An approach to constructing a homogeneous time series of soil moisture using SMOS

D. Leroux, Yann H. Kerr, E. F. Wood, A.K. Sahoo, R. Bindlish, T.J. Jackson

An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS

Delphine J. Leroux, Yann H. Kerr, *Fellow, IEEE*, Eric F. Wood, Alok K. Sahoo, Rajat Bindlish, *Senior Member, IEEE*, and Thomas J. Jackson, *Fellow, IEEE*

***Abstract*—Overlapping soil moisture time series derived from two satellite microwave radiometers (the Soil Moisture and Ocean Salinity and the Advanced Microwave Scanning Radiometer-Earth Observing System) are used to generate a soil moisture time series from 2003 to 2010. Two statistical methodologies for generating long homogeneous time series of soil moisture are considered. Generated soil moisture time series using only morning satellite overpasses are compared to ground measurements from four watersheds in the U.S.A. with different climatologies. The two methods, cumulative density function (CDF) matching and copulas, are based on the same statistical theory, but the first makes the assumption that the two data sets are ordered the same way, which is not needed by the second. Both methods are calibrated in 2010, and the calibrated parameters are applied to the soil moisture data from 2003 to 2009. Results from these two methods compare well with ground measurements. However, CDF matching improves the correlation, whereas copulas improve the root-mean-square error.**

***Index Terms*—Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), cumulative density function (CDF) matching, copulas, Soil Moisture and Ocean Salinity (SMOS), soil moisture, time series.**

I. Introduction

Soil moisture is an important variable and is now considered as an essential climate variable by the World Meteorological Organization [1]. It has a crucial role in the transfers of water and energy between the soil and the atmosphere. Soil moisture is also an input variable for land surface modeling in determining the evaporative fraction at the surface and the infiltration in the root zone. For both agriculture and water resource management, soil moisture information is essential at local and regional scales. At global scales, soil moisture is of great value for weather forecasting [2], climate change [3], and monitoring extreme events such as floods and droughts.

Soil Moisture and Ocean Salinity (SMOS) [4] was successfully launched by the European Space Agency in November 2009 and since has been providing global maps of soil moisture every three days at a nominal spatial resolution of 43 km with an accuracy of 0.04 m^3/m^3. SMOS is the first mission specifically designed for soil moisture monitoring. The Soil Moisture Active Passive (SMAP) mission [5] is scheduled for launch in October 2014 by the National Aeronautics and Space Administration. SMAP will continue the time series of soil moisture based on 1.4-GHz radiometer observations that began with SMOS. The 1.4-GHz frequency channel is the most suitable frequency for soil moisture retrieval [6].

Longer time series of satellite-based soil moisture would be of value in climate-related analysis. Utilizing the data from the previous generations of satellite sensors involves resolving numerous issues. Some of the platforms and approaches have been developed to retrieve soil moisture using the higher frequencies, which has been the only option until now. These include the Scanning Multichannel Microwave Radiometer (1978–1987) [7], the Special Sensor Microwave/Imager (1987–current) [7], the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (2002–2011) [7], [8], WindSat (2003–current) [9], and the European Remote Sensing-Advanced Scatterometer (1991–current) [10]. Although their lowest frequencies (5–20 GHz) are not the most suitable for soil moisture retrievals (higher sensitivity to vegetation growth and atmospheric conditions), they remain a valuable time series from 1978 until now. Applications such as data assimilation or climate change assessment require consistent products. The products referenced earlier have been retrieved using different sensors with different algorithms, and as a result, the time series is not homogeneous. This heterogeneity can be interpreted as a bias and is a problem in the data assimilation process. To avoid this issue, these products need to be processed to correct for any bias or amplitude variation between the data sets.

Many previous studies have developed various methods for the homogenization of time series. Vincent *et al.* [11] developed a method to harmonize temperature time series with gaps. The first step was to determine if the series was homogeneous by comparing its anomalies to those of a reference series. The identification of the gaps and their magnitude was performed by successively fitting a linear model with different magnitude values with the best fit being indicated by the minimum sum of square errors. Homogeneous temperature and precipitation time series were developed by Begert *et al.* [12] using statistical

Manuscript received November 28, 2011; revised May 14, 2012 and October 22, 2012; accepted December 22, 2012. This work was supported in part by Telespazio France and in part by TOSCA.

D. J. Leroux is with the Centre d'Etudes Spatiales de la Biosphere, 31400 Toulouse, France, and also with Telespazio France, 31023 Toulouse Cedex 1, France (e-mail: delphine.j.leroux@gmail.com).

Y. H. Kerr is with the Centre d'Etudes Spatiales de la Biosphere, 31400 Toulouse, France (e-mail: yann.kerr@cesbio.cnes.fr).

E. F. Wood and A. K. Sahoo are with the Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: efwood@princeton.edu; aksahoo2004@gmail.com).

R. Bindlish and T. J. Jackson are with the USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705 USA (e-mail: rajat.bindlish@ars.usda.gov; tom.jackson@ars.usda.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2240691

methods to detect potential inhomogeneity. In that study, a reference time series was necessary in order to detect and compute the magnitude of the shifts. Picard and Fily [13] proposed a method to simulate a homogeneous time series of the cumulative melting surface in Antarctica. Using satellite observations from different sensors and acquisition times was the biggest challenge. Correcting for the effect of the observing time was accomplished in two steps. First, a sinusoidal function with a 24-h periodicity was fitted, and then, an optimal interpolation to refine this first guess model to *force* it to be closer was applied to the observations and to provide very low uncertainty around observation time and larger uncertainty when there is no available observation.

Matching the cumulative density functions (CDFs) of two data sets has been used in several studies to merge time series. Reichle and Koster [14] and Choi and Jacobs [15] merged soil moisture derived from satellite observations with model data, and Li *et al.* [16] corrected the bias of precipitation and temperature products derived from different models. CDF matching was also used as a preliminary step of the assimilation process [17] and to produce long time series of soil moisture [18], [19].

Over the last few years, a new method based on copula functions has been developed. It allows the derivation of bivariate distributions without making the assumptions required when dealing with multivariate frequency distributions, e.g., the same type of marginal distribution for both variables, a joint normal distribution, and independent variables. One of the major advantages of the copula method is that the marginal distributions can be of any form [20]. The first comprehensive treatment of copulas was by Nelsen [21]. He presented methods to construct copulas and discussed the role played by copulas in modeling and dependence. Since then, copulas have been applied in various applications with the majority of the literature dedicated to the financial sector [22], [23]. In the field of hydrology, some applications have emerged. Genest and Favre [24] summarized the existing methods to detect and evaluate the dependence between the data sets through copulas (analytically and graphically) and enumerated the various methods to choose the best copula family and estimate their parameters. Favre *et al.* [25] applied copulas to peak flows and volumes from two watersheds, Salvadori and De Michele [26] to storm and rainfall time series, Dupuis [27] to the volume and duration of low flows of two rivers, Zhang and Singh [28] to rainfall frequency, Serinaldi and Grimaldi [29] to flood and sea frequency, and Laux *et al.* [30] to precipitation data. Gao *et al.* [31] used copulas as a preprocessing step for the assimilation process on soil moisture data.

Joint statistical analysis has already been applied when the sources of the soil moisture measurements come from different observation systems (e.g., AMSR-E surface soil moisture and 10-cm soil moisture from a land surface model [14]). Similarly, joint statistical methods form the basis for data assimilation of satellite soil moisture into land surface models [31]. There are many other studies related to joint probability, including where the variables are physically different but where their statistical relationships are useful (e.g., rainfall storm intensity and storm duration [32]).

The goal of this paper is to estimate for all the AMSR-E period (2003–2010) SMOS-equivalent observations that can be used to develop a statistical representation of SMOS retrieval so that current and future SMOS retrievals can be used in applications like drought monitoring based on percentiles. However, matching 130 am C-/X-band (AMSR-E) observations with 600 am L-band (SMOS) observations presents some issues: 1) The crossing times are different, and rainfalls may occur between the two acquisitions; and 2) the frequencies are different, so the sensing depths are not similar.

The statistical impact of the rainfalls that could occur between 130 am and 600 am is to lower the correlation. However, if the correlation is sufficiently high, a statistical relationship can be established to estimate an equivalent SMOS value from an AMSR-E observation. This high correlation implies that the occurrence of precipitation between the SMOS and AMSR-E overpasses is rare. Moreover, it is well known that soil moisture has a long temporal correlation time scale, so the overpass time differences will have a minimal effect on the analysis.

The impact of the different frequencies between AMSR-E and SMOS is, in most situations, not significant. The higher AMSR-E frequency (10.7 GHz) results in a more superficial emission depth than the SMOS observations, so while the retrieved values may be different, their relative values will be similar (both dry or wet). The correlation between paired observations depends on their relative values (with their individual time series) and not absolute values, and in the case of copula-based joint distributions, the correlation is represented by the Kendall tau whose calculation is based on ranks.

If the two sensing depths were to be reconciled physically, given the soil property variability (spatially and with depth) with different wetting and drying properties, a physical model would introduce significant uncertainty that could be very difficult to estimate afterward. If the SMOS (or AMSR-E) data were adjusted to the AMSR-E (or SMOS) emission depth through data assimilation into a land surface model for example, then the complete record would have to be adjusted with the added uncertainty of the data assimilation step. With any of the suggested adjustments, there is a mismatch with the past or with the future. Only by treating the original data sets and determining the information content between them can a consistent approach be represented.

Data assimilation could, however, deal with the precipitation and the difference in sensing depth issues, but that would imply other uncertainties such as the space–time variability of the precipitation data sets, as well as other meteorological issues. Building a homogeneous time series based on data assimilation into a land surface model can be seen as a competing approach.

In this paper, we show two statistical methods to obtain this homogeneous time series. The satellite data and the four watersheds where the time series are simulated are presented in Section II. The two statistical methods for generating homogeneous time series are presented in Section III which includes the general theory and how to apply them to real data. Simulated time series over the four watersheds are presented in Section IV. Conclusions and perspectives are described in the last section.

II. Regions of Interest and Satellite Data

A. SMOS

With its L-band radiometer, SMOS [4] has been providing soil moisture data for almost three years and global coverage every three days with a 43-km resolution. The satellite is polar orbiting with equator crossing times of 6 am (local solar time (LST), ascending) and 6 pm (LST, descending). The signal at L-band is mainly influenced by the water content at the surface of the soil (around 5 cm).

SMOS acquires brightness temperatures at multiple incidence angles, from 0° to 55° with full polarization. The angular signature is a key element of the retrieval algorithm that provides soil moisture and the vegetation optical thickness through the minimization of a cost function between modeled and acquired brightness temperatures [33], [34]. This estimated soil moisture is referred as the Level 2 product [34] and is AQ12 available on the Icosahedral Snyder Equal Area-4h9 grid [35]. The nodes of this grid are equally spaced at about 15 km. In this paper, the 2010 SMOS Level 2 version 4 products have been used.

Currently, numerous studies are underway on the validation of SMOS soil moisture product with *in situ* measurements and estimates of other sensors and models. Bitar *et al.* [36] used the Soil Climate Analysis Network [37] and the Snowpack Telemetry sites in North America to compare SMOS soil moisture retrievals and ground measurements. That study showed that SMOS soil moisture had a very good dynamic response but tended to underestimate the values. However, the new version of the product (V4) significantly improved the general results. Jackson *et al.* [38] studied SMOS soil moisture and vegetation optical depth over four watersheds in the U.S. They concluded that SMOS almost met the accuracy AQ13 requirement with root-mean-square errors (rmses) of 0.043 and 0.047 m^3/m^3 in the morning and afternoon, respectively, whereas the vegetation optical depth retrievals were not reliable AQ14 yet for use in vegetation analyses. Leroux *et al.* [39] compared SMOS data with other satellite and model output products over the same four watersheds for the year 2010. It showed that SMOS soil moisture data were closer to the ground measurements than the other data sets. Even though the correlation coefficient was not the best, the bias was extremely small.

After the results of the validation activities, the European Center for Medium-Range Weather Forecasts has decided and is now ready to process SMOS data in near real time into their Integrated Forecast System. It is expected to have an impact on the weather forecast at short and medium ranges [40].

B. AMSR-E

The AMSR-E was launched in June 2002 on the Aqua satellite. This radiometer acquires data with a single 55° incidence angle at six different frequencies: 6.9, 10.7, 18.7, 23.8, 36.5, and 89.0 GHz, all dual polarized. The crossing times are respectively 1:30 am (LST, descending) and 1:30 pm (LST, ascending).

There are several soil moisture products available that are based on AMSR-E data. Many studies have already showed

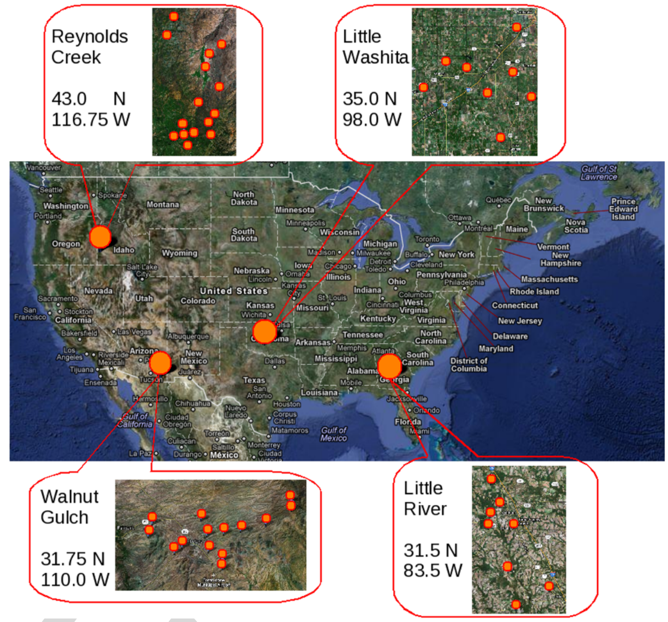


Fig. 1. Map of the four sites: WG, AZ; LW, OK; LR, GA; and RC, ID.

that the NASA product [41] is not able to reproduce low values AQ15 of soil moisture and has low dynamic range [42]–[46]. The soil moisture data produced by the joint collaboration of the Vrije University of Amsterdam and NASA (whereafter called the Land Parameter Retrieval Model (LPRM) [7]) were chosen in this study. AQ16

The LPRM [7] retrieves soil moisture and optical thickness using the C- and X-band AMSR-E channels (combined product) and 36.5 GHz to estimate the surface temperature. This algorithm is based on a microwave radiative transfer model with *a priori* information about soil characteristics. The products are available on a 0.25° × 0.25° grid only for the descending orbit. These data have been quality controlled, and the contaminated estimates due to high topography and extreme weather conditions such as snow have been flagged and not been considered in this study. AQ17

C. Study Areas

Four watersheds located in the United States were selected for this study: Walnut Gulch (WG) in Arizona, Little Washita (LW) in Oklahoma, Little River (LR) in Georgia, and Reynolds Creek (RC) in Idaho (see Fig. 1). They represent different types of climate (from semiarid to humid) and land use patterns [47]. These four watersheds have been used as calibration and validation sites for comparison of AMSR-E satellite product [47] and SMOS product [38], [39].

WG is located in the Southeast Arizona. Most of the watershed is covered by shrubs and grass, which is typical of the region. The annual mean temperature is 17.6 °C (at Tombstone), and the annual mean precipitation is 320 mm (mainly from high intensity convective thunderstorms in the late summer). The uppermost 10 cm of the soil profile contains up to 60% gravel, and the underlying horizons usually contain less than 40% gravel.

TABLE I

AQ18 WATERSHED CHARACTERISTICS AND THE COORDINATES OF THE BOX CONTAINING THE POINTS USED FOR STATISTICS

Watershed	Number of stations	Climate	Annual rainfall (mm)	Topography	Land use	Box for statistics (corners coord.)
Walnut Gulch AZ	14	semi-arid	320	rolling	range	31.3 N - 110.5 W 32.3 N - 109.5 W
Little Washita OK	8	sub-humid	750	rolling	range/wheat	34.4 N - 98.5 W 35.4 N - 97.5 W
Little River GA	8	humid	1200	flat	row crop/forest	31.0 N - 84.0 W 32.0 N - 83.0 W
Reynolds Creek ID	15	semi-arid	500	mountainous	range	34.7 N - 98.7 W 35.7 N - 97.7 W

TABLE II

CORRELATION COEFFICIENTS (R) BETWEEN THE IN SITU MEASUREMENTS AT 130 AM AND 600 AM FOR THE FOUR WATERSHEDS. N IS THE NUMBER OF AVAILABLE DATES, AND CI IS THE 95% CONFIDENCE INTERVAL

WG			LW		
R	N	CI	R	N	CI
0.96	365	[0.95-0.97]	0.97	365	[0.96-0.98]
LR			**RC**		
R	N	CI	R	N	CI
0.95	365	[0.94-0.96]	0.99	328	[0.99-0.99]

LW is located in Southwest Oklahoma in the Southern Great Plains region of the U.S. The climate is subhumid with an average annual rainfall of 750 mm (mainly during the spring and fall seasons). Topography is moderately rolling with a maximum relief of less than 200 m. Land use is dominated by rangeland and pasture (63%).

LR is located in the Southern Georgia near Tifton. With an average annual precipitation of 1200 mm, the climate is humid. The LR watershed is typical of the heavily vegetated slow-moving stream systems in the Coastal Plain region of the U.S. The topography over this watershed is relatively flat. Approximately 40% of the watershed is forest with 40% crops and 15% pasture.

RC is located in a mountainous area of Southwest Idaho. The topography is high with a relief of over 1000 m that results in diverse climates. Soils and vegetations are typical in this part of the Rocky Mountains. The climate is considered as semiarid with an annual precipitation of 500 mm. Approximately 75% of the annual precipitation at high elevation is snow, whereas only 25% is snow at low elevation.

Surface soil moisture and temperature sensors (0–5 cm) have been acquiring data since 2002 for the four watersheds. The data used in this study are the means and standard deviations of the soil moisture and surface temperature acquired every 30 min from 2009 to 2010 (hourly for RC). The averages are based on 14/8/8/15 sensors for WG/LW/LR/RC, respectively, after eliminating sensors with poor and suspicious AQ19 performances. Weighting coefficients have been derived for each sensor with a Thiessen polygon. Table I summarizes the characteristics of each watershed [47].

In order to estimate the effect of the rainfalls that could occur between 130 am and 600 am, the correlation coefficients between the measurements at 130 am and 600 am have been computed for the four watersheds (see Table II and Fig. 2). They range from 0.95 to 0.99, and based on the fact that rainfalls would lower the correlation, we can assess that precipitations that do not affect significantly the analysis.

