,t}$ can be characterized by standard deviation of $\varepsilon_{j,t}$,

$$\sigma(\varepsilon_{j,t}) = \left(\overline{\varepsilon_{i,j}^2} \right)^{1/2}. \tag{7}$$

It is customary to express the sampling uncertainty in a dimensionless form as

$$S = \frac{\sigma(\varepsilon_{j,t})}{\overline{\theta}_{j,t}} 100\%. \tag{8}$$

We are interested in how sampling uncertainty (S) varies as a function of the number of sensors, sensor network configurations (random, stratified and currently deployed), and wetness condition. To address this, we simulated several realizations of the sampling error ($\varepsilon_{j,t}$) via resampling experiments on the PSR/CX soil moisture fields, as discussed below.

To perform the resampling experiment for randomly selected sensor networks, we followed the following procedure:

i. Generate two random numbers x and y ($1 \le x \le 32$; $1 \le y \le 32$) from a uniform distribution to decide on the sensor location. Note that there are 32 × 32 PSR/CX pixels in a given 25.6-km by 25.6-km region. To locate m sensors we need to generate 2 m random numbers. We carefully selected randomly uniform networks (i.e. we did not consider clustered networks as this would require case-dependent solutions). Each sensor is assumed to represent the PSR/CX footprint.
ii. Select m PSR/CX soil moisture data at the sensor locations.
iii. Calculate the sample mean soil moisture content by averaging the selected m soil moisture data as per Eq. (5).
iv. Calculate the sampling error as per Eq. (6) by taking the difference between the sample mean obtained in step (iii) and the regional population mean value obtained from Eq. (2).
v. Repeat steps (i)−(iv) 500 times.
vi. Repeat steps (i)−(v) with different number of sensors m.
vii. Repeat steps (i)–(vi) with each image over each region.

The 500 realizations for each PSR/CX soil moisture image, each region and each number of sensors will enable us to calculate the corresponding sampling uncertainty.

To perform the resampling experiment for evenly spaced sensor networks, we followed a similar procedure with the exception that here the sampling locations are specified by the evenly spaced sensors network (see Fig. 10) and not by random sampling.

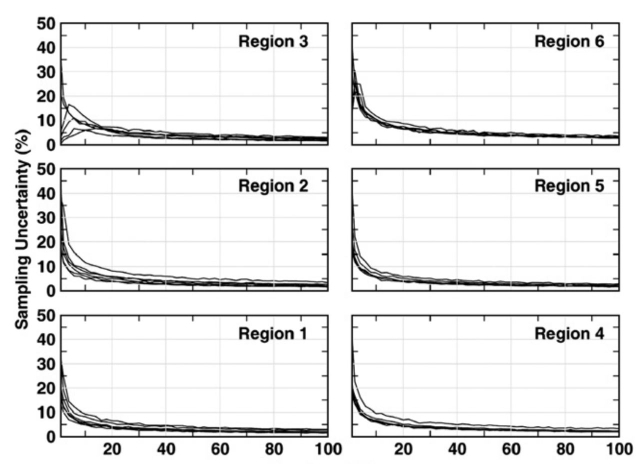

Fig. 7. Sampling uncertainty (S, Eq. (8)) as a function of the randomly located ground-based sensor network density for each region and date. Each panel shows the results for each region, and the different curves in within each panel correspond to the different PSR/CX images. The sampling uncertainty values are all valid for the 25.6-km × 25.6-km grid boxes.

5. Simulation results: random ground-based sensor networks

In Fig. 7, we show the sampling uncertainty (S, Eq. (8)) results as a function of randomly located number of sensors, for each region and soil moisture condition as obtained from the PSR/CX image. The sampling uncertainty values are all valid for 25.6-km × 25.6-km grid boxes that represent AMSR-E footprints. The sampling uncertainty decreases rapidly when the number of sensors increases until a threshold number is reached, beyond which the sampling uncertainty converges asymptotically. An exception occurs in region 3 when small number of sensors (<5) is used. For this range (1 to 5 sensors), the sampling uncertainty increases with increasing number of sensors. This unusual behavior in region 3 is caused by the fact that the region is characterized by vastly large dry areas, and hence any sampling from small number of sensors will most likely result in an estimate that exhibits little variation from one realization to another, leading to unusually small sampling uncertainty values. Generally, 10 to 50 sensors in each region are needed to bring the sampling uncertainty within 5% in all cases. The sampling uncertainty for each number of sensors varies over a range of values depending on the soil moisture condition. This range is relatively wide for small number of sensors, and narrows down with increasing number of sensors.

In Fig. 8, we show the sampling uncertainty as a function of the mean soil moisture content, for three quantities of sensors (2, 10 and 100). The sampling uncertainty and mean soil moisture values are all for 25.6-km × 25.6-km grid boxes. For each sample size, there are a total of 36 pairs of sampling uncertainty and mean soil moisture content resulting from six regions and six PSR/CX images per region. Results show that the sampling uncertainty decreases with increasing wetness, for mean soil moisture content exceeding 3.2%vsm. However, the rate of decay depends on the number of sensors: the decay is faster for smaller number of sensors than for larger number of sensors. For land surfaces with moisture content less than 3.2% vsm, the sampling uncertainty increases with increasing wetness, which could be attributed to the fact that sampling of soil moisture in largely dry areas underestimates the areal mean value but with smaller sampling uncertainty (i.e. the estimate changes little from one realization to another).

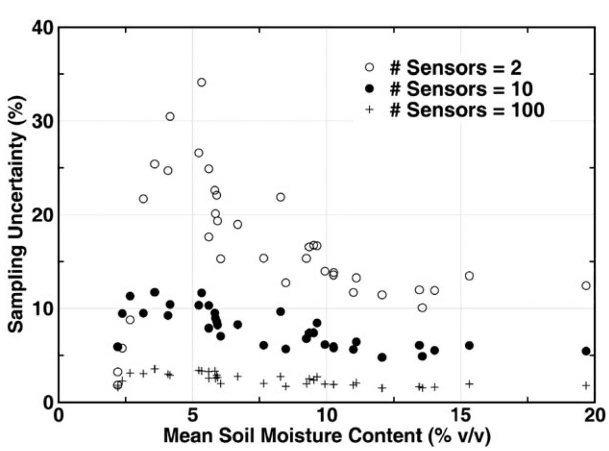

Fig. 8. Sampling uncertainty as a function of the mean soil moisture content, for three different numbers of sensors (2, 10 and 100). The sampling uncertainties and mean soil moisture values are all for 25.6-km × 25.6-km grid boxes.

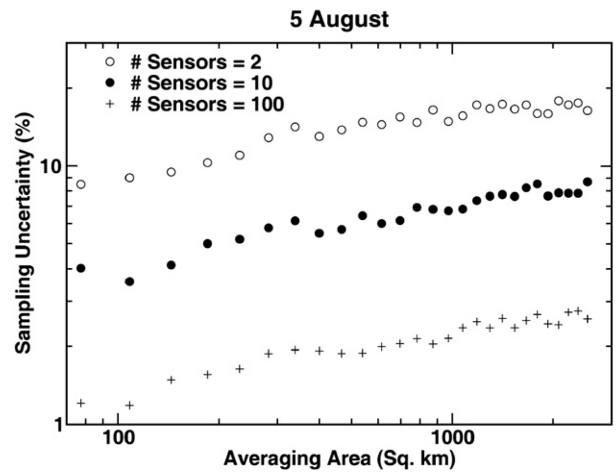

Fig. 9. Sampling uncertainty as a function of the averaging area, obtained from simulation experiments on the PSR/CX image dated 5 August, for three different numbers of sensors.

Our purpose is to develop a simple model for estimating temporal sampling uncertainty from easily available data (large-scale mean soil moisture value available from satellites), and hence the focus on mean soil moisture content. Having said this, we have also looked at the sub-grid scale variability in terms of the correlation function. Our results discussed in Section 3 show that the correlation function changes with soil moisture condition, and so there is no a single 'climatological correlation function" that can be used for sampling uncertainty estimation.

Let us now look at the dependence of the sampling uncertainty on soil moisture averaging areas. We selected 39 different averaging areas so that the largest area corresponds to the dimensions of 50.4 km × 50.4 km (i.e. 63 × 63 PSR/CX pixels, recall that each PSR/CX pixel has a dimension of 0.8 km × 0.8 km) and the smallest area encompasses a dimension of 8.8 km × 8.8 km (i.e. 11 × 11 PSR/CX pixels). These spatial scales (ranging from 50 km to 8 km) are relevant for future missions (e.g., SMOS).