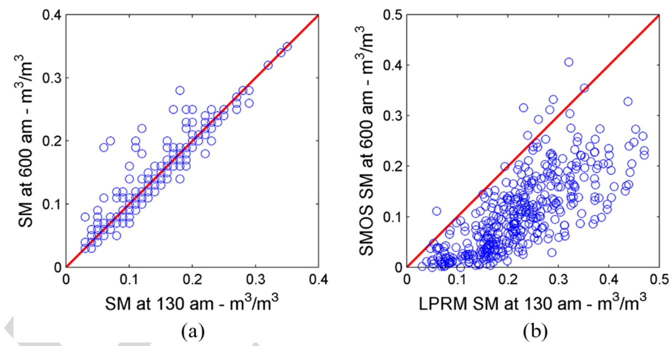


Fig. 2. Comparison between the 130 am and the 600 am soil moisture: *In situ* observations and satellite products for the four watersheds. (a) *In situ* soil moisture at 130 am and 600 am. (b) LPRM (130 am) and SMOS (600 am) soil moisture.

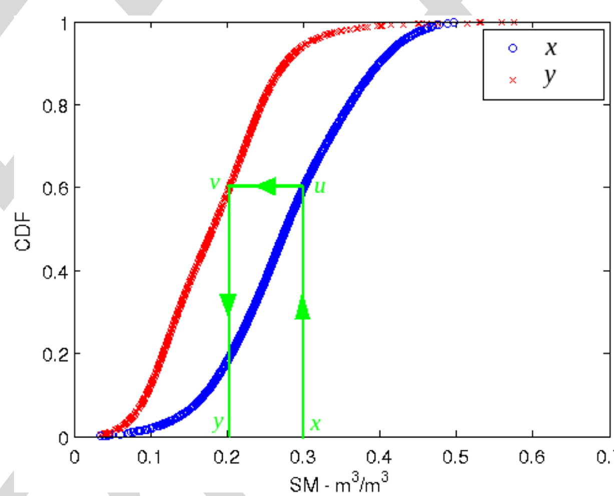


Fig. 3. Principle of CDF matching by setting the probabilities equal. For a given x, find y such that $G_Y(y) = F_X(x)$.

III. TWO STATISTICAL METHODS FOR GENERATING HOMOGENEOUS TIME SERIES

Two statistical methods were used to create a homogeneous time series of soil moisture. CDF matching has been widely used in previous studies to merge time series [14], [15], [18], [19], whereas copulas have just started to be used recently for environmental purposes.

A. CDF Matching

The CDF is the probability that a random variable X takes a value less than or equal to a given number x

$$F_X(x) = \Pr[X \leq x] \quad (1)$$

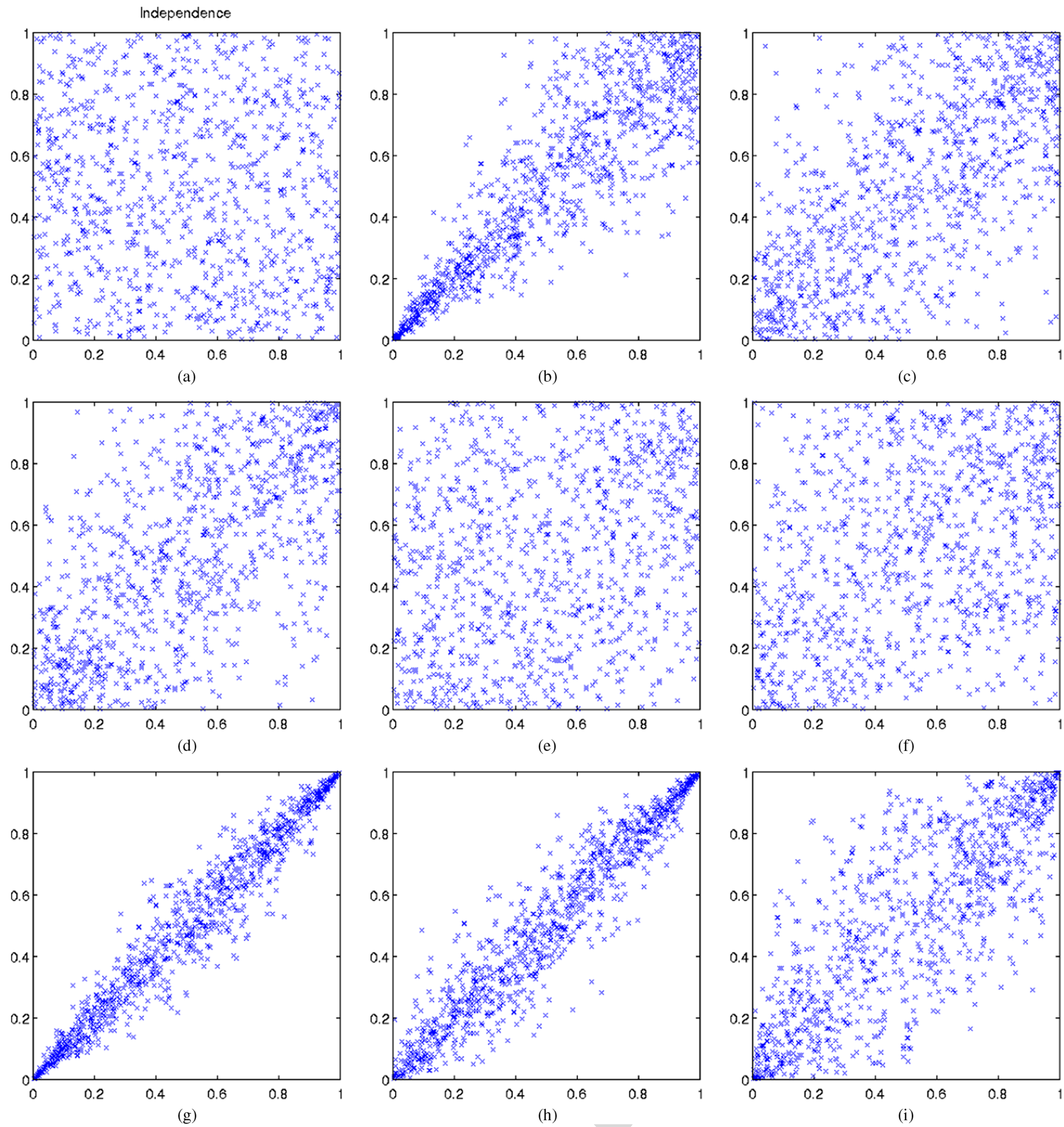


Fig. 4. Representations of the nine copulas showing their characteristics in the form of the point cloud (x-axis: CDF of the first data set; y-axis: CDF of the second data set).

where F_X is the CDF of the random variable X. If two time series are considered, the CDF matching consists of matching the CDF of each data set by setting their probabilities equal (see Fig. 3). The following approach has been applied here to the soil moisture data.

1) Compute the CDF of both data sets X and Y: F_X and G_Y.
2) Given a value x of X, find y such that $G_Y(y) = F_X(x)$.

However, the assumption that the probabilities $F_X(x)$ and $G_Y(y)$ are equal is never confirmed, and most of the time, they are scattered like in Fig. 4. The copula method models this dependence between the probabilities.

For the rest of this paper, we use the variable u to represent $F_X(x)$ and v for $G_Y(y)$. U and V are data sets, whereas u and v are values of these data sets.

B. Copulas

The copula theory is a very useful and powerful tool to model the dependence structure between two sets of random variables.

TABLE III
NINE COPULAS TESTED IN THE STUDY: DEFINITION, PARAMETER RANGE, AND FAMILY

Copula	$C_\theta(u,v)$	$\theta \in$	Family
Independent	$u \cdot v$	-	-
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	Archimedean
Frank	$\frac{-1}{\theta} \ln \left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$(-\infty, \infty)/0$	Archimedean
Gumbel	$\exp\left(-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{1/\theta}\right)$	$[1, \infty)$	Archimedean
FGM	$uv + \theta uv(1-u)(1-v)$	$[-1, 1]$	Elliptical
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$[-1, 1]$	Archimedean
Arch12	$\left(1 + \left((u^{-1}-1)^\theta + (v^{-1}-1)^\theta\right)^{1/\theta}\right)^{-1}$	$[1, \infty)$	Archimedean
Arch14	$\left(1 + \left((u^{-1/\theta}-1)^\theta + (v^{-1/\theta}-1)^\theta\right)^{1/\theta}\right)^{-1/\theta}$	$[1, \infty)$	Archimedean
Gaussian	$\frac{\int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \exp\left(\frac{2\theta s\omega - s^2 - \omega^2}{2(1-\theta^2)}\right)}{2\pi\sqrt{1-\theta^2}}$	$[-1, 1]$	Elliptical

Like the CDF matching, copulas separate the marginal behavior of variables from the dependence structure by using distribution functions. Instead of setting the probabilities u and v equal, the variables U and V are compared and analyzed. The copula function binds the two variables together.

There are many families of copulas which exhibit very different properties. The form of the scatter of U and V is controlled by the family choice, and the width of the tail of this scatter is controlled by the single parameter θ. Most of the definitions that follow in this section are based on [21].

1) General Theory: A copula is a function that generates a multivariate cumulative distribution function from 1-D marginal CDFs. Given two random variables, X and Y, with marginal CDFs F_X and G_Y, then, Sklar's theorem states

$$H_{XY}(x,y) = C_{XY}\left(F_X(x), G_Y(y)\right) = \Pr[X \le x, Y \le y] \quad (2)$$

where H_{XY} is the joint CDF of X and Y and C_{XY} is the associated copula function. It is then possible to derive conditional distributions, $H_{XY}(y|x)$, i.e., the joint CDF knowing x. Let $u = F_X(x)$ and $v = G_Y(y)$. Then, $H_{XY}(y|x)$ can be derived by

$$C_{V|U} = \frac{\partial C(u,v)}{\partial u}. \quad (3)$$

Schweizer and Wolff [48] established that the copula function accounts for all the dependence between the two variables. They demonstrated that transformations of the variables X and Y do not affect their associated variables. Thus, the way that X and Y evolve together is captured by the copula, regardless of the scale in which each variable is measured.

2) Some Copula Families: The product copula corresponds to the independence between X and Y

$$C(u,v) = u \cdot v. \quad (4)$$

A copula of the Archimedean family takes the following form:

$$C(u,v) = \phi^{-1}\left(\phi(u) + \phi(v)\right) \quad (5)$$

where ϕ is the generator function that goes from [0, 1] to $(0, \infty)$. It satisfies three conditions: $\phi(1) = 0$, ϕ strictly decreasing, and ϕ convex.

Elliptical copulas have distributions with elliptic contours. The main advantage of elliptical distributions is that the level of correlation between the variables U and V can be specified. The disadvantages are that elliptical copulas do not have closed-form expressions and are restricted to have radial symmetry.

In this paper, nine copulas were used: the product copula, Clayton, Frank, Gumbel, Farlie–Gumbel–Moregenstern (FGM), Ali–Mikhail–Haq, Arch12 (the 12th copula presented in [21]), Arch14 (the 14th copula presented in [21]), and the Gaussian copula. The nine copulas are described in Table III and Fig. 4 and have their own characteristics.

1) Clayton: Strong left tail dependence and relatively weak right tail dependence (i.e., u and v are strongly linked for low values, whereas they are not for high values).
2) Frank: Dependence is symmetric in both tails, weak in both tails, and stronger in the center of the distribution.
3) Gumbel: Strong right tail dependence and relatively weak left tail dependence (the opposite of Clayton).
4) FGM: Useful when the dependence between U and V is modest in amplitude.
5) Gaussian: Flexible as it allows for positive and negative dependences.

Hafner and Reznikova [23] and Wang and Pham [49] developed a method that includes the time into the copula formula to create a dynamic copula evolving with time. In this paper, time was not included, but the year 2010 was divided into four seasons as different statistical behaviors were expected: December–January–February, March–April–May (MAM), June–July–August (JJA), and September–October–November (SON).

3) How to Select a Family: Since copulas separate marginal distributions from dependence structures, the appropriate copula for a particular application is the one that best captures the dependence features of the data [22]. Dupuis [27] examined the effects of model misspecification and highlighted the dangers of improper copula selection. Genest and Rivest [50] proposed a method to select the most appropriate copula, but this method is only relevant for Archimedean copulas. Other methods were developed to compare any type of copulas [51]–[54]. Genest *et al.* [55] and Berg [54] compared some of them and concluded that there was no universal test and that some procedures performed better in some situations but never in all the situations.

The method proposed by Huard *et al.* [56] is based on a Bayesian approach where any type of copula can be tested. It does not perform perfectly well in all the situations (with small correlation coefficients or with small sample size) but has the advantage to be a very fast method. This method was chosen in this study to select the copula that provides the best fit to the data.

4) Method Used for Simulations: The key to generating simulations from a copula is to understand that a copula is a joint distribution and that it obeys to the same rules. A conditional copula $C_{V|U}(u,v)$ is the probability that the random variable V is less than or equal to a value v knowing that the random variable U is equal to a value u

$$C_{V|U}(u,v) = \Pr[V \leq v \mid U = u] = t \sim \mathcal{U}(0,1). \quad (6)$$

Simulating a uniform variable t is necessary in order to generate simulations from a copula. To retrieve $V|U$, the function $C_{V|U}$ needs to be inverted such that $v = C^{-1}_{V|U}(t)$, or the equation $C_{V|U}(v) = t$ needs to be solved numerically. For each value of t, a value for v is retrieved. The following approach was used here to simulate data with the copulas.

1) Compute F_X and G_Y from the two original data sets X and Y with (1).
2) Choose the appropriate copula C by applying Huard's method and fitting the parameter θ to the original data.
3) Derive the conditional copula $C_{V|U}$ with (3).
4) Generate 1000 simulations $t \sim \mathcal{U}(0,1)$.
5) Compute v with $v = C^{-1}_{V|U}(t)$ and y with $y = G^{-1}_Y(v)$.
6) The mean and standard deviation from the 1000 simulations can be computed.

IV. Methodology

For the CDF matching and the copula methods, 2010 data were used for calibration. The CDFs of SMOS and LPRM were calculated for the 2010 data sets. The two algorithms were then applied to the data from previous years. It should be noted that the consequence of using 2010 as a calibration year is that only the soil moisture range from 2010 is taken into account. If an extreme event occurred in the previous years, it might not be well described with these methods as they are only based on statistics and not on physical models. By looking at the *in situ* soil moisture time series in Fig. 7, 2010 did not have enough wet values over LR to estimate correctly the strong rainfalls of 2004, 2005, and 2009, not enough wet values over LW for rainfalls in 2007 and not enough dry values as well for 2003 and 2006, and again not enough dry values over RC for all the previous years. AQ23

The two methods were applied to data contained in a $1^\circ \times 1^\circ$ box around each watershed in order to have enough points for computing reliable statistics. The coordinates of each box are indicated in Table I. Only the satellite morning overpasses were selected for this study (6:00 am for SMOS and 1:30 am for AMSR-E, LST) since LPRM retrievals were only available for this overpass.

The 2010 calibration year was divided into four seasons: December–January–February, MAM, JJA, and SON. This subdivision was done in order to better capture the seasonal dynamic that can be very different depending on the time of the year, particularly in vegetated areas. However, not enough points were available during the winter period (December–January–February) to compute reliable statistics, so no estimation was performed for this season.

When comparing either two different remote sensing products or *in situ* data with remote sensing products, there is the issue of the scale effect, as the products may have significantly different spatial resolutions. Moreover, the spatial variability varies with the seasons and the heterogeneity. So as to reduce the problem, we used in this study averaged *in situ* data sets (8 to 15 stations that were several miles away) which were especially produced to be representative of 50-km spatial resolution or so [47]. Also, statistics were applied to all the points contained in a $1^\circ \times 1^\circ$ box (more than 50 grid points).

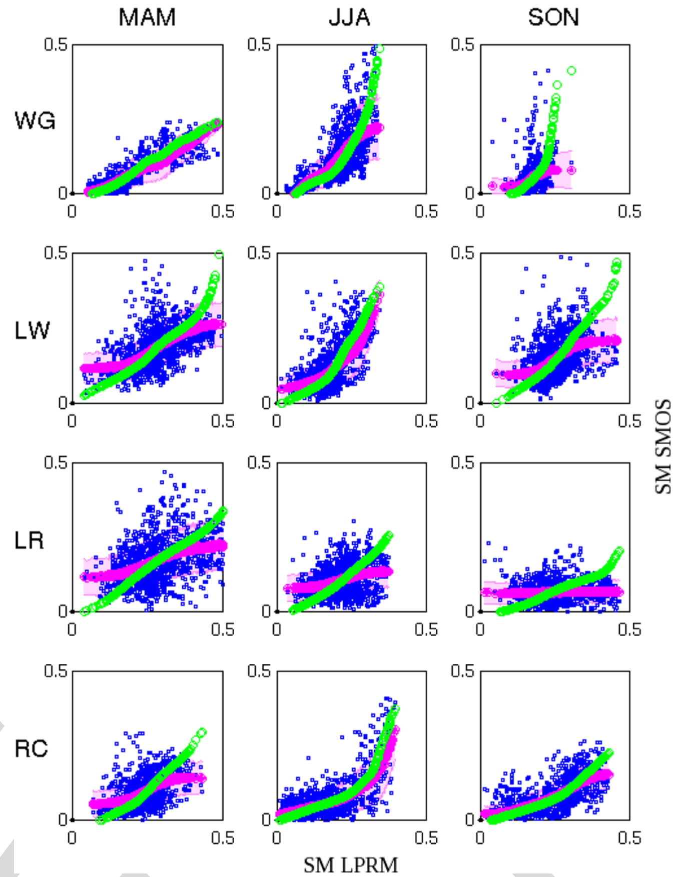


Fig. 5. Discrepancies in the simulations of soil moisture between CDF matching and copulas in 2010. Original soil moisture LPRM data are represented by blue points, and simulated data with CDF matching and copulas are in green and red, respectively. The standard deviation of the copula simulations is represented in shadowed red. Each row corresponds to a site, and each column corresponds to a season. x-axis: LPRM soil moisture. y-axis: SMOS soil moisture. AQ24

V. Generated Homogeneous Time Series

The year 2010 was used to compute the CDFs of each data set (SMOS and LPRM) for both methods and the joint CDF based on fitting and selecting copula functions as described previously. The soil moisture data were estimated using

AQ25

TABLE IV
STATISTICAL RESULTS OF THE SIMULATIONS FROM COPULAS AND CDF MATCHING. THE SIMULATIONS WERE COMPARED TO GROUND MEASUREMENTS OVER 2010 DIVIDED INTO FOUR SEASONS: MAM, JJA, SON, BUT NOT ENOUGH DATA AVAILABLE FOR WINTER SEASON. THE BEST RESULTS ARE WRITTEN IN BOLD, AND RMSES ARE IN m^3/m^3

		SMOS		LPRM		Copula method			CDF matching		# points
		R	RMSE	R	RMSE	Fam(θ)	R	RMSE	R	RMSE	
WG	MAM	0.80	0.032	0.82	0.125	Gumbel (2.18)	**0.89**	**0.020**	0.87	0.031	43
	JJA	0.86	0.053	0.86	0.126	Clayton(2.63)	0.76	**0.076**	**0.81**	0.090	45
	SON	0.64	0.029	0.79	0.133	Frank (3.13)	**0.64**	**0.012**	0.53	0.029	42
	total	0.84	0.040	0.79	0.139	-	0.79	**0.043**	**0.82**	0.054	159
LW	MAM	0.70	0.068	0.48	0.166	Frank (4.40)	0.55	**0.057**	**0.57**	0.075	44
	JJA	0.85	0.037	0.58	0.085	Gumbel (1.66)	**0.77**	**0.042**	0.76	0.050	44
	SON	0.80	0.041	0.80	0.122	Frank (3.61)	**0.75**	**0.023**	0.72	0.048	46
	total	0.78	0.049	0.59	0.148	-	**0.71**	**0.043**	**0.71**	0.059	162
LR	MAM	0.77	0.080	0.54	0.175	Frank (2.82)	**0.59**	**0.063**	0.58	0.067	39
	JJA	0.57	0.053	0.67	0.131	Frank (2.00)	0.65	0.034	**0.66**	**0.033**	40
	SON	0.59	0.032	0.37	0.174	FGM (0.31)	**0.17**	**0.033**	0.16	0.037	39
	total	0.74	0.060	0.65	0.178	-	0.51	**0.045**	**0.59**	0.048	147
RC	MAM	0.14	0.097	0.11	0.096	Frank (3.10)	0.26	**0.089**	**0.27**	0.105	47
	JJA	0.63	0.055	0.81	0.070	Gumbel (1.81)	**0.84**	**0.047**	0.83	0.052	42
	SON	0.14	0.070	0.52	0.144	Frank (6.30)	**0.34**	**0.056**	0.29	0.066	39
	total	0.55	0.081	0.73	0.099	-	**0.80**	**0.059**	0.70	0.067	142

the conditional distribution (conditional on LPRM retrievals). While the copula procedure has the potential to generate an ensemble of SMOS-like soil moisture estimates, given the LPRM estimated soil moisture, we only use the mean estimate. The ensembles could be used to provide uncertainty estimates. It should be noted that CDF matching can only provide a single SMOS estimate. The resulting time series will result in a statistically homogeneous time series under the assumption that 2010 LPRM retrievals and the underlying AMSR-E brightness temperatures are temporally consistent. The resulting SMOS-like estimated soil moisture is then compared to ground measurements.

A. Calibration Year 2010 and Comparison With Ground Measurements

2010 is the year with both SMOS data and LPRM data. CDFs were computed for both variables. CDF matching and copula methods were then applied, and these produced different SMOS-like estimates. In Fig. 5, the original data (SMOS and LPRM) are represented by the blue point cloud, CDF matching and copula estimates are in green and red colors, respectively, and standard deviations from copula simulations are in red AQ26 shadows. This standard deviation can be interpreted as the uncertainty associated to the copula simulations, which can be not produced by CDF matching estimation.

Over WG in the MAM season, there was no obvious difference between the two simulation methods. However, in the JJA and SON seasons, there were differences for the high values of soil moisture: The CDF matching method produced higher simulated values than the copula method. Similar behavior can also be seen for all seasons in the other three sites, i.e., LW, LR, and RC. Discrepancies can also be observed for small values of soil moisture over LW, LR, and RC (MAM) where copulas generated higher values of soil moisture.

Standard deviations of soil moisture simulations from copulas were also computed (see Fig. 5). This standard deviation is directly related to the width of the tail of the chosen copula which is controlled by the θ parameter. A high value of the standard deviation corresponds to a large tail, meaning that the two variables are weakly linked to each other, whereas a small value corresponds to a strong link. The differences in the simulations can also be observed in the 2010 time series (see Table IV and Fig. 6). Compared to the original LPRM data, the estimated soil moisture was close to the SMOS level and comparable to the ground measurements. The bias between LPRM and SMOS was corrected by both methods.

Over WG, CDF matching and copula simulations were not very different except in the summer season when the CDF matching simulations were higher than the copulas. Considering the entire year, both simulation methods improved the original statistics from the LPRM data set. The correlation coefficient did not change significantly ($R = 0.79$ for LPRM and $R = 0.79/0.82$ for copulas/CDF matching), but the rmse was highly improved going from 0.139 m^3/m^3 (original LPRM data) to 0.054 m^3/m^3 with CDF matching and 0.043 m^3/m^3 with copula, which represents an improvement of a factor of 3.

Over LW, simulations responded very well to the successive rain events throughout the year and exhibited a pattern of decrease following a rain event. The first two months (March–April) exhibited more noisy simulations, and the statistics were impacted by this behavior ($R = 0.55/0.57$ and rmse $= 0.057/0.075$ m^3/m^3 for copulas/CDF matching). The other two seasons gave good results in terms of statistics. For the entire year, the R value was highly improved ($R = 0.59$ for LPRM and $R = 0.71/0.71$ for copulas/CDF matching), and the rmse was reduced by a factor of 3 (rmse $= 0.148$ m^3/m^3 for LPRM and rmse $= 0.043/0.059$ m^3/m^3 for copulas/CDF matching).

The LR watershed is the site with the highest rainfall frequency (events of small amplitude). The successive rainfall events were not well captured by the simulations, particularly during the fall season when both simulations exhibited only small variations, which resulted in very poor statistics ($R = 0.17/0.16$ for copulas/CDF matching). Unfortunately, even if the rain events were captured by the original data sets, none was captured by both data sets at the same time, so only the nonraining periods were taken into account by the statistics. Therefore, the simulations can only be representative of the dry periods. It should be noted that the statistics of LPRM were

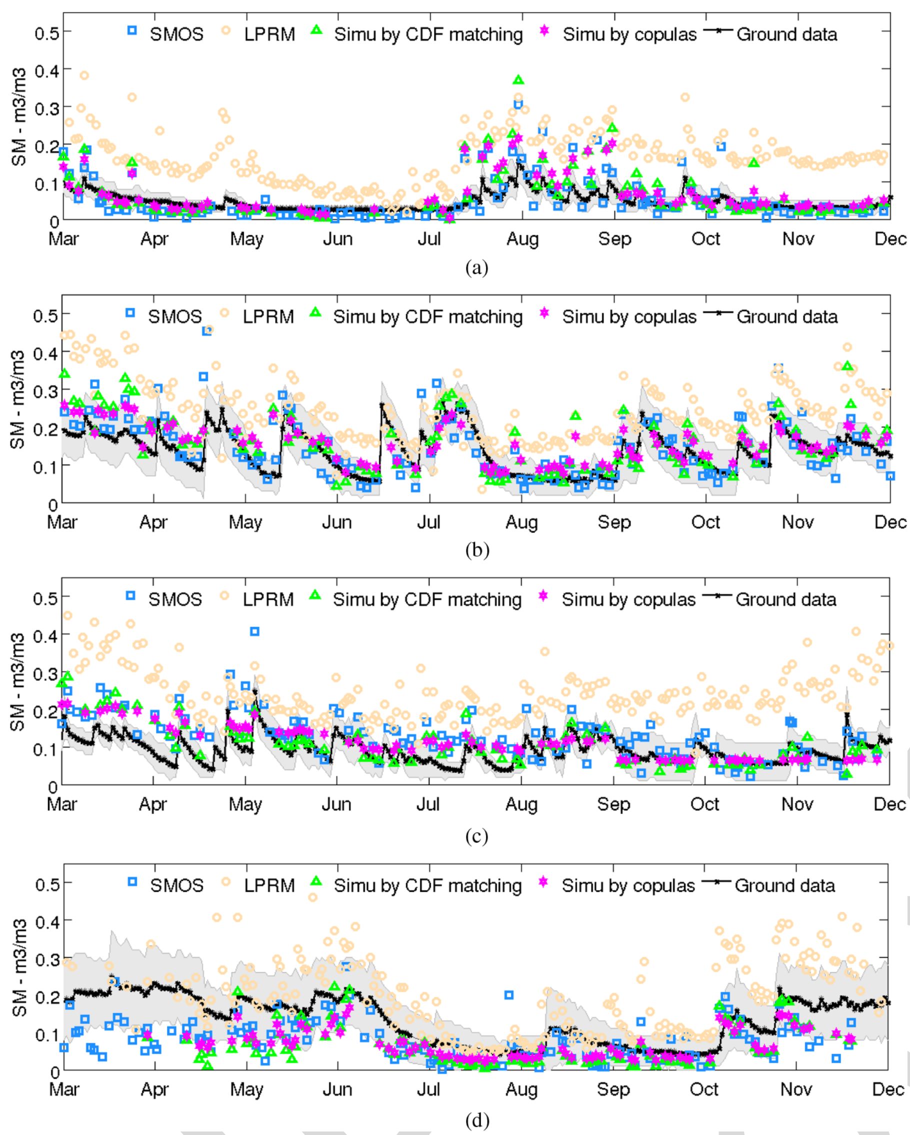


Fig. 6. Simulations for 2010: SMOS, LPRM, simulated soil moisture data from CDF matching and copulas, and ground measurements over the four watersheds. Since the *in situ* data are the mean of several ground measurements, their standard deviations are represented in gray shadows showing the spatial variability. (a) WG. (b) LW. (c) LR. (d) RC.

already not good during this season ($R = 0.37$ and rmse $=$ $0.174\ \mathrm{m}^3/\mathrm{m}^3$). During the spring season, SMOS overestimated the *in situ* soil moisture measurements, so as a result, the copulas and CDF matching estimates overestimated the *in situ* measurements as well.

RC is located in a mountainous region and is subject to frequent snow and frozen soil events. The satellite-based soil moisture was not comparable to the ground measurements until late May. After this winter period, the simulations captured accurately the soil moisture evolution and improved the original statistics and especially the rmse ($0.099\ \mathrm{m}^3/\mathrm{m}^3$ for LPRM and $0.059/0.067\ \mathrm{m}^3/\mathrm{m}^3$ for copulas/CDF matching).

B. *Times Series 2003–2010 and Comparison With Ground Measurements*

Soil moisture from 2003 to 2010 was simulated from the LPRM retrievals (2003–2010) using the copulas and CDF matching relationships developed for 2010. Fig. 7 and Table V show the entire time series and the associated statistics (R and rmse) between the original data, CDF matching simulations, copula simulations, and ground measurements.

WG is the driest site and did not have a lot of rain events. These rain events were well described by the simulated soil moisture even though they were sometimes largely overestimated, particularly by CDF matching simulations. Artifacts at the extremities of the seasons can be seen at the beginning of 2006 and 2008. The correlation coefficient was improved using the CDF matching for each year, whereas the errors were reduced by a factor larger than 2 with the copulas.

The overestimation of the soil moisture after the rain events with CDF matching can be found as well over LW, but the temporal evolution was well captured by both methods. For this watershed, CDF matching overestimated the high soil moisture values and underestimated the low values. CDF matching produced soil moisture with a higher dynamic range than copulas.

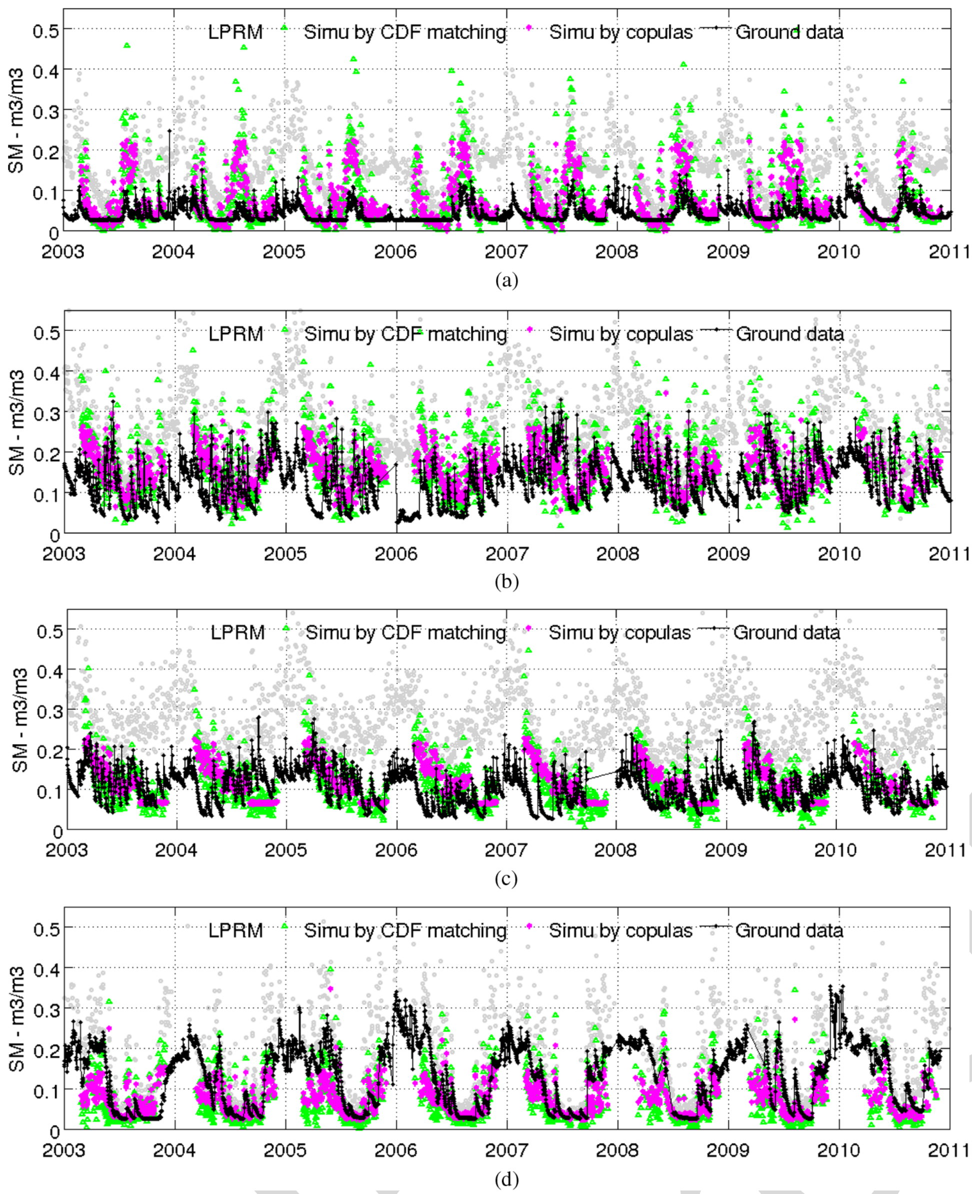


Fig. 7. Simulated time series from 2003 to 2010 with ground measurements for the four watersheds. (a) WG. (b) LW. (c) LR. (d) RC.

This was reflected in the total rmse value (0.079 m^3/m^3), whereas the rmse of the copula simulations was of 0.066 m^3/m^3 (original LPRM rmse: 0.160 $\mathrm{m}^3/\mathrm{m}^3$).

LR is the site with the largest number of rain events, and as mentioned in the previous section, this high rain frequency was not properly captured during the fall season of 2010; this can be seen as well in the entire time series where all the copulas and CDF matching estimates were flat during fall seasons. Moreover, since SMOS was overestimating the soil moisture during the spring season of 2010, both statistical estimates had this behavior. Even though the tendency of the simulations was correct, the dynamic behavior was not well represented, which resulted in a very poor correlation coefficient (negative values in 2004 and 2007).

RC is a very complicated site because of the frequent snow and frozen soil events occurring during half of the year. However, statistical results were improved for the entire year with copula simulations (rmse $= 0.099\ \mathrm{m}^3/\mathrm{m}^3$ for LPRM and rmse $= 0.056/0.062\ \mathrm{m}^3/\mathrm{m}^3$ for copulas/CDF matching).

VI. Conclusion and Perspectives

The main goal of this study was to propose a new method to generate a long homogeneous time series (2003–2010) of soil moisture from two overlapping time series.