In Fig. 9, we show the dependence of the sampling uncertainty on the soil moisture averaging areas, obtained from our simulation on the PSR/CX image dated 5 August. The plot is on a log–log scale in which straight lines are characteristics of scaling functions. Fig. 9 reveals that the sampling uncertainty obeys scaling law with respect to the averaging area, and it increases with increasing averaging area. We also found similar simulation results based on the other PSR/CX images.

In the preceding paragraphs, we have shown that the sampling uncertainty results are scalable with respect to the number of sensors (N), mean soil moisture content ($\overline{\theta}$) and averaging area (A). We will now develop a simple model that can be used to estimate sampling uncertainty for any combination of

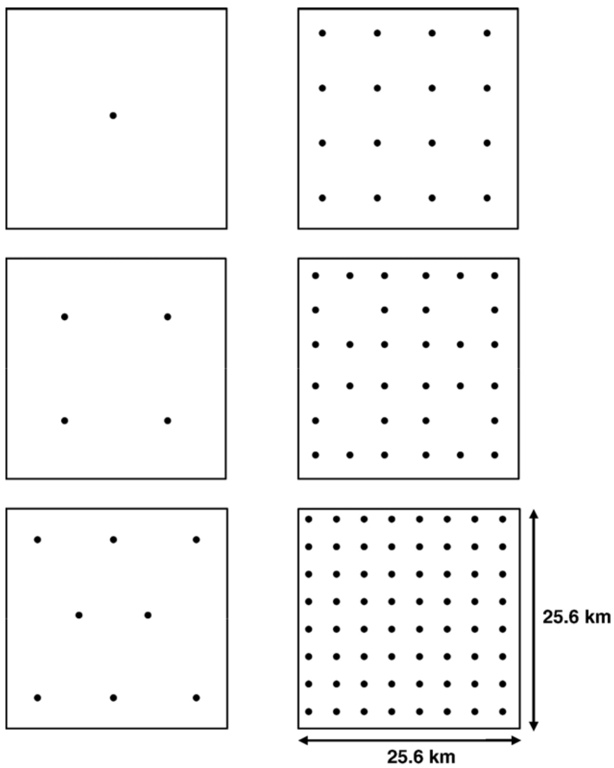

Fig. 10. The evenly spaced networks used to study the effect of various factors on the sampling uncertainty of soil moisture fields. The numbers of sensors shown on the plots are 1, 4, 8, 16, 32 and 64.

the N, $\overline{\theta}$ and A. The various scaling laws that our results suggest could be written as

$$S = f\{N^{-\alpha}, \overline{\theta}^{-\beta}, A^{-\gamma}\}. \tag{9}$$

However, we do not know the model form $f\{.\}$. Following Steiner et al. (2003) and Gebremichael and Krajewski (2004), who assumed a multiplicative model for the temporal sampling uncertainty in satellite-based rainfall estimates, we assumed the following simple empirical form:

$$S = aN^{-\alpha}\overline{\theta}^{-\beta}A^{\gamma}, \tag{10}$$

where A is in km^2, $\overline{\theta}$ is in %vsm, S is the sampling uncertainty in percent of mean, and parameters a, α, β and γ are obtained from regression fit. Using simulations results from all PSR/CX images, all regions, sample sizes ranging from 1 to 100 and averaging areas ranging from 8.8 km × 8.8 km to 50.4 km × 50.4 km, we estimated the parameters involved in Eq. (10) by multiple regression analysis after taking the logarithms of both sides. We found the following results: a = 11.3, α = 0.49, β = 0.19 and γ = 0.16. The associated standard errors are negligible (<0.01). These parameter estimates are obtained so that they minimize the mean square difference between the fitted and actual uncertainty estimates. We assessed the performance of the sampling uncertainty model (i.e. Eq. (10)), by comparing its estimates to the actual sampling uncertainty estimates. For 90% of the cases, the estimates made by Eq. (10) fall within the range 0.80–1.20 times the actual sampling uncertainty estimates, whereas this range narrows down to 0.91–1.06 times for 50% of the cases. Note that the scaling exponent corresponding to the number of sensors is about −0.50, indicating that the samples gained by increasing the number of sensors are uncorrelated (from statistical theory). We add the caveat that Eq. (10) fails to capture the relationship between sampling uncertainty and number of sensors, for small number of sensors (<5) in vastly dry regions (see the first paragraph in this section for additional information).