For that purpose, two statistical tools, the CDF matching and the copulas, were tested over four watersheds in the U.S. By using CDF matching, the assumption that the two studied data sets are ranked in the same way is made, which the copulas do not require. The two analyzed data sets (SMOS and LPRM) were jointly available only for 2010, so data from 2010 were used to estimate the CDFs that are used as references to estimate SMOS soil moisture for previous years. The novelty of the approach is its application: establishing the statistical relationship between

TABLE V

STATISTICAL RESULTS FROM THE COMPARISON BETWEEN THE SIMULATED TIME SERIES OF SOIL MOISTURE FROM 2003 TO 2010. ORIGINAL SOIL MOISTURE TIMES ARE REPRESENTED BY LPRM. THE BEST RESULTS ARE INDICATED IN BOLD, AND THE RMSE ARE IN m^3/m^3. (a) WG. (b) LW. (c) LR. (d) RC

(a)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.070	0.76	0.82	0.66	0.81	0.68	0.65	0.79	0.73
	RMSE	0.129	0.141	0.146	0.133	0.147	0.138	0.129	0.139	0.138
Copula	R	0.62	0.55	0.82	0.64	0.81	**0.75**	0.76	0.79	0.69
	RMSE	**0.059**	**0.059**	**0.059**	**0.060**	**0.054**	**0.053**	**0.060**	**0.043**	**0.057**
CDF m.	R	**0.73**	**0.62**	**0.88**	**0.72**	**0.89**	**0.75**	**0.79**	**0.82**	**0.74**
	RMSE	0.070	0.074	0.071	0.073	0.067	0.067	0.077	0.054	0.071

(b)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.56	0.71	0.48	0.67	0.32	0.42	0.52	0.58	0.55
	RMSE	0.163	0.149	0.187	0.149	0.173	0.158	0.149	0.149	0.160
Copula	R	0.56	0.47	0.19	0.62	0.41	**0.64**	**0.58**	**0.71**	0.47
	RMSE	**0.071**	**0.064**	**0.088**	**0.077**	**0.060**	**0.056**	**0.051**	**0.044**	**0.066**
CDF m.	R	**0.59**	**0.60**	**0.34**	**0.63**	**0.49**	0.61	0.53	**0.71**	**0.51**
	RMSE	0.083	0.070	0.101	0.092	0.069	0.076	0.069	0.059	0.079

(c)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.51	0.60	0.46	0.75	0.64	0.70	0.49	0.65	0.58
	RMSE	0.171	0.148	0.181	0.185	0.180	0.166	0.187	0.178	0.174
Copula	R	0.54	-0.48	**0.73**	0.01	-0.14	0.20	0.43	0.51	0.19
	RMSE	**0.042**	**0.079**	**0.036**	**0.069**	**0.081**	0.054	**0.047**	**0.045**	**0.059**
CDF m.	R	**0.68**	**-0.16**	0.72	**0.28**	**0.18**	**0.50**	**0.55**	**0.59**	**0.37**
	RMSE	0.044	0.080	0.042	0.070	0.085	**0.050**	0.048	0.048	0.061

(d)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.78	0.76	0.74	0.80	0.84	0.69	0.78	0.73	0.77
	RMSE	0.093	0.085	0.110	0.099	0.102	0.106	0.099	0.099	0.099
Copula	R	**0.53**	**0.78**	**0.70**	**0.68**	**0.72**	0.75	**0.72**	**0.80**	**0.69**
	RMSE	**0.065**	**0.045**	**0.065**	**0.060**	**0.051**	**0.047**	**0.052**	**0.059**	**0.056**
CDF m.	R	0.42	0.69	0.65	0.63	0.70	**0.65**	0.71	0.70	0.63
	RMSE	0.073	0.051	0.070	0.063	0.055	0.056	0.056	0.067	0.062

AMSR-E and SMOS retrieved soil moisture values and using this relationship to estimate the *equivalent* SMOS value for the AMSR-E period prior to the SMOS launch.

The first analysis of these simulations over 2010 showed that the simulated data sets were very similar to the SMOS estimates and reproduced SMOS behavior accurately except over the LR watershed where numerous rain events occurred. This high rainfall frequency was interpreted statistically as noise, and hence, the simulations did not describe the soil moisture evolution over this site very well. RC was also a very complicated site due to the local topography and seasonal climate conditions. Soil moisture derived from satellite observations was not able to accurately reproduce the dynamics as found in the *in situ* data, and as a result, the simulated soil moisture did not either. However, the total rmse for the simulated soil moisture from copulas was reduced by a factor of almost 2. The WG and LW sites were well represented by the simulations, and copulas improved the error by a factor of 3, whereas CDF matching improved the correlation.

The time series of soil moisture were estimated from 2003 to 2010 and were compared to *in situ* measurements at all four watersheds. Since simulated soil moisture data in 2010 over the LR watershed had very little dynamic range, they remained the same for the entire time series and showed very poor statistical results. Even though the rmse values were improved by a factor of 3, the total correlation was not good. For the three other sites, the correlation coefficient was a bit degraded compared to the original LPRM data, but the rmse was highly improved with copulas by a factor of 2 to 3. In general, CDF matching gave better results in terms of correlation, and copulas gave better results in terms of errors compared to the ground measurements.

As a more general conclusion, CDF matching gives good results but does not take into account the structure of the dependence between the two data sets, whereas the copulas allow to model this structure. Through the choice of the family and the parameter θ (which controls the width of the tail of the scatter), it is possible to model all kinds of structures, from the perfect dependence (CDF matching), right or left dependence, to complete independence. This is why copulas produce better results for the extreme values (very low and very high values) than CDF matching. Copulas can also estimate the uncertainty of the soil moisture simulations given the LPRM value and can be seen as a quality information in the simulation process. However, the copula method is time consuming. It is quick to choose the copula family and its associated parameter as it is based on a Bayesian approach; however, it is very time consuming to generate the 1000 simulations, particularly if the chosen copula does not have an analytic inversion form. In the latter case, 1000 equations need to be resolved numerically.

Nevertheless, these simulations represent an advantage since it is possible to compute a mean and a standard deviation. The limitations are the same for both methods and even for any general statistical methods using a specific year as a reference: Only the variable range of this particular year can be well represented. Therefore, if an event in a previous year occurs and is out of the range found in the specific year of reference (such as drought or flood events), then that event will not be well represented in the simulated results.

In order to improve this methodology, applying a moving window of three months would provide more accurate results instead of dividing the year into four seasons. This would also avoid the artifacts and gaps generally noticed at the transition between the seasons. Another solution would be to introduce the time in the copulas, but the level of complexity in the copula manipulation would increase as well.

In this paper, the attempt to build a homogeneous soil moisture time series has been based on statistical methods only. Of course, other methods exist to reconcile different sensor acquisitions, and because SMOS and AMSR-E do not operate at the same frequencies and not at the same crossing times, using physical models to tackle these discrepancies is an alternative to statistical methods. Moreover, matching observations acquired at 130 am and 600 am can trigger some questions, particularly regarding the precipitations that could occur in between. The present study is a first step toward a unified and homogeneous soil moisture time series, and mixing physical and statistical models to do so would be a breakthrough for climate studies.

The next step of this study is to build a homogeneous time series of soil moisture at the global scale. Hence, the results of this study will be extended in the future to build a global map of the copula family choice and to study if there exists any relationship between the chosen copulas and the soil characteristics or land use data. This would allow us to derive soil moisture time series from LPRM data within SMOS soil moisture range over the entire globe.

Acknowledgment

The authors would like to thank the USDA research team for providing the *in situ* data and D. Huard from McGill University for the helpful inputs on the copula method. USDA is an Equal Employment Opportunity employer.

References

[1] "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC," World Meteorological Org., Geneva, Switzerland, Tech. Rep., 2010.

[2] M. Drusch, "Initializing numerical weather predictions models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *J. Geophys. Res.*, vol. 113, no. D3, p. D03 102, Feb. 2007.

[3] H. Douville and F. Chauvin, "Relevance of soil moisture for seasonal climate predictions: A preliminary study," *Climate Dyn.*, vol. 16, no. 10/11, pp. 719–736, Oct. 2000.

[4] Y. Kerr, P. Waldteufel, J. Wigneron, S. Delwart, F. Cabot, J. Boutin, M. Escorihuela, J. Font, N. Reul, C. Gruhier, S. Juglea, M. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.

[5] D. Entekhabi, E. Njoku, P. O'Neill, K. Kellogg, W. Crow, W. Edelstein, J. Entin, S. Goodman, T. Jackson, J. Johnson, J. Kimball, J. Piepmeier, R. Koster, N. Martin, K. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. Shi, M. Spencer, S. Thurman, L. Tsang, and J. Zyl, "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[6] Y. Kerr, P. Waldteufel, J. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.

[7] M. Owe, R. de Jeu, and J. Walker, "A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1643–1654, Aug. 2001.

[8] E. Njoku, T. Jackson, V. Lakshmi, T. Chan, and S. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215–229, Feb. 2003.

[9] L. Li, P. Gaiser, B. Gao, R. Bevilacqua, T. Jackson, E. Njoku, C. Rüdiger, J. Calvet, and R. Bindlish, "Windsat global soil moisture retrieval and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2224–2241, May 2010.

[10] V. Naeimi, K. Scipal, Z. Bartalis, and S. H. W. Wagner, "An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 1999–2013, Jul. 2009.

[11] L. Vincent, X. Zhang, B. Bonsal, and W. Hogg, "Homogeneization of daily temperatures over Canada," *J. Climate*, vol. 15, no. 11, pp. 1322–1334, Jun. 2002.

[12] M. Begert, T. Schlegel, and W. Kirchhofer, "Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000," *Int. J. Climatol.*, vol. 25, no. 1, pp. 65–80, Jan. 2005.

[13] G. Picard and M. Fily, "Surface melting observations in Antarctica by microwave radiometers: Correcting 26-year time series from changes in acquisition hours," *Remote Sens. Environ.*, vol. 104, no. 3, pp. 325–336, Oct. 2006.

[14] R. Reichle and R. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, no. 19, p. L19 501, Oct. 2004.

[15] M. Choi and J. Jacobs, "Temporal variability corrections for advanced microwave scanning radiometer E (AMSR-E) surface soil moisture: Case study in little river region, Georgia, U.S.," *Sensors*, vol. 8, no. 4, pp. 2617–2627, Apr. 2008.

[16] H. Li, J. Sheffield, and E. Wood, "Bias correction of monthly precipitation and temperature fields from IPCC AR4 models using equidistant quantile matching," *J. Geophys. Res., Atmosp.*, vol. 115, no. D10, p. D10 101, May 2010.

[17] M. Drusch, E. Wood, and H. Gao, "Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture," *Geophys. Res. Lett.*, vol. 32, no. 15, p. L15 403, Aug. 2005.

[18] Y. Liu, A. van Dijk, R. de Jeu, and T. Holmes, "An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia," *Water Resour. Res.*, vol. 45, no. 7, p. W07 405, Jul. 2009.

[19] Y. Liu, R. Parinussa, W. Dorigo, R. D. Jeu, W. Wagner, A. V. Dijk, M. McCabe, and J. Evans, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 2, pp. 425–436, Feb. 2011.

[20] V. Singh and W. Strupczewski, "Editorial," *J. Hydrol. Eng.*, vol. 12, no. 4, p. 345, Jul. 2007.

[21] R. Nelsen, "An introduction to copulas," in *Springer Series in Statistics*. New York, NY, USA: Springer-Verlag, 1998.

[22] P. Trivedi and D. Zimmer, *Copula Modeling: An Introduction for Practitioners*, vol. 1, *Foundations and Trends in Econometrics*. Hanover, MA, USA: Now Publ. Inc., 2005, pp. 1–111.

[23] C. Hafner and O. Reznikova, "Efficient estimation of a semiparametric synamic copula model," *Comput. Stat. Data Anal.*, vol. 54, no. 11, pp. 2609–2627, Nov. 2010.

[24] C. Genest and A. Favre, "Everything you always wanted to know about copula modeling but were afraid to ask," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 347–368, Jul./Aug. 2007.

[25] A. Favre, S. E. Adlouni, L. Perreault, N. Thiémonge, and B. Bobée, "Multivariate hydrological frequency analysis using copulas," *Water Resour. Res.*, vol. 40, no. 1, p. W01 101, Jan. 2004.

[26] G. Salvadori and C. de Michele, "On the use of copulas in hydrology: Theory and practice," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 369–380, Jul. 2007.

[27] D. Dupuis, "Using Copulas in hydrology: Benefits, cautions and issues," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 381–393, Jul. 2007.

[28] L. Zhang and V. Singh, "Trivariate flood frequency analysis using the Gumbel-Hougaard copula," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 431–439, Jul. 2007.

[29] F. Serinaldi and S. Grimaldi, "Fully nested 3-copula: Procedure and application on hydrological data," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 420–430, Jul. 2007.

[30] P. Laux, S. Vogl, W. Qiu, H. Knoche, and H. Kunstmann, "Copula-based statistical refinement of precipitation in RCM simulations over complex terrain," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 7, pp. 2401–2419, Jul. 2011.

[31] H. Gao, E. Wood, M. Drusch, and M. McCabe, "Copula-derived observation operators for assimilating TMI and AMSR-E retrieved soil moisture into land surface models," *J. Hydrometreol.*, vol. 8, no. 3, pp. 413–429, Jun. 2007.

[32] C. de Michele and G. Salvadori, "A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas," *J. Geophys. Res.*, vol. 108, no. D2, p. 4067, Jan. 2003.

[33] J. Wigneron, Y. Kerr, P. Waldteufel, K. Saleh, M. Escorihuela, P. Richaume, P. Ferrazzoli, P. de Rosnay, R. Gurney, J. Calvet, J. Grant, M. Guglielmetti, B. Hornbuckle, C. Mätzler, T. Pellarin, and M. Schwank, "L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields," *Remote Sens. Environ.*, vol. 107, no. 4, pp. 639–655, Apr. 2007.

[34] Y. Kerr, P. Waldteufel, P. Richaume, J. Wigneron, P. Ferrazzoli, A. Mahmoodi, A. A. Bitar, F. Cabot, C. Gruhier, S. Juglea, D. Leroux, A. Mialon, and S. Delwart, "The SMOS soil moisture retrieval algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1384–1403, May 2012.

[35] D. Carr, R. Kahn, K. Sahr, and T. Olsen, "ISEA discrete global grids," *Stat. Comput. Stat. Graph. Newl.*, vol. 8, no. 2/3, pp. 31–39, 1997.

[36] A. A. Bitar, D. Leroux, Y. Kerr, O. Merlin, P. Richaume, A. Sahoo, and E. Wood, "Evaluation of SMOS soil moisture products over continental US using SCAN/SNOTEL network," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1572–1586, May 2012.

[37] G. Schaefer, M. Cosh, and T. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *J. Atmosp. Ocean. Technol.*, vol. 24, no. 12, pp. 2073–2077, Dec. 2007.

[38] T. Jackson, R. Bindlish, M. Cosh, T. Zhao, P. Starks, D. Bosch, M. Seyfried, S. Moran, Y. Kerr, and D. Leroux, "Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1530–1543, May 2012.

[39] D. Leroux, Y. Kerr, A. A. Bitar, C. Gruhier, R. Bindlish, T. Jackson, B. Berthelot, and G. Portet, "Comparison between SMOS, VUA, ASCAT and ECMWF soil moisture products over four watersheds in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, submitted for publication. AQ27

[40] J. Sabatier, A. Fouilloux, and P. de Rosnay, "Technical implementation of SMOS data in the ECMWF integrated forecast system," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 252–256, Mar. 2012.

[41] [Online]. Available: ftp://n4ftl01u.ecs.nasa.gov/SAN/AMSA/

[42] C. Gruhier, P. de Rosnay, Y. Kerr, E. Mougin, E. Ceschia, J. Calvet, and P. Richaume, "Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions," *Geophys. Res. Lett.*, vol. 35, no. 10, p. L10 405, May 2008.

[43] C. Rudiger, J. C. Calvet, C. Gruhier, T. Holmes, R. de Jeu, and W. Wagner, "An intercomparison of ERS-SCAT and AMSR-E soil moisture observations with model simulations over France," *Amer. Meteorol. Soc.*, vol. 10, no. 2, pp. 431–447, Apr. 2009.

[44] C. Draper, J. Walker, P. Steinle, R. de Jeu, and T. Holmes, "An evaluation of AMSR-E derived soil moisture over Australia," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 703–710, Apr. 2009.

[45] C. Gruhier, P. de Rosnay, S. Hasenauer, T. Holmes, R. de Jeu, Y. Kerr, E. Mougin, E. Njoku, F. Timouk, W. Wagner, and M. Zribi, "Soil moisture active and passive microwave products: Intercomparison and evaluation over a Sahelian site," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 1, pp. 141–156, Jan. 2010.

[46] S. Chaurasia, D. Tung, P. Thapliyal, and P. Joshi, "Assessment of AMSR-E soil moisture product over India," *Int. J. Remote Sens.*, vol. 32, no. 23, pp. 7955–7970, Dec. 2011.

[47] T. Jackson, M. Cosh, R. Bindlish, P. Starks, D. Bosch, M. Seyfried, D. Goodrich, M. Moran, and J. Du, "Validation of advanced microwave scanning radiometer soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4256–4272, Dec. 2010.

[48] B. Schweizer and E. Wolff, "On nonparametric measures of dependence for random variables," *Ann. Stat.*, vol. 9, no. 4, pp. 879–885, Jul. 1981.

[49] Y. Wang and H. Pham, "Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 13–22, Mar. 2012.

[50] C. Genest and L.-P. Rivest, "Statistical inference procedures for bivariate Archimedean copulas," *J. Amer. Stat. Assoc.*, vol. 88, no. 423, pp. 1034–1043, Sep. 1993.

[51] J. Fermanian, "Goodness of fit tests for copulas," *J. Multivariate Anal.*, vol. 95, no. 1, pp. 119–152, Jul. 2005.

[52] C. Genest, J.-F. Quessy, and B. Rémillard, "Goodness-of-fit procedures for copula models based on the probability integral transform," *Scand. J. Stat.*, vol. 33, no. 2, pp. 337–366, Jun. 2006.

[53] C. Genest and B. Rémillard, "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models," *Annales Henri Poincaré*, vol. 44, no. 6, pp. 1096–1127, 2008.

[54] D. Berg, "Copula goodness-of-fit testing: An overview and power comparison," *Eur. J. Finance*, vol. 15, no. 7/8, pp. 675–701, 2009.

[55] C. Genest, B. Rémillard, and D. Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Math. Econom.*, vol. 44, no. 2, pp. 199–213, Apr. 2009.

[56] D. Huard, G. Evin, and A. Favre, "Bayesian copula selection," *Comput. Stat. Data Anal.*, vol. 51, no. 2, pp. 809–822, Nov. 2006.

Delphine J. Leroux received the M.S. degree in applied mathematics from Institut National des Sciences Appliquees, Toulouse, France, in 2009 and the Ph.D. degree in spatial hydrology from Universite Paul Sabatier, Toulouse, in 2012.

From 2009 to 2012, she was with the Centre d'Etudes Spatiales de la Biosphere, Toulouse, where she worked on the validation of the Soil Moisture and Ocean Salinity soil moisture product at the local and global scales by using statistical and physical methods. She is currently with the Jet Propulsion Laboratory (JPL), Pasadena, CA, USA, on the Soil Moisture Active Passive mission. AQ28

Yann H. Kerr (M'88–SM'01–F'13) received the engineering degree from Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, the M.Sc. degree in electronics and electrical engineering from Glasgow University, Glasgow, U.K., and the Ph.D. degree in astrophysique geophysique et techniques spatiales from Université Paul Sabatier, Toulouse. AQ29

He is currently the Director of the Centre d'Etudes Spatiales de la Biosphere, Toulouse. His fields of interest are in the theory and techniques for microwave and thermal infrared remote sensing of the Earth, with emphasis on hydrology and water resource management. He was an EOS Principal Investigator (interdisciplinary investigations) and PI and precursor of the use of the SCAT over land. In 1990, he started to work on the interferometric concept applied to passive microwave Earth observation and was subsequently the Science Lead on the MIRAS project for ESA. In 1997, he proposed the Soil Moisture and Ocean Salinity (SMOS) mission, the natural outcome of the previous MIRAS work. He is currently involved in the exploitation of SMOS data, in the Cal Val activities and related level 2 soil moisture and level 3 and 4 developments. He is also working on the SMOS next concept. AQ30 AQ31 AQ32 AQ33 AQ34 AQ35 AQ36

Dr. Kerr received the World Meteorological Organization 1st prize (Norbert Gerbier), the USDA Secretary's team award for excellence (Salsa Program), and the GRSS certificate of recognition for leadership in the development of the first synthetic aperture microwave radiometer in space and success of the SMOS mission and is a Distinguished Lecturer for GRSS. AQ37 AQ38 AQ39

Eric F. Wood received the B.A.Sc. degree in civil engineering from the University of British Columbia, Vancouver, BC, Canada, in 1970 and the S.M., C.E., and Sc.D. degrees in civil engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1972, 1973, and 1974, respectively.