6. Simulation results: evenly spaced ground-based sensor networks

We constructed six networks (Fig. 10) to quantify the sampling error for evenly spaced sensor networks. The networks are composed of 1, 4, 8, 16, 32 and 64 sensors inside each 25.6-km × 25.6-km region. As discussed in Section 4, fixing the sensor locations gives us only one realization of sampling error ($\varepsilon_{j,t}$ Eq. (6)) for each regional PSR/CX image at a fixed sample size. For this reason, the results shown in this section and the next are sampling error values ($\varepsilon_{j,t}$) and not sampling uncertainties (S).

In Fig. 11, we show the sampling error in percent of mean (i.e. $100\varepsilon_{i,j}/\overline{\theta}_{i,j}$) as a function of the number of sensors, for each regional PSR/CX image. The dependence of the sampling error on the number of sensors shows little intra-regional variability and large inter-regional variability. For example in region 4, the sampling errors are mostly negative, the absolute value of which decreases with increasing number of sensors. In region 5, the sampling errors are positive for small number of sensors (1 to

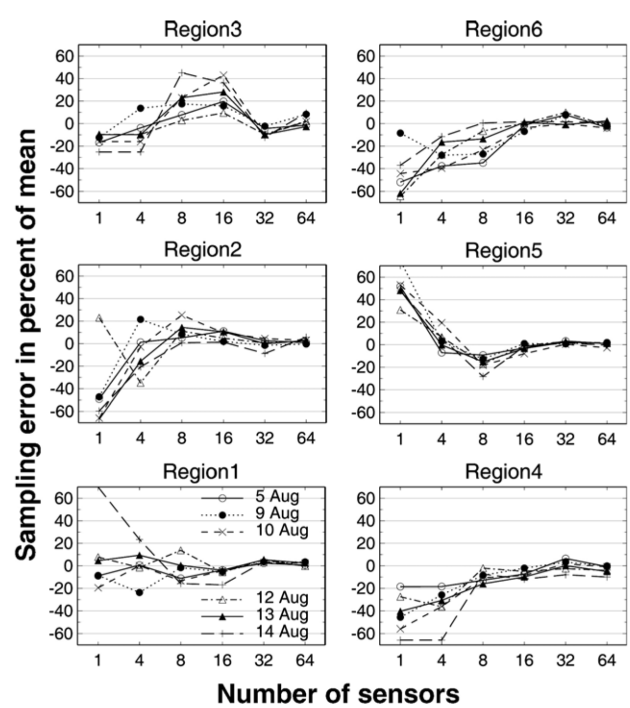

Fig. 11. Sampling error in percent of mean ($100\varepsilon_{i,j}/\overline{\theta}_{i,j}$) as a function of the number of sensors with locations corresponding to Fig. 10, for each regional PSR/CX image. Each panel shows the results for one region, and the lines within each panel correspond to different PSR/CX images.

4), become negative for medium number of sensors (8 to 16), and get close to zero beyond 32 sensors. Overall, the sampling error varies within ±70% for one sensor, and within ±10% for 32 sensors. The actual value of the estimated sampling error depends on the region and the underlying soil moisture field.

7. PSR/CX versus ground-based sensors

The study of the sampling error using PSR/CX footprints (800 m × 800 m) may not reflect the true magnitude of the sampling error one would encounter using ground-based sensor observations (0.05 m × 0.05 m). The PSR/CX footprint observations exhibit less variability than the ground-based sensors, and hence the use of PSR/CX footprints tends to underestimate the sampling error. The sampling uncertainty model, i.e. Eq. (10) with the estimated parameter values, derived from the PSR/CX footprints therefore needs to be modified to obtain results applicable for ground-based sensors:

$$S = 11.3 f_r N^{-0.49} \overline{\theta}^{\,-0.19} A^{0.16}, \tag{11}$$

where f_r is the correction factor that accounts for the area discrepancy between the ground-based sensors and PSR/CX footprints. Further study, based on dense networks of ground-based sensors, is needed to estimate the value of f_r.

According to our model (Eq. (10)), there are only two parameters that reflect the effect of grid-box size on the sampling uncertainty: γ (estimated to be 0.16) and a (estimated to be 11.3). γ is estimated based on a large range of grid-box sizes, and hence it may not change significantly as one goes beyond the range of the spatial scales used in this analysis. Having said this, we believe that the parameter a will change when using the ground-based sensor (instead of PSR/CX) for reasons discussed in the preceding paragraph. We have introduced the correction factor f_r to account for this (see Eq. (11)).

8. Summary and conclusions

Satellite-based soil moisture estimation holds great promise to enhance characterization of near-surface soil moisture state across the globe. The usability of the products requires validation against independent and more accurate soil moisture estimates. Although ground-based sensors provide direct measurements of soil moisture (in contrast to satellite-based sensors that can only detect electromagnetic signals indirectly related to soil moisture), their footprint is of the order of 5 cm × 5 cm. On the other hand, area resolution (or grid box size) of satellite-based products could range from 10 km × 10 km to 50 km × 50 km. Therefore, in order to use validation data from ground-based sensor networks, we need to quantify the spatial sampling error for soil moisture estimates derived from ground-based sensor networks, as a function of number of sensors, averaging region and the characteristics of the soil moisture field.

We investigated the sampling error in the instantaneous areal mean soil moisture estimates by making use of six PSR/CX-derived soil moisture images in a semiarid monsoon region of Sonora, Mexico. Using a resampling experiment, we estimated the sampling error as a function of ground-based sensor network type, number of sensors, grid box size, and areal mean soil moisture.

A key result of our analysis is that the sampling uncertainty scales with the number of sensors, grid box size and areal mean soil moisture. We developed a model that can be used to estimate the sampling uncertainty. The estimated sampling uncertainty (S, expressed as standard deviation of sampling error as a percentage of the areal mean soil moisture) is inversely proportional to the number of sensors and the areal mean soil moisture but proportional to the grid box size

$$S = 11.3 f_r N^{-0.49} \overline{\theta}^{\,-0.19} A^{0.16}.$$

The factor f_r is introduced to account for the area discrepancy between the PSR/CX footprints used in this study and the ground-based sensors. Further investigation using dense network of ground-based sensors is required to estimate f_r. This model is valid for five or more ground-based sensors. We note that this is the first time that the sampling uncertainty model shown above is proposed for soil moisture applications.

The main advantage of this model is that it allows calculation of the sampling uncertainty from the coarse-grid satellite estimates (i.e. without the knowledge of sub-grid scale variability). We acknowledge that such a model is subject to error, that's why we treated the model in a probabilistic framework. Our results have shown that for 90% of the spatial soil moisture realizations, our simple model produces sampling uncertainty estimates that have standard errors to within ±20%.

The model mentioned above can be used to estimate the sampling uncertainty arising from the use of ground-based sensor network in estimating areal average soil moisture for areas ranging from about 10 km × 10 km to 50 km × 50 km. The results are therefore applicable to existing satellite-based soil moisture products, such as the AMSR-E products. The resulting sampling uncertainty estimates are of paramount importance for designing new validation networks, quantifying the error in existing validation networks and filtering out this error from the validation of satellite estimates, and assimilating network estimates into satellite soil moisture algorithms. Furthermore, the method presented here is applicable to other study regions (Oklahoma, Iowa, Alabama, Georgia, and Arizona) where soil moisture fields have been mapped at high resolution using airborne passive microwave remote sensors (e.g., Cosh et al., 2004, 2005; Jackson et al., 2005; Jacobs et al., 2004; Mohanty & Skaggs, 2001).

We note that the results of this study (e.g., the regression parameter estimates) may not be necessarily applicable to other regions characterized by different climatology, topography and land surface conditions. Further study is therefore required to assess the applicability and limitation of this sampling uncertainty model in other regions.

Acknowledgments

We would like to acknowledge funding for this work from NSF NM EPSCoR project. We particularly thank Thomas J.

Jackson and Rajat Bindlish (USDA-ARS) for providing the PSR/CX soil moisture data over the SMEX04 Sonora region. In addition, we thank David Gochis (NCAR), Julio C. Rodríguez (IMADES), Christopher J. Watts (Universidad de Sonora) and Jaime Garatuza-Payán (ITSON) for providing useful ancillary data sets used in this study. This effort would not have been possible without the organization and implementation of the SMEX04 experiment in Sonora. We are grateful to all the participants in the field activities and initial data processing.

References

Baachi, B., & Kottegoda, N. T. (1995). Identification and calibration of spatial correlation patterns of rainfall. *Journal of Hydrology*, *165*, 311−348.