He is currently a Professor in the Department of Civil and Environmental Engineering, Princeton University, NJ, USA. His recent contributions include the macroscale hydrologic prediction of the coupled water and energy balances of the land surface, remote sensing as an integral tool for observation and modeling of the hydrologic cycle, and developing and evaluating seasonal hydrological forecasts based on coupled seasonal climate forecasts.

Dr. Wood is a fellow of the American Geophysical Union and the American Meteorological Society. He received, among other honors, the Robert E. Horton Award from the Hydrology Section of the American Geophysical Union (1977), the John Dalton Medal from the European Geophysical Union (2007), and the Jules G. Charney Award from the American Meteorological Society (2010).

Alok K. Sahoo received the Ph.D. degree in computational sciences from George Mason University, Fairfax, VA, USA, in 2008.

After finishing his Ph.D., he worked as a Research Associate at the Institute of Global Environment and Society, Beltsville, MD, USA, from 2008 to 2009. He is currently an Associate Research Hydrologist with the Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA. His research interests involve the application of microwave remote sensing in hydrologic applications. His current research projects include soil moisture estimation from microwave sensors and checking consistency among hydrologic cycle variables for drought monitoring.

Rajat Bindlish (S'98–AM'99–M'03–SM'05) received the B.S. degree in civil engineering from the Indian Institute of Technology, Bombay in 1993 and the M.S. and Ph.D. degrees in civil engineering from The Pennsylvania State University in 1996 and 2000, respectively.

He is currently with SSAI, working at USDA Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. His research interests involve the application of microwave remote sensing in hydrology. He is currently working on soil moisture estimation from microwave sensors and their subsequent application in land surface hydrology.

Thomas J.Jackson (SM'96–F'02) received the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 1976.

He is a Research Hydrologist with the U.S. Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. His research involves the application and development of remote sensing technology in hydrology and agriculture, primarily the microwave measurement of soil moisture.

Dr. Jackson is or has been a member of the science and validation teams of the Aqua, ADEOS-II, Radarsat, Oceansat-1, Envisat, ALOS, Soil Moisture and Ocean Salinity, Aquarius, GCOM-W, and Soil Moisture Active Passive remote sensing satellites. He is a fellow of the Society of Photo-Optical Instrumentation Engineers, the American Meteorological Society, and the American Geophysical Union. In 2003, he received the William T. Pecora Award (NASA and Department of Interior) for outstanding contributions toward understanding the Earth by means of remote sensing and the AGU Hydrologic Sciences Award for outstanding contributions to the science of hydrology. He received the IEEE Geoscience and Remote Sensing Society Distinguished Achievement Award in 2011.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Please be aware that the authors are required to pay overlength page charges ($200 per page) if the paper is longer than 6 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

AQ1 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ2 = Please provide the expanded form of “USDA.”
AQ3 = Please provide the expanded form of “ARS.”
AQ4 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ5 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ6 = “AMSR-E” is defined as “Advanced Microwave Scanning Radiometer-Earth Observing System” for consistency. Please check if appropriate, and correct if necessary.
AQ7 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ8 = “Picard *et al.*” was changed to “Picard and Fily.” Please check if appropriate, and correct if necessary."
AQ9 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ10 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ11 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ12 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ13 = “rmse” is defined as “root-mean-square error.” Please check if appropriate, and correct if necessary."
AQ14 = The sentence was rephrased. Please check if the original thought was retained, and correct if necessary."
AQ15 = Please provide the expanded form of “NASA.”
AQ16 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ17 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ18 = The caption was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ19 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ20 = “CDF” is defined as “cumulative density function” for consistency. Please check if appropriate, and correct if necessary.
AQ21 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.

AQ22 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ23 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ24 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ25 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ26 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ27 = Please provide publication update in Ref. [39].
AQ28 = The author's current affiliation indicated in the footnote did not correspond to the current affiliation provided in the curriculum vitae. Please check.
AQ29 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ30 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ31 = Please provide the expanded form of "EOS."
AQ32 = Please provide the expanded form of "PI."
AQ33 = Please provide the expanded form of "SCAT."
AQ34 = Please provide the expanded form of "MIRAS."
AQ35 = Please provide the expanded form of "ESA."
AQ36 = Please provide the expanded form of "Cal Val."
AQ37 = "Dr." was inserted as the title for author Yann H. Kerr. Please check if appropriate, and correct if necessary.
AQ38 = Please provide the expanded form of "GRSS."
AQ39 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ40 = Please provide the specific location of the Indian Institute of Technology, Bombay.
AQ41 = Please provide the specific location of The Pennsylvania State University.
AQ42 = Please provide the expanded form of "SSAI."
AQ43 = Please provide the expanded form of "ADEOS-II."
AQ44 = Please provide the expanded form of "ALOS."
AQ45 = Please provide the expanded form of "GCOM-W."
AQ46 = Please provide the expanded form of "AGU."

END OF ALL QUERIES

An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS

Delphine J. Leroux, Yann H. Kerr, *Fellow, IEEE*, Eric F. Wood, Alok K. Sahoo, Rajat Bindlish, *Senior Member, IEEE*, and Thomas J. Jackson, *Fellow, IEEE*

***Abstract*—Overlapping soil moisture time series derived from two satellite microwave radiometers (the Soil Moisture and Ocean Salinity and the Advanced Microwave Scanning Radiometer-Earth Observing System) are used to generate a soil moisture time series from 2003 to 2010. Two statistical methodologies for generating long homogeneous time series of soil moisture are considered. Generated soil moisture time series using only morning satellite overpasses are compared to ground measurements from four watersheds in the U.S.A. with different climatologies. The two methods, cumulative density function (CDF) matching and copulas, are based on the same statistical theory, but the first makes the assumption that the two data sets are ordered the same way, which is not needed by the second. Both methods are calibrated in 2010, and the calibrated parameters are applied to the soil moisture data from 2003 to 2009. Results from these two methods compare well with ground measurements. However, CDF matching improves the correlation, whereas copulas improve the root-mean-square error.**

***Index Terms*—Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), cumulative density function (CDF) matching, copulas, Soil Moisture and Ocean Salinity (SMOS), soil moisture, time series.**

I. Introduction

SOIL moisture is an important variable and is now considered as an essential climate variable by the World Meteorological Organization [1]. It has a crucial role in the transfers of water and energy between the soil and the atmosphere. Soil moisture is also an input variable for land surface modeling in determining the evaporative fraction at the surface and the infiltration in the root zone. For both agriculture and water resource management, soil moisture information is essential at local and regional scales. At global scales, soil moisture is of great value for weather forecasting [2], climate change [3], and monitoring extreme events such as floods and droughts.

Soil Moisture and Ocean Salinity (SMOS) [4] was successfully launched by the European Space Agency in November 2009 and since has been providing global maps of soil moisture every three days at a nominal spatial resolution of 43 km with an accuracy of 0.04 m^3/m^3. SMOS is the first mission specifically designed for soil moisture monitoring. The Soil Moisture Active Passive (SMAP) mission [5] is scheduled for launch in October 2014 by the National Aeronautics and Space Administration. SMAP will continue the time series of soil moisture based on 1.4-GHz radiometer observations that began with SMOS. The 1.4-GHz frequency channel is the most suitable frequency for soil moisture retrieval [6].

Longer time series of satellite-based soil moisture would be of value in climate-related analysis. Utilizing the data from the previous generations of satellite sensors involves resolving numerous issues. Some of the platforms and approaches have been developed to retrieve soil moisture using the higher frequencies, which has been the only option until now. These include the Scanning Multichannel Microwave Radiometer (1978–1987) [7], the Special Sensor Microwave/Imager (1987–current) [7], the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (2002–2011) [7], [8], WindSat (2003–current) [9], and the European Remote Sensing-Advanced Scatterometer (1991–current) [10]. Although their lowest frequencies (5–20 GHz) are not the most suitable for soil moisture retrievals (higher sensitivity to vegetation growth and atmospheric conditions), they remain a valuable time series from 1978 until now. Applications such as data assimilation or climate change assessment require consistent products. The products referenced earlier have been retrieved using different sensors with different algorithms, and as a result, the time series is not homogeneous. This heterogeneity can be interpreted as a bias and is a problem in the data assimilation process. To avoid this issue, these products need to be processed to correct for any bias or amplitude variation between the data sets.

Many previous studies have developed various methods for the homogenization of time series. Vincent *et al.* [11] developed a method to harmonize temperature time series with gaps. The first step was to determine if the series was homogeneous by comparing its anomalies to those of a reference series. The identification of the gaps and their magnitude was performed by successively fitting a linear model with different magnitude values with the best fit being indicated by the minimum sum of square errors. Homogeneous temperature and precipitation time series were developed by Begert *et al.* [12] using statistical

Manuscript received November 28, 2011; revised May 14, 2012 and October 22, 2012; accepted December 22, 2012. This work was supported in part by Telespazio France and in part by TOSCA.

D. J. Leroux is with the Centre d'Etudes Spatiales de la Biosphere, 31400 Toulouse, France, and also with Telespazio France, 31023 Toulouse Cedex 1, France (e-mail: delphine.j.leroux@gmail.com).

Y. H. Kerr is with the Centre d'Etudes Spatiales de la Biosphere, 31400 Toulouse, France (e-mail: yann.kerr@cesbio.cnes.fr).

E. F. Wood and A. K. Sahoo are with the Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: efwood@princeton.edu; aksahoo2004@gmail.com).

R. Bindlish and T. J. Jackson are with the USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705 USA (e-mail: rajat.bindlish@ars.usda.gov; tom.jackson@ars.usda.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2240691

methods to detect potential inhomogeneity. In that study, a reference time series was necessary in order to detect and AQ8 compute the magnitude of the shifts. Picard and Fily [13] proposed a method to simulate a homogeneous time series of the cumulative melting surface in Antarctica. Using satellite observations from different sensors and acquisition times was the biggest challenge. Correcting for the effect of the observing time was accomplished in two steps. First, a sinusoidal function with a 24-h periodicity was fitted, and then, an optimal interpolation to refine this first guess model to *force* it to be closer was applied to the observations and to provide very low uncertainty around observation time and larger uncertainty when there is no available observation.

Matching the cumulative density functions (CDFs) of two data sets has been used in several studies to merge time series. Reichle and Koster [14] and Choi and Jacobs [15] merged soil moisture derived from satellite observations with model data, and Li *et al.* [16] corrected the bias of precipitation and temperature products derived from different models. CDF matching was also used as a preliminary step of the assimilation process [17] and to produce long time series of soil moisture [18], [19].

Over the last few years, a new method based on copula functions has been developed. It allows the derivation of bivariate distributions without making the assumptions required when dealing with multivariate frequency distributions, e.g., the same type of marginal distribution for both variables, a joint normal distribution, and independent variables. One of the major advantages of the copula method is that the marginal distributions can be of any form [20]. The first comprehensive treatment of copulas was by Nelsen [21]. He presented methods to construct copulas and discussed the role played by copulas in modeling and dependence. Since then, copulas have been applied in various applications with the majority of the literature dedicated to the financial sector [22], [23]. In the field of hydrology, some applications have emerged. Genest and Favre [24] summarized the existing methods to detect and evaluate the dependence between the data sets through copulas (analytically and graphically) and enumerated the various methods to choose the best copula family and estimate their parameters. Favre *et al.* [25] applied copulas to peak flows and volumes from two watersheds, Salvadori and De Michele [26] to storm and rainfall time series, Dupuis [27] to the volume and duration of low flows of two rivers, Zhang and Singh [28] to rainfall frequency, Serinaldi and Grimaldi [29] to flood and sea frequency, and Laux *et al.* [30] to precipitation data. Gao *et al.* [31] used copulas as a preprocessing step for the assimilation process on soil moisture data.

Joint statistical analysis has already been applied when the sources of the soil moisture measurements come from different observation systems (e.g., AMSR-E surface soil moisture and 10-cm soil moisture from a land surface model [14]). Similarly, joint statistical methods form the basis for data assimilation of satellite soil moisture into land surface models [31]. There are many other studies related to joint probability, including where the variables are physically different but where their statistical relationships are useful (e.g., rainfall storm intensity and storm duration [32]).

The goal of this paper is to estimate for all the AMSR-E period (2003–2010) SMOS-equivalent observations that can be used to develop a statistical representation of SMOS retrieval so that current and future SMOS retrievals can be used in applications like drought monitoring based on percentiles. However, AQ9 matching 130 am C-/X-band (AMSR-E) observations with 600 am L-band (SMOS) observations presents some issues: 1) The crossing times are different, and rainfalls may occur between the two acquisitions; and 2) the frequencies are different, so the sensing depths are not similar.

The statistical impact of the rainfalls that could occur between 130 am and 600 am is to lower the correlation. However, if the correlation is sufficiently high, a statistical relationship can be established to estimate an equivalent SMOS value from an AMSR-E observation. This high correlation implies that the occurrence of precipitation between the SMOS and AMSR-E overpasses is rare. Moreover, it is well known that soil moisture has a long temporal correlation time scale, so the overpass time differences will have a minimal effect on the analysis. AQ10

The impact of the different frequencies between AMSR-E and SMOS is, in most situations, not significant. The higher AMSR-E frequency (10.7 GHz) results in a more superficial emission depth than the SMOS observations, so while the retrieved values may be different, their relative values will be similar (both dry or wet). The correlation between paired observations depends on their relative values (with their individual time series) and not absolute values, and in the case of copula-based joint distributions, the correlation is represented by the Kendall tau whose calculation is based on ranks.

If the two sensing depths were to be reconciled physically, given the soil property variability (spatially and with depth) with different wetting and drying properties, a physical model would introduce significant uncertainty that could be very difficult to estimate afterward. If the SMOS (or AMSR-E) data were adjusted to the AMSR-E (or SMOS) emission depth through data assimilation into a land surface model for example, then the complete record would have to be adjusted with the added uncertainty of the data assimilation step. With any of the suggested adjustments, there is a mismatch with the past or with the future. Only by treating the original data sets and determining the information content between them can a consistent approach be represented. AQ11

Data assimilation could, however, deal with the precipitation and the difference in sensing depth issues, but that would imply other uncertainties such as the space–time variability of the precipitation data sets, as well as other meteorological issues. Building a homogeneous time series based on data assimilation into a land surface model can be seen as a competing approach.

In this paper, we show two statistical methods to obtain this homogeneous time series. The satellite data and the four watersheds where the time series are simulated are presented in Section II. The two statistical methods for generating homogeneous time series are presented in Section III which includes the general theory and how to apply them to real data. Simulated time series over the four watersheds are presented in Section IV. Conclusions and perspectives are described in the last section.

II. Regions of Interest and Satellite Data

A. SMOS

With its L-band radiometer, SMOS [4] has been providing soil moisture data for almost three years and global coverage every three days with a 43-km resolution. The satellite is polar orbiting with equator crossing times of 6 am (local solar time (LST), ascending) and 6 pm (LST, descending). The signal at L-band is mainly influenced by the water content at the surface of the soil (around 5 cm).

SMOS acquires brightness temperatures at multiple incidence angles, from 0° to 55° with full polarization. The angular signature is a key element of the retrieval algorithm that provides soil moisture and the vegetation optical thickness through the minimization of a cost function between modeled and acquired brightness temperatures [33], [34]. This estimated soil moisture is referred as the Level 2 product [34] and is AQ12 available on the Icosahedral Snyder Equal Area-4h9 grid [35]. The nodes of this grid are equally spaced at about 15 km. In this paper, the 2010 SMOS Level 2 version 4 products have been used.

Currently, numerous studies are underway on the validation of SMOS soil moisture product with *in situ* measurements and estimates of other sensors and models. Bitar *et al.* [36] used the Soil Climate Analysis Network [37] and the Snowpack Telemetry sites in North America to compare SMOS soil moisture retrievals and ground measurements. That study showed that SMOS soil moisture had a very good dynamic response but tended to underestimate the values. However, the new version of the product (V4) significantly improved the general results. Jackson *et al.* [38] studied SMOS soil moisture and vegetation optical depth over four watersheds in the U.S. They concluded that SMOS almost met the accuracy AQ13 requirement with root-mean-square errors (rmses) of 0.043 and 0.047 m^3/m^3 in the morning and afternoon, respectively, whereas the vegetation optical depth retrievals were not reliable AQ14 yet for use in vegetation analyses. Leroux *et al.* [39] compared SMOS data with other satellite and model output products over the same four watersheds for the year 2010. It showed that SMOS soil moisture data were closer to the ground measurements than the other data sets. Even though the correlation coefficient was not the best, the bias was extremely small.

After the results of the validation activities, the European Center for Medium-Range Weather Forecasts has decided and is now ready to process SMOS data in near real time into their Integrated Forecast System. It is expected to have an impact on the weather forecast at short and medium ranges [40].

B. AMSR-E

The AMSR-E was launched in June 2002 on the Aqua satellite. This radiometer acquires data with a single 55° incidence angle at six different frequencies: 6.9, 10.7, 18.7, 23.8, 36.5, and 89.0 GHz, all dual polarized. The crossing times are respectively 1:30 am (LST, descending) and 1:30 pm (LST, ascending).

There are several soil moisture products available that are based on AMSR-E data. Many studies have already showed

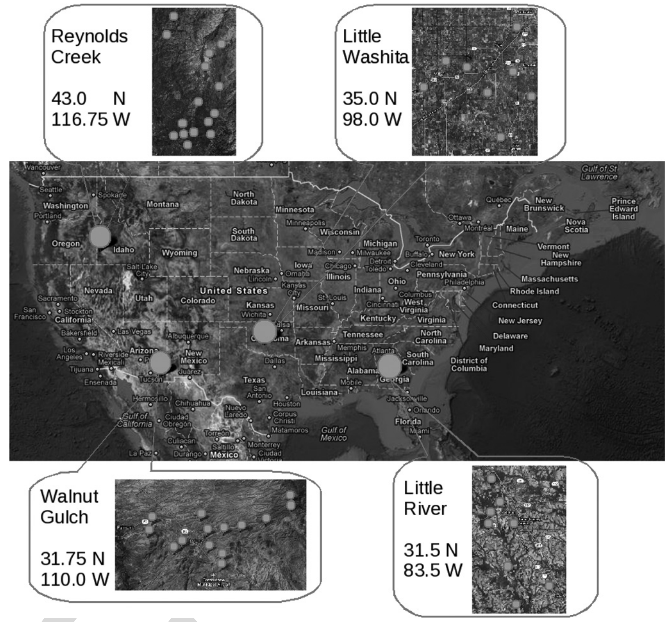


Fig. 1. Map of the four sites: WG, AZ; LW, OK; LR, GA; and RC, ID.

that the NASA product [41] is not able to reproduce low values AQ15 of soil moisture and has low dynamic range [42]–[46]. The soil moisture data produced by the joint collaboration of the Vrije University of Amsterdam and NASA (whereafter called the Land Parameter Retrieval Model (LPRM) [7]) were chosen in this study. AQ16

The LPRM [7] retrieves soil moisture and optical thickness using the C- and X-band AMSR-E channels (combined product) and 36.5 GHz to estimate the surface temperature. This algorithm is based on a microwave radiative transfer model with *a priori* information about soil characteristics. The products are available on a $0.25° \times 0.25°$ grid only for the descending orbit. These data have been quality controlled, and the contaminated estimates due to high topography and extreme weather conditions such as snow have been flagged and not been considered in this study. AQ17

C. Study Areas

Four watersheds located in the United States were selected for this study: Walnut Gulch (WG) in Arizona, Little Washita (LW) in Oklahoma, Little River (LR) in Georgia, and Reynolds Creek (RC) in Idaho (see Fig. 1). They represent different types of climate (from semiarid to humid) and land use patterns [47]. These four watersheds have been used as calibration and validation sites for comparison of AMSR-E satellite product [47] and SMOS product [38], [39].

WG is located in the Southeast Arizona. Most of the watershed is covered by shrubs and grass, which is typical of the region. The annual mean temperature is 17.6 °C (at Tombstone), and the annual mean precipitation is 320 mm (mainly from high intensity convective thunderstorms in the late summer). The uppermost 10 cm of the soil profile contains up to 60% gravel, and the underlying horizons usually contain less than 40% gravel.

AQ18 TABLE I
WATERSHED CHARACTERISTICS AND THE COORDINATES OF THE BOX CONTAINING THE POINTS USED FOR STATISTICS

Watershed	Number of stations	Climate	Annual rainfall (mm)	Topography	Land use	Box for statistics (corners coord.)
Walnut Gulch AZ	14	semi-arid	320	rolling	range	31.3 N - 110.5 W 32.3 N - 109.5 W
Little Washita OK	8	sub-humid	750	rolling	range/wheat	34.4 N - 98.5 W 35.4 N - 97.5 W
Little River GA	8	humid	1200	flat	row crop/forest	31.0 N - 84.0 W 32.0 N - 83.0 W
Reynolds Creek ID	15	semi-arid	500	mountainous	range	34.7 N - 98.7 W 35.7 N - 97.7 W

TABLE II
CORRELATION COEFFICIENTS (R) BETWEEN THE IN SITU MEASUREMENTS AT 130 AM AND 600 AM FOR THE FOUR WATERSHEDS. N IS THE NUMBER OF AVAILABLE DATES, AND CI IS THE 95% CONFIDENCE INTERVAL

WG			LW		
R	N	CI	R	N	CI
0.96	365	[0.95-0.97]	0.97	365	[0.96-0.98]
LR			RC		
R	N	CI	R	N	CI
0.95	365	[0.94-0.96]	0.99	328	[0.99-0.99]

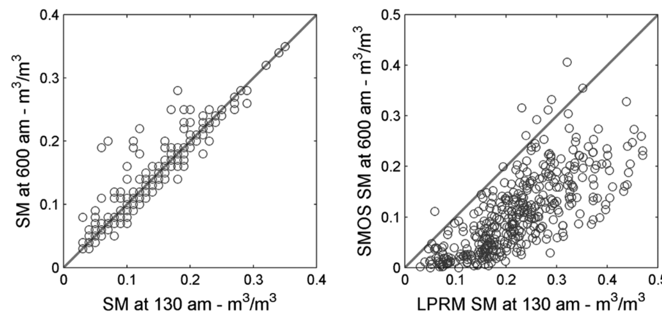
LW is located in Southwest Oklahoma in the Southern Great Plains region of the U.S. The climate is subhumid with an average annual rainfall of 750 mm (mainly during the spring and fall seasons). Topography is moderately rolling with a maximum relief of less than 200 m. Land use is dominated by rangeland and pasture (63%).

LR is located in the Southern Georgia near Tifton. With an average annual precipitation of 1200 mm, the climate is humid. The LR watershed is typical of the heavily vegetated slow-moving stream systems in the Coastal Plain region of the U.S. The topography over this watershed is relatively flat. Approximately 40% of the watershed is forest with 40% crops and 15% pasture.

RC is located in a mountainous area of Southwest Idaho. The topography is high with a relief of over 1000 m that results in diverse climates. Soils and vegetations are typical in this part of the Rocky Mountains. The climate is considered as semiarid with an annual precipitation of 500 mm. Approximately 75% of the annual precipitation at high elevation is snow, whereas only 25% is snow at low elevation.

Surface soil moisture and temperature sensors (0–5 cm) have been acquiring data since 2002 for the four watersheds. The data used in this study are the means and standard deviations of the soil moisture and surface temperature acquired every 30 min from 2009 to 2010 (hourly for RC). The averages are based on 14/8/8/15 sensors for WG/LW/LR/RC, respectively, after eliminating sensors with poor and suspicious
AQ19 performances. Weighting coefficients have been derived for each sensor with a Thiessen polygon. Table I summarizes the characteristics of each watershed [47].

In order to estimate the effect of the rainfalls that could occur between 130 am and 600 am, the correlation coefficients between the measurements at 130 am and 600 am have been computed for the four watersheds (see Table II and Fig. 2). They range from 0.95 to 0.99, and based on the fact that rainfalls would lower the correlation, we can assess that precipitations that do not affect significantly the analysis.



(a) (b)

Fig. 2. Comparison between the 130 am and the 600 am soil moisture: *In situ* observations and satellite products for the four watersheds. (a) *In situ* soil moisture at 130 am and 600 am. (b) LPRM (130 am) and SMOS (600 am) soil moisture.

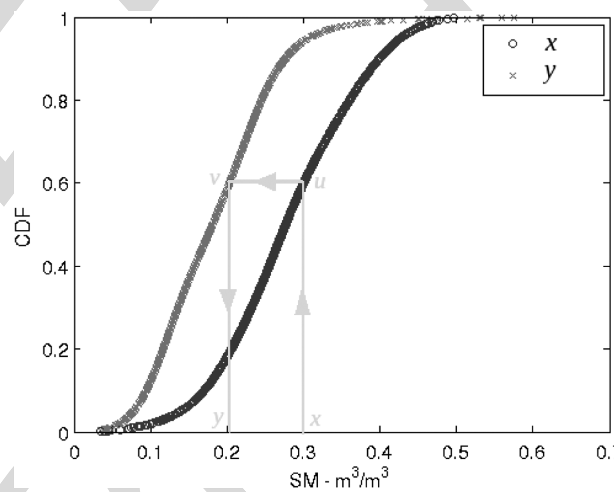


Fig. 3. Principle of CDF matching by setting the probabilities equal. For a given x, find y such that $G_Y(y) = F_X(x)$.

III. TWO STATISTICAL METHODS FOR GENERATING HOMOGENEOUS TIME SERIES

Two statistical methods were used to create a homogeneous time series of soil moisture. CDF matching has been widely used in previous studies to merge time series [14], [15], [18], [19], whereas copulas have just started to be used recently for environmental purposes.

A. CDF Matching

The CDF is the probability that a random variable X takes a value less than or equal to a given number x

$$F_X(x) = \Pr[X \leq x] \quad (1)$$

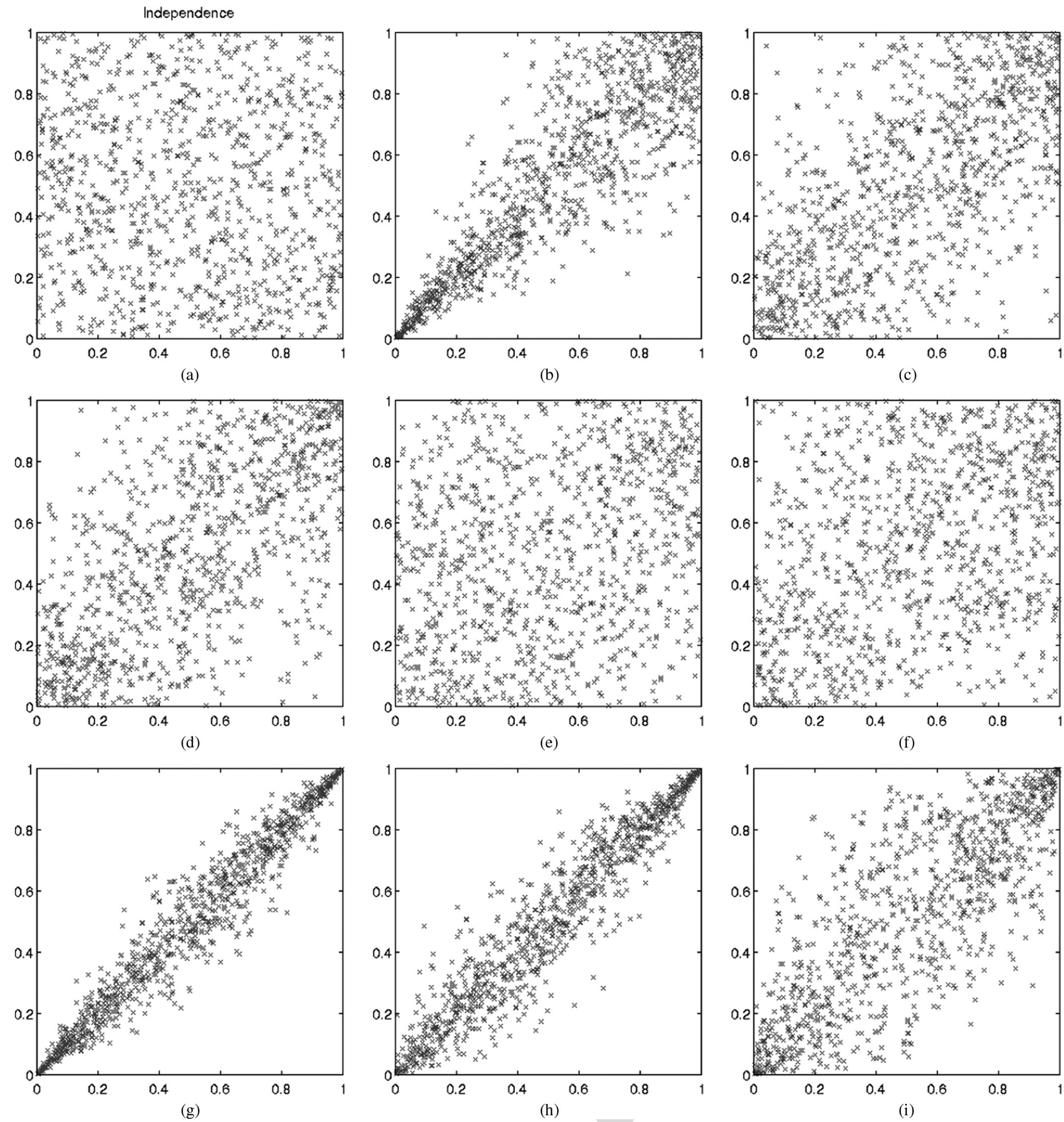


Fig. 4. Representations of the nine copulas showing their characteristics in the form of the point cloud (x-axis: CDF of the first data set; y-axis: CDF of the second data set).

where F_X is the CDF of the random variable X. If two time series are considered, the CDF matching consists of matching the CDF of each data set by setting their probabilities equal (see Fig. 3). The following approach has been applied here to the soil moisture data.

1) Compute the CDF of both data sets X and Y: F_X and G_Y.
2) Given a value x of X, find y such that $G_Y(y) = F_X(x)$.

However, the assumption that the probabilities $F_X(x)$ and $G_Y(y)$ are equal is never confirmed, and most of the time, they are scattered like in Fig. 4. The copula method models this dependence between the probabilities.

For the rest of this paper, we use the variable u to represent $F_X(x)$ and v for $G_Y(y)$. U and V are data sets, whereas u and v are values of these data sets.

B. Copulas

The copula theory is a very useful and powerful tool to model the dependence structure between two sets of random variables.

TABLE III
NINE COPULAS TESTED IN THE STUDY: DEFINITION, PARAMETER RANGE, AND FAMILY

Copula	$C_\theta(u,v)$	$\theta \in$	Family
Independent	$u \cdot v$	-	-
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$(0,\infty)$	Archimedean
Frank	$\frac{-1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$(-\infty,\infty)/0$	Archimedean
Gumbel	$\exp\left(-\left((-\ln u)^\theta+(-\ln v)^\theta\right)^{1/\theta}\right)$	$[1,\infty)$	Archimedean
FGM	$uv+\theta uv(1-u)(1-v)$	$[-1,1]$	Elliptical
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$[-1,1]$	Archimedean
Arch12	$\left(1+\left((u^{-1}-1)^\theta+(v^{-1}-1)^\theta\right)^{1/\theta}\right)^{-1}$	$[1,\infty)$	Archimedean
Arch14	$\left(1+\left((u^{-1/\theta}-1)^\theta+(v^{-1/\theta}-1)^\theta\right)^{1/\theta}\right)^{-1/\theta}$	$[1,\infty)$	Archimedean
Gaussian	$\frac{\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\exp\left(\frac{2\theta s\omega-s^2-\omega^2}{2(1-\theta^2)}\right)}{2\pi\sqrt{1-\theta^2}}$	$[-1,1]$	Elliptical

Like the CDF matching, copulas separate the marginal behavior of variables from the dependence structure by using distribution functions. Instead of setting the probabilities u and v equal, the variables U and V are compared and analyzed. The copula function binds the two variables together.

There are many families of copulas which exhibit very different properties. The form of the scatter of U and V is controlled by the family choice, and the width of the tail of this scatter is controlled by the single parameter θ. Most of the definitions that follow in this section are based on [21].

1) General Theory: A copula is a function that generates a multivariate cumulative distribution function from 1-D marginal CDFs. Given two random variables, X and Y, with marginal CDFs F_X and G_Y, then, Sklar's theorem states AQ20

$$H_{XY}(x,y)=C_{XY}\left(F_X(x),G_Y(y)\right)=\Pr[X\leq x, Y\leq y] \quad (2)$$

where H_{XY} is the joint CDF of X and Y and C_{XY} is the associated copula function. It is then possible to derive conditional distributions, $H_{XY}(y|x)$, i.e., the joint CDF knowing x. Let $u=F_X(x)$ and $v=G_Y(y)$. Then, $H_{XY}(y|x)$ can be derived by

$$C_{V|U}=\frac{\partial C(u,v)}{\partial u}. \quad (3)$$

Schweizer and Wolff [48] established that the copula function accounts for all the dependence between the two variables. They demonstrated that transformations of the variables X and Y do not affect their associated variables. Thus, the way that X and Y evolve together is captured by the copula, regardless of the scale in which each variable is measured.

2) Some Copula Families: The product copula corresponds to the independence between X and Y

$$C(u,v)=u\cdot v. \quad (4)$$

A copula of the Archimedean family takes the following form:

$$C(u,v)=\phi^{-1}\left(\phi(u)+\phi(v)\right) \quad (5)$$

where ϕ is the generator function that goes from [0, 1] to $(0,\infty)$. It satisfies three conditions: $\phi(1)=0$, ϕ strictly decreasing, and ϕ convex.

Elliptical copulas have distributions with elliptic contours. The main advantage of elliptical distributions is that the level of correlation between the variables U and V can be specified. The disadvantages are that elliptical copulas do not have closed-form expressions and are restricted to have radial symmetry. AQ21

In this paper, nine copulas were used: the product copula, Clayton, Frank, Gumbel, Farlie–Gumbel–Moregenstern (FGM), Ali–Mikhail–Haq, Arch12 (the 12th copula presented in [21]), Arch14 (the 14th copula presented in [21]), and the Gaussian copula. The nine copulas are described in Table III and Fig. 4 and have their own characteristics.

1) Clayton: Strong left tail dependence and relatively weak right tail dependence (i.e., u and v are strongly linked for low values, whereas they are not for high values).
2) Frank: Dependence is symmetric in both tails, weak in both tails, and stronger in the center of the distribution. AQ22
3) Gumbel: Strong right tail dependence and relatively weak left tail dependence (the opposite of Clayton).
4) FGM: Useful when the dependence between U and V is modest in amplitude.
5) Gaussian: Flexible as it allows for positive and negative dependences.

Hafner and Reznikova [23] and Wang and Pham [49] developed a method that includes the time into the copula formula to create a dynamic copula evolving with time. In this paper, time was not included, but the year 2010 was divided into four seasons as different statistical behaviors were expected: December–January–February, March–April–May (MAM), June–July–August (JJA), and September–October–November (SON).

3) How to Select a Family: Since copulas separate marginal distributions from dependence structures, the appropriate copula for a particular application is the one that best captures the dependence features of the data [22]. Dupuis [27] examined the effects of model misspecification and highlighted the dangers of improper copula selection. Genest and Rivest [50] proposed a method to select the most appropriate copula, but this method is only relevant for Archimedean copulas. Other methods were developed to compare any type of copulas [51]–[54]. Genest *et al.* [55] and Berg [54] compared some of them and concluded that there was no universal test and that some procedures performed better in some situations but never in all the situations.

The method proposed by Huard *et al.* [56] is based on a Bayesian approach where any type of copula can be tested. It does not perform perfectly well in all the situations (with small correlation coefficients or with small sample size) but has the advantage to be a very fast method. This method was chosen in this study to select the copula that provides the best fit to the data.