Bell, K. R., Blanchard, B. J., Schmugge, T. J., & Witczak, M. W. (1980). Analysis of surface moisture variations within large field sites. *Water Resources Research*, *16*, 796−810.

Berndtsson, R. (1987). On the use of the cross-correlation analysis in studies of patterns of rainfall variability. *Journal of Hydrology*, *93*, 113−134.

Bindlish, R., Jackson, T. J., Gasiewski, A., Stankov, B., Klein, M., Cosh, M. H., Mlandenova, I., Watts, C., Vivoni, E., Lakshmi, V., Bolten, J., & Keefer, T. (2008). Aircraft based soil moisture retrievals under mixed vegetation and topographic conditions. *Remote Sensing of Environment*, *112*, 375−390 (this issue). doi:10.1016/j.rse.2007.01.024.

Brown, D. E. (1994). *Biotic communities Southwestern United States and Northwestern Mexico.* Salt Lake City, UT: University of Utah Press 342 pp.

Charpentier, M. A., & Groffman, P. M. (1992). Soil moisture variability within remote sensing pixels. *Journal of Geophysical Research*, *97*, 18,987−18,995.

Coblentz, D. D., & Riitters, K. H. (2004). Topographic controls on the regional-scale biodiversity of the south-western USA. *Journal of Biogeography*, *31* (7), 1125−1138.

Cosh, M. H., Jackson, T. J., Bindlish, R., Famiglietti, J. S., & Ryu, D. (2005). Calibration of an impedance probe for estimation of surface soil water content over large regions. *Journal of Hydrology*, *311*, 49−58.

Cosh, M. H., Jackson, T. J., Bindlish, R., & Prueger, J. H. (2004). Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment*, *92*, 427−435.

Douglas, M. W., Maddox, R., Howard, K., & Reyes, S. (1993). The Mexican Monsoon. *Journal of Climate*, *6*, 1665−1677.

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

Gebremichael, M., & Krajewski, W. F. (2004). Characterization of the temporal sampling error in space-time averaged rainfall estimates from satellites. *Journal of Geophysical Research*, *109*, D11110. doi:10.1029/2004JD004509

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

INIFAP, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. (2001). *Inventario Nacional de Suelos 1:250,000*.

Jackson, T. J., Bindlish, R., Gasiewski, A. J., Stankov, B., Klein, M., Njoku, E. G., et al. (2005). Polarimetric scanning radiometer C- and X-band microwave observations during SMEX03. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(11), 2418−2430.

Jacobs, J. M., Mohanty, B. P., Hsu, E. -C., & Miller, D. (2004). SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sensing of Environment*, *92*, 436−446.

Mohanty, B. P., & Skaggs, T. H. (2001). Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Advances in Water Resources*, *24*, 1051−1067.

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

Owe, M., Jones, E. B., & Schmugge, T. J. (1982). Soil moisture variation patterns observed in Hand County, South Dakota. *Water Resources Bulletin*, *18*, 949−954.

Piepmeier, J. R., & Gasiewski, A. J. (2001). High-resolution passive microwave polarimetric mapping of ocean surface wind vector fields. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 606−622.

SIUE-IMADES, Instituto del Medio Ambiente y Desarrollo Sustentable del Estado de Sonora (1998). *Proyecto de Ordenamiento Ecológico del Territorio del Estado de Sonora.* 577 pp.

Steiner, M., Bell, T. L., Zhang, Y., & Wood, E. F. (2003). Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar dataset. *Journal of Climate*, *16*(22), 3759−3778.

Vivoni, E. R., Gebremichael, M., Watts, C. J., Bindlish, R., & Jackson, T. J. (2008). Comparison of ground-based and remotely-sensed surface soil moisture estimates over complex terrain using SMEX04. *Remote Sensing of Environment*, *112*, 314−325 (this issue). doi:10.1016/j.rse.2006.10.028.

Vivoni, E. R., Gutiérrez-Jurado, H. A., Aragón, C. A., Méndez-Barroso, L. A., Rinehart, A. J., Wyckoff, R. L., et al. (2007). Variation of hydrometeorological conditions along a topographic transect in northwestern México during the North American monsoon. *Journal of Climate*, *20*(9), 1792−1809.

Yoo, C. (2001). Sampling of soil moisture fields and related errors: Implication to the optimal sampling design. *Advances in Water Resources*, *24*(5), 521−530.