4) Method Used for Simulations: The key to generating simulations from a copula is to understand that a copula is a joint distribution and that it obeys to the same rules. A conditional copula $C_{V|U}(u,v)$ is the probability that the random variable V is less than or equal to a value v knowing that the random variable U is equal to a value u

$$C_{V|U}(u,v) = \Pr[V \leq v \mid U = u] = t \sim \mathcal{U}(0,1). \quad (6)$$

Simulating a uniform variable t is necessary in order to generate simulations from a copula. To retrieve $V|U$, the function $C_{V|U}$ needs to be inverted such that $v = C_{V|U}^{-1}(t)$, or the equation $C_{V|U}(v) = t$ needs to be solved numerically. For each value of t, a value for v is retrieved. The following approach was used here to simulate data with the copulas.

1) Compute F_X and G_Y from the two original data sets X and Y with (1).
2) Choose the appropriate copula C by applying Huard's method and fitting the parameter θ to the original data.
3) Derive the conditional copula $C_{V|U}$ with (3).
4) Generate 1000 simulations $t \sim \mathcal{U}(0,1)$.
5) Compute v with $v = C_{V|U}^{-1}(t)$ and y with $y = G_Y^{-1}(v)$.
6) The mean and standard deviation from the 1000 simulations can be computed.

IV. Methodology

For the CDF matching and the copula methods, 2010 data were used for calibration. The CDFs of SMOS and LPRM were calculated for the 2010 data sets. The two algorithms were then applied to the data from previous years. It should be noted that the consequence of using 2010 as a calibration year is that only the soil moisture range from 2010 is taken into account. If an extreme event occurred in the previous years, it might not be well described with these methods as they are only based on statistics and not on physical models. By looking at the *in situ* soil moisture time series in Fig. 7, 2010 did not have enough wet values over LR to estimate correctly the strong rainfalls of 2004, 2005, and 2009, not enough wet values over LW for rainfalls in 2007 and not enough dry values as well for 2003 and 2006, and again not enough dry values over RC for all the AQ23 previous years.

The two methods were applied to data contained in a $1^\circ \times 1^\circ$ box around each watershed in order to have enough points for computing reliable statistics. The coordinates of each box are indicated in Table I. Only the satellite morning overpasses were selected for this study (6:00 am for SMOS and 1:30 am for AMSR-E, LST) since LPRM retrievals were only available for this overpass.

The 2010 calibration year was divided into four seasons: December–January–February, MAM, JJA, and SON. This subdivision was done in order to better capture the seasonal dynamic that can be very different depending on the time of the year, particularly in vegetated areas. However, not enough points were available during the winter period (December–January–February) to compute reliable statistics, so no estimation was performed for this season.

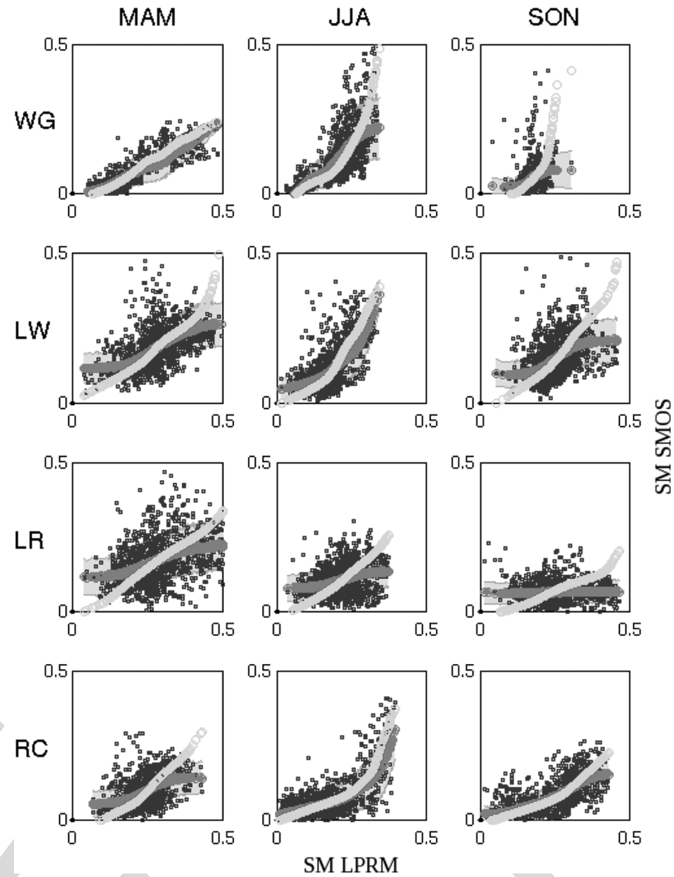


Fig. 5. Discrepancies in the simulations of soil moisture between CDF matching and copulas in 2010. Original soil moisture LPRM data are represented by blue points, and simulated data with CDF matching and copulas are in green and red, respectively. The standard deviation of the copula simulations is represented in shadowed red. Each row corresponds to a site, and each column corresponds to a season. x-axis: LPRM soil moisture. y-axis: SMOS soil moisture. AQ24

When comparing either two different remote sensing products or *in situ* data with remote sensing products, there is the issue of the scale effect, as the products may have significantly different spatial resolutions. Moreover, the spatial variability varies with the seasons and the heterogeneity. So as to reduce the problem, we used in this study averaged *in situ* data sets (8 to 15 stations that were several miles away) which were especially produced to be representative of 50-km spatial resolution or so [47]. Also, statistics were applied to all the points contained in a $1^\circ \times 1^\circ$ box (more than 50 grid points).

V. Generated Homogeneous Time Series

The year 2010 was used to compute the CDFs of each data set (SMOS and LPRM) for both methods and the joint CDF based on fitting and selecting copula functions as described previously. The soil moisture data were estimated using

TABLE IV
STATISTICAL RESULTS OF THE SIMULATIONS FROM COPULAS AND CDF MATCHING. THE SIMULATIONS WERE COMPARED TO GROUND MEASUREMENTS OVER 2010 DIVIDED INTO FOUR SEASONS: MAM, JJA, SON, BUT NOT ENOUGH DATA AVAILABLE FOR WINTER SEASON. THE BEST RESULTS ARE WRITTEN IN BOLD, AND RMSES ARE IN m^3/m^3

AQ25

		SMOS		LPRM		Copula method			CDF matching		# points
		R	RMSE	R	RMSE	Fam(θ)	R	RMSE	R	RMSE	
WG	MAM	0.80	0.032	0.82	0.125	Gumbel (2.18)	**0.89**	**0.020**	0.87	0.031	43
	JJA	0.86	0.053	0.86	0.126	Clayton(2.63)	0.76	**0.076**	**0.81**	0.090	45
	SON	0.64	0.029	0.79	0.133	Frank (3.13)	**0.64**	**0.012**	0.53	0.029	42
	total	0.84	0.040	0.79	0.139	-	0.79	**0.043**	**0.82**	0.054	159
LW	MAM	0.70	0.068	0.48	0.166	Frank (4.40)	0.55	**0.057**	**0.57**	0.075	44
	JJA	0.85	0.037	0.58	0.085	Gumbel (1.66)	**0.77**	**0.042**	0.76	0.050	44
	SON	0.80	0.041	0.80	0.122	Frank (3.61)	**0.75**	**0.023**	0.72	0.048	46
	total	0.78	0.049	0.59	0.148	-	**0.71**	**0.043**	**0.71**	0.059	162
LR	MAM	0.77	0.080	0.54	0.175	Frank (2.82)	**0.59**	**0.063**	0.58	0.067	39
	JJA	0.57	0.053	0.67	0.131	Frank (2.00)	0.65	0.034	**0.66**	**0.033**	40
	SON	0.59	0.032	0.37	0.174	FGM (0.31)	**0.17**	**0.033**	0.16	0.037	39
	total	0.74	0.060	0.65	0.178	-	0.51	**0.045**	**0.59**	0.048	147
RC	MAM	0.14	0.097	0.11	0.096	Frank (3.10)	0.26	**0.089**	**0.27**	0.105	47
	JJA	0.63	0.055	0.81	0.070	Gumbel (1.81)	**0.84**	**0.047**	0.83	0.052	42
	SON	0.14	0.070	0.52	0.144	Frank (6.30)	**0.34**	**0.056**	0.29	0.066	39
	total	0.55	0.081	0.73	0.099	-	**0.80**	**0.059**	0.70	0.067	142

the conditional distribution (conditional on LPRM retrievals). While the copula procedure has the potential to generate an ensemble of SMOS-like soil moisture estimates, given the LPRM estimated soil moisture, we only use the mean estimate. The ensembles could be used to provide uncertainty estimates. It should be noted that CDF matching can only provide a single SMOS estimate. The resulting time series will result in a statistically homogeneous time series under the assumption that 2010 LPRM retrievals and the underlying AMSR-E brightness temperatures are temporally consistent. The resulting SMOS-like estimated soil moisture is then compared to ground measurements.

A. *Calibration Year 2010 and Comparison With Ground Measurements*

2010 is the year with both SMOS data and LPRM data. CDFs were computed for both variables. CDF matching and copula methods were then applied, and these produced different SMOS-like estimates. In Fig. 5, the original data (SMOS and LPRM) are represented by the blue point cloud, CDF matching and copula estimates are in green and red colors, respectively, and standard deviations from copula simulations are in red shadows. AQ26 This standard deviation can be interpreted as the uncertainty associated to the copula simulations, which can be not produced by CDF matching estimation.

Over WG in the MAM season, there was no obvious difference between the two simulation methods. However, in the JJA and SON seasons, there were differences for the high values of soil moisture: The CDF matching method produced higher simulated values than the copula method. Similar behavior can also be seen for all seasons in the other three sites, i.e., LW, LR, and RC. Discrepancies can also be observed for small values of soil moisture over LW, LR, and RC (MAM) where copulas generated higher values of soil moisture.

Standard deviations of soil moisture simulations from copulas were also computed (see Fig. 5). This standard deviation is directly related to the width of the tail of the chosen copula which is controlled by the θ parameter. A high value of the standard deviation corresponds to a large tail, meaning that the two variables are weakly linked to each other, whereas a small value corresponds to a strong link. The differences in the simulations can also be observed in the 2010 time series (see Table IV and Fig. 6). Compared to the original LPRM data, the estimated soil moisture was close to the SMOS level and comparable to the ground measurements. The bias between LPRM and SMOS was corrected by both methods.

Over WG, CDF matching and copula simulations were not very different except in the summer season when the CDF matching simulations were higher than the copulas. Considering the entire year, both simulation methods improved the original statistics from the LPRM data set. The correlation coefficient did not change significantly ($R = 0.79$ for LPRM and $R = 0.79/0.82$ for copulas/CDF matching), but the rmse was highly improved going from 0.139 m^3/m^3 (original LPRM data) to 0.054 m^3/m^3 with CDF matching and 0.043 m^3/m^3 with copula, which represents an improvement of a factor of 3.

Over LW, simulations responded very well to the successive rain events throughout the year and exhibited a pattern of decrease following a rain event. The first two months (March–April) exhibited more noisy simulations, and the statistics were impacted by this behavior ($R = 0.55/0.57$ and rmse $= 0.057/0.075$ m^3/m^3 for copulas/CDF matching). The other two seasons gave good results in terms of statistics. For the entire year, the R value was highly improved ($R = 0.59$ for LPRM and $R = 0.71/0.71$ for copulas/CDF matching), and the rmse was reduced by a factor of 3 (rmse $= 0.148$ m^3/m^3 for LPRM and rmse $= 0.043/0.059$ m^3/m^3 for copulas/CDF matching).

The LR watershed is the site with the highest rainfall frequency (events of small amplitude). The successive rainfall events were not well captured by the simulations, particularly during the fall season when both simulations exhibited only small variations, which resulted in very poor statistics ($R = 0.17/0.16$ for copulas/CDF matching). Unfortunately, even if the rain events were captured by the original data sets, none was captured by both data sets at the same time, so only the nonraining periods were taken into account by the statistics. Therefore, the simulations can only be representative of the dry periods. It should be noted that the statistics of LPRM were

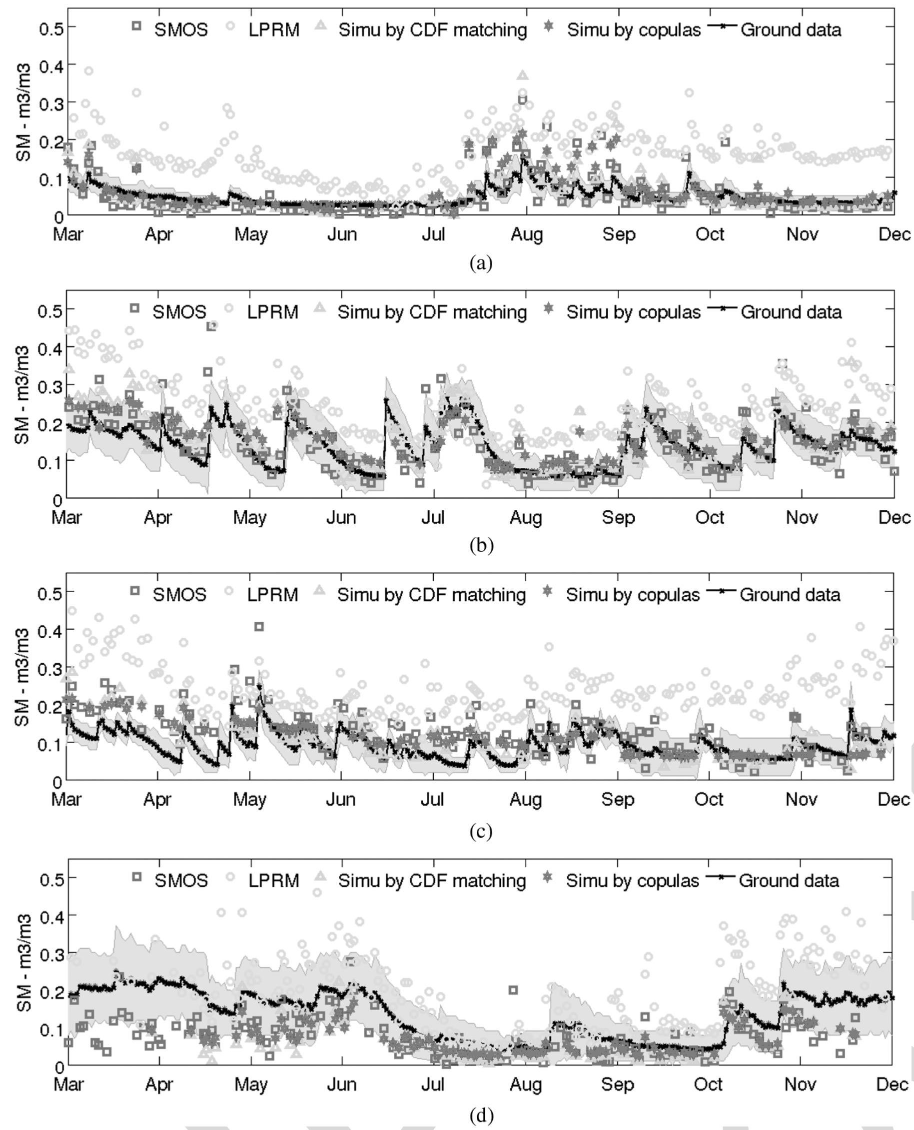


Fig. 6. Simulations for 2010: SMOS, LPRM, simulated soil moisture data from CDF matching and copulas, and ground measurements over the four watersheds. Since the *in situ* data are the mean of several ground measurements, their standard deviations are represented in gray shadows showing the spatial variability. (a) WG. (b) LW. (c) LR. (d) RC.

already not good during this season ($R = 0.37$ and rmse $=$ $0.174\ \mathrm{m}^3/\mathrm{m}^3$). During the spring season, SMOS overestimated the *in situ* soil moisture measurements, so as a result, the copulas and CDF matching estimates overestimated the *in situ* measurements as well.

RC is located in a mountainous region and is subject to frequent snow and frozen soil events. The satellite-based soil moisture was not comparable to the ground measurements until late May. After this winter period, the simulations captured accurately the soil moisture evolution and improved the original statistics and especially the rmse ($0.099\ \mathrm{m}^3/\mathrm{m}^3$ for LPRM and $0.059/0.067\ \mathrm{m}^3/\mathrm{m}^3$ for copulas/CDF matching).

B. Times Series 2003–2010 and Comparison With Ground Measurements

Soil moisture from 2003 to 2010 was simulated from the LPRM retrievals (2003–2010) using the copulas and CDF matching relationships developed for 2010. Fig. 7 and Table V show the entire time series and the associated statistics (R and rmse) between the original data, CDF matching simulations, copula simulations, and ground measurements.

WG is the driest site and did not have a lot of rain events. These rain events were well described by the simulated soil moisture even though they were sometimes largely overestimated, particularly by CDF matching simulations. Artifacts at the extremities of the seasons can be seen at the beginning of 2006 and 2008. The correlation coefficient was improved using the CDF matching for each year, whereas the errors were reduced by a factor larger than 2 with the copulas.

The overestimation of the soil moisture after the rain events with CDF matching can be found as well over LW, but the temporal evolution was well captured by both methods. For this watershed, CDF matching overestimated the high soil moisture values and underestimated the low values. CDF matching produced soil moisture with a higher dynamic range than copulas.

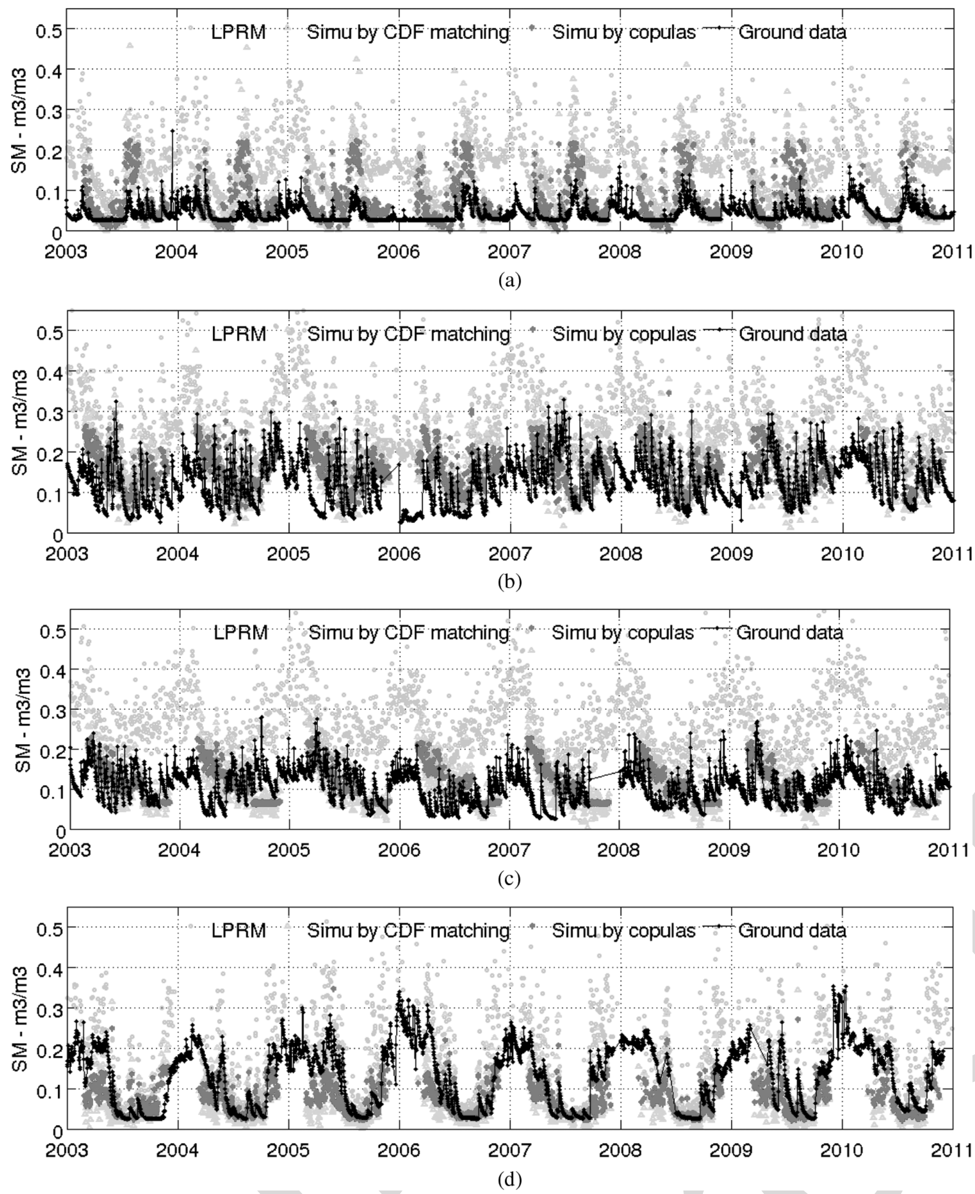


Fig. 7. Simulated time series from 2003 to 2010 with ground measurements for the four watersheds. (a) WG. (b) LW. (c) LR. (d) RC.

This was reflected in the total rmse value (0.079 m^3/m^3), whereas the rmse of the copula simulations was of 0.066 m^3/m^3 (original LPRM rmse: 0.160 m^3/m^3).

LR is the site with the largest number of rain events, and as mentioned in the previous section, this high rain frequency was not properly captured during the fall season of 2010; this can be seen as well in the entire time series where all the copulas and CDF matching estimates were flat during fall seasons. Moreover, since SMOS was overestimating the soil moisture during the spring season of 2010, both statistical estimates had this behavior. Even though the tendency of the simulations was correct, the dynamic behavior was not well represented, which resulted in a very poor correlation coefficient (negative values in 2004 and 2007).

RC is a very complicated site because of the frequent snow and frozen soil events occurring during half of the year. However, statistical results were improved for the entire year with copula simulations (rmse $= 0.099$ m^3/m^3 for LPRM and rmse $= 0.056/0.062$ m^3/m^3 for copulas/CDF matching).

VI. Conclusion and Perspectives

The main goal of this study was to propose a new method to generate a long homogeneous time series (2003–2010) of soil moisture from two overlapping time series.

For that purpose, two statistical tools, the CDF matching and the copulas, were tested over four watersheds in the U.S. By using CDF matching, the assumption that the two studied data sets are ranked in the same way is made, which the copulas do not require. The two analyzed data sets (SMOS and LPRM) were jointly available only for 2010, so data from 2010 were used to estimate the CDFs that are used as references to estimate SMOS soil moisture for previous years. The novelty of the approach is its application: establishing the statistical relationship between

TABLE V
STATISTICAL RESULTS FROM THE COMPARISON BETWEEN THE SIMULATED TIME SERIES OF SOIL MOISTURE FROM 2003 TO 2010. ORIGINAL SOIL MOISTURE TIMES ARE REPRESENTED BY LPRM. THE BEST RESULTS ARE INDICATED IN BOLD, AND THE RMSE ARE IN m^3/m^3. (a) WG. (b) LW. (c) LR. (d) RC

(a)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.070	0.76	0.82	0.66	0.81	0.68	0.65	0.79	0.73
	RMSE	0.129	0.141	0.146	0.133	0.147	0.138	0.129	0.139	0.138
Copula	R	0.62	0.55	0.82	0.64	0.81	**0.75**	0.76	0.79	0.69
	RMSE	**0.059**	**0.059**	**0.059**	**0.060**	**0.054**	**0.053**	**0.060**	**0.043**	**0.057**
CDF m.	R	**0.73**	**0.62**	**0.88**	**0.72**	**0.89**	**0.75**	**0.79**	**0.82**	**0.74**
	RMSE	0.070	0.074	0.071	0.073	0.067	0.067	0.077	0.054	0.071

(b)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.56	0.71	0.48	0.67	0.32	0.42	0.52	0.58	0.55
	RMSE	0.163	0.149	0.187	0.149	0.173	0.158	0.149	0.149	0.160
Copula	R	0.56	0.47	0.19	0.62	0.41	**0.64**	**0.58**	**0.71**	0.47
	RMSE	**0.071**	**0.064**	**0.088**	**0.077**	**0.060**	**0.056**	**0.051**	**0.044**	**0.066**
CDF m.	R	**0.59**	**0.60**	**0.34**	**0.63**	**0.49**	0.61	0.53	**0.71**	**0.51**
	RMSE	0.083	0.070	0.101	0.092	0.069	0.076	0.069	0.059	0.079

(c)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.51	0.60	0.46	0.75	0.64	0.70	0.49	0.65	0.58
	RMSE	0.171	0.148	0.181	0.185	0.180	0.166	0.187	0.178	0.174
Copula	R	0.54	-0.48	**0.73**	0.01	-0.14	0.20	0.43	0.51	0.19
	RMSE	**0.042**	**0.079**	**0.036**	**0.069**	**0.081**	0.054	**0.047**	**0.045**	**0.059**
CDF m.	R	**0.68**	**-0.16**	0.72	**0.28**	**0.18**	**0.50**	**0.55**	**0.59**	**0.37**
	RMSE	0.044	0.080	0.042	0.070	0.085	**0.050**	0.048	0.048	0.061

(d)

		2003	2004	2005	2006	2007	2008	2009	2010	Total
LPRM	R	0.78	0.76	0.74	0.80	0.84	0.69	0.78	0.73	0.77
	RMSE	0.093	0.085	0.110	0.099	0.102	0.106	0.099	0.099	0.099
Copula	R	**0.53**	**0.78**	**0.70**	**0.68**	**0.72**	0.75	**0.72**	**0.80**	**0.69**
	RMSE	**0.065**	**0.045**	**0.065**	**0.060**	**0.051**	**0.047**	**0.052**	**0.059**	**0.056**
CDF m.	R	0.42	0.69	0.65	0.63	0.70	**0.65**	0.71	0.70	0.63
	RMSE	0.073	0.051	0.070	0.063	0.055	0.056	0.056	0.067	0.062

AMSR-E and SMOS retrieved soil moisture values and using this relationship to estimate the *equivalent* SMOS value for the AMSR-E period prior to the SMOS launch.

The first analysis of these simulations over 2010 showed that the simulated data sets were very similar to the SMOS estimates and reproduced SMOS behavior accurately except over the LR watershed where numerous rain events occurred. This high rainfall frequency was interpreted statistically as noise, and hence, the simulations did not describe the soil moisture evolution over this site very well. RC was also a very complicated site due to the local topography and seasonal climate conditions. Soil moisture derived from satellite observations was not able to accurately reproduce the dynamics as found in the *in situ* data, and as a result, the simulated soil moisture did not either. However, the total rmse for the simulated soil moisture from copulas was reduced by a factor of almost 2. The WG and LW sites were well represented by the simulations, and copulas improved the error by a factor of 3, whereas CDF matching improved the correlation.

The time series of soil moisture were estimated from 2003 to 2010 and were compared to *in situ* measurements at all four watersheds. Since simulated soil moisture data in 2010 over the LR watershed had very little dynamic range, they remained the same for the entire time series and showed very poor statistical results. Even though the rmse values were improved by a factor of 3, the total correlation was not good. For the three other sites, the correlation coefficient was a bit degraded compared to the original LPRM data, but the rmse was highly improved with copulas by a factor of 2 to 3. In general, CDF matching gave better results in terms of correlation, and copulas gave better results in terms of errors compared to the ground measurements.

As a more general conclusion, CDF matching gives good results but does not take into account the structure of the dependence between the two data sets, whereas the copulas allow to model this structure. Through the choice of the family and the parameter θ (which controls the width of the tail of the scatter), it is possible to model all kinds of structures, from the perfect dependence (CDF matching), right or left dependence, to complete independence. This is why copulas produce better results for the extreme values (very low and very high values) than CDF matching. Copulas can also estimate the uncertainty of the soil moisture simulations given the LPRM value and can be seen as a quality information in the simulation process. However, the copula method is time consuming. It is quick to choose the copula family and its associated parameter as it is based on a Bayesian approach; however, it is very time consuming to generate the 1000 simulations, particularly if the chosen copula does not have an analytic inversion form. In the latter case, 1000 equations need to be resolved numerically.

Nevertheless, these simulations represent an advantage since it is possible to compute a mean and a standard deviation. The limitations are the same for both methods and even for any general statistical methods using a specific year as a reference: Only the variable range of this particular year can be well represented. Therefore, if an event in a previous year occurs and is out of the range found in the specific year of reference (such as drought or flood events), then that event will not be well represented in the simulated results.

In order to improve this methodology, applying a moving window of three months would provide more accurate results instead of dividing the year into four seasons. This would also avoid the artifacts and gaps generally noticed at the transition between the seasons. Another solution would be to introduce the time in the copulas, but the level of complexity in the copula manipulation would increase as well.

In this paper, the attempt to build a homogeneous soil moisture time series has been based on statistical methods only. Of course, other methods exist to reconcile different sensor acquisitions, and because SMOS and AMSR-E do not operate at the same frequencies and not at the same crossing times, using physical models to tackle these discrepancies is an alternative to statistical methods. Moreover, matching observations acquired at 130 am and 600 am can trigger some questions, particularly regarding the precipitations that could occur in between. The present study is a first step toward a unified and homogeneous soil moisture time series, and mixing physical and statistical models to do so would be a breakthrough for climate studies.

The next step of this study is to build a homogeneous time series of soil moisture at the global scale. Hence, the results of this study will be extended in the future to build a global map of the copula family choice and to study if there exists any relationship between the chosen copulas and the soil characteristics or land use data. This would allow us to derive soil moisture time series from LPRM data within SMOS soil moisture range over the entire globe.

ACKNOWLEDGMENT

The authors would like to thank the USDA research team for providing the *in situ* data and D. Huard from McGill University for the helpful inputs on the copula method. USDA is an Equal Employment Opportunity employer.

REFERENCES

[1] "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC," World Meteorological Org., Geneva, Switzerland, Tech. Rep., 2010.

[2] M. Drusch, "Initializing numerical weather predictions models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *J. Geophys. Res.*, vol. 113, no. D3, p. D03 102, Feb. 2007.

[3] H. Douville and F. Chauvin, "Relevance of soil moisture for seasonal climate predictions: A preliminary study," *Climate Dyn.*, vol. 16, no. 10/11, pp. 719–736, Oct. 2000.

[4] Y. Kerr, P. Waldteufel, J. Wigneron, S. Delwart, F. Cabot, J. Boutin, M. Escorihuela, J. Font, N. Reul, C. Gruhier, S. Juglea, M. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.

[5] D. Entekhabi, E. Njoku, P. O'Neill, K. Kellogg, W. Crow, W. Edelstein, J. Entin, S. Goodman, T. Jackson, J. Johnson, J. Kimball, J. Piepmeier, R. Koster, N. Martin, K. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. Shi, M. Spencer, S. Thurman, L. Tsang, and J. Zyl, "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[6] Y. Kerr, P. Waldteufel, J. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.

[7] M. Owe, R. de Jeu, and J. Walker, "A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1643–1654, Aug. 2001.

[8] E. Njoku, T. Jackson, V. Lakshmi, T. Chan, and S. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215–229, Feb. 2003.

[9] L. Li, P. Gaiser, B. Gao, R. Bevilacqua, T. Jackson, E. Njoku, C. Rüdiger, J. Calvet, and R. Bindlish, "Windsat global soil moisture retrieval and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2224–2241, May 2010.

[10] V. Naeimi, K. Scipal, Z. Bartalis, and S. H. W. Wagner, "An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 1999–2013, Jul. 2009.

[11] L. Vincent, X. Zhang, B. Bonsal, and W. Hogg, "Homogeneization of daily temperatures over Canada," *J. Climate*, vol. 15, no. 11, pp. 1322–1334, Jun. 2002.

[12] M. Begert, T. Schlegel, and W. Kirchhofer, "Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000," *Int. J. Climatol.*, vol. 25, no. 1, pp. 65–80, Jan. 2005.

[13] G. Picard and M. Fily, "Surface melting observations in Antarctica by microwave radiometers: Correcting 26-year time series from changes in acquisition hours," *Remote Sens. Environ.*, vol. 104, no. 3, pp. 325–336, Oct. 2006.

[14] R. Reichle and R. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, no. 19, p. L19 501, Oct. 2004.

[15] M. Choi and J. Jacobs, "Temporal variability corrections for advanced microwave scanning radiometer E (AMSR-E) surface soil moisture: Case study in little river region, Georgia, U.S.," *Sensors*, vol. 8, no. 4, pp. 2617–2627, Apr. 2008.

[16] H. Li, J. Sheffield, and E. Wood, "Bias correction of monthly precipitation and temperature fields from IPCC AR4 models using equidistant quantile matching," *J. Geophys. Res., Atmosp.*, vol. 115, no. D10, p. D10 101, May 2010.

[17] M. Drusch, E. Wood, and H. Gao, "Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture," *Geophys. Res. Lett.*, vol. 32, no. 15, p. L15 403, Aug. 2005.

[18] Y. Liu, A. van Dijk, R. de Jeu, and T. Holmes, "An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia," *Water Resour. Res.*, vol. 45, no. 7, p. W07 405, Jul. 2009.

[19] Y. Liu, R. Parinussa, W. Dorigo, R. D. Jeu, W. Wagner, A. V. Dijk, M. McCabe, and J. Evans, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 2, pp. 425–436, Feb. 2011.

[20] V. Singh and W. Strupczewski, "Editorial," *J. Hydrol. Eng.*, vol. 12, no. 4, p. 345, Jul. 2007.

[21] R. Nelsen, "An introduction to copulas," in *Springer Series in Statistics*. New York, NY, USA: Springer-Verlag, 1998.

[22] P. Trivedi and D. Zimmer, *Copula Modeling: An Introduction for Practitioners*, vol. 1, *Foundations and Trends in Econometrics*. Hanover, MA, USA: Now Publ. Inc., 2005, pp. 1–111.

[23] C. Hafner and O. Reznikova, "Efficient estimation of a semiparametric synamic copula model," *Comput. Stat. Data Anal.*, vol. 54, no. 11, pp. 2609–2627, Nov. 2010.

[24] C. Genest and A. Favre, "Everything you always wanted to know about copula modeling but were afraid to ask," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 347–368, Jul./Aug. 2007.

[25] A. Favre, S. E. Adlouni, L. Perreault, N. Thiémonge, and B. Bobée, "Multivariate hydrological frequency analysis using copulas," *Water Resour. Res.*, vol. 40, no. 1, p. W01 101, Jan. 2004.

[26] G. Salvadori and C. de Michele, "On the use of copulas in hydrology: Theory and practice," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 369–380, Jul. 2007.

[27] D. Dupuis, "Using Copulas in hydrology: Benefits, cautions and issues," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 381–393, Jul. 2007.

[28] L. Zhang and V. Singh, "Trivariate flood frequency analysis using the Gumbel-Hougaard copula," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 431–439, Jul. 2007.

[29] F. Serinaldi and S. Grimaldi, "Fully nested 3-copula: Procedure and application on hydrological data," *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 420–430, Jul. 2007.

[30] P. Laux, S. Vogl, W. Qiu, H. Knoche, and H. Kunstmann, "Copula-based statistical refinement of precipitation in RCM simulations over complex terrain," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 7, pp. 2401–2419, Jul. 2011.

[31] H. Gao, E. Wood, M. Drusch, and M. McCabe, "Copula-derived observation operators for assimilating TMI and AMSR-E retrieved soil moisture into land surface models," *J. Hydrometreol.*, vol. 8, no. 3, pp. 413–429, Jun. 2007.

[32] C. de Michele and G. Salvadori, "A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas," *J. Geophys. Res.*, vol. 108, no. D2, p. 4067, Jan. 2003.

[33] J. Wigneron, Y. Kerr, P. Waldteufel, K. Saleh, M. Escorihuela, P. Richaume, P. Ferrazzoli, P. de Rosnay, R. Gurney, J. Calvet, J. Grant, M. Guglielmetti, B. Hornbuckle, C. Mätzler, T. Pellarin, and M. Schwank, "L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields," *Remote Sens. Environ.*, vol. 107, no. 4, pp. 639–655, Apr. 2007.

[34] Y. Kerr, P. Waldteufel, P. Richaume, J. Wigneron, P. Ferrazzoli, A. Mahmoodi, A. A. Bitar, F. Cabot, C. Gruhier, S. Juglea, D. Leroux, A. Mialon, and S. Delwart, "The SMOS soil moisture retrieval algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1384–1403, May 2012.

[35] D. Carr, R. Kahn, K. Sahr, and T. Olsen, "ISEA discrete global grids," *Stat. Comput. Stat. Graph. Newl.*, vol. 8, no. 2/3, pp. 31–39, 1997.

[36] A. A. Bitar, D. Leroux, Y. Kerr, O. Merlin, P. Richaume, A. Sahoo, and E. Wood, "Evaluation of SMOS soil moisture products over continental US using SCAN/SNOTEL network," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1572–1586, May 2012.

[37] G. Schaefer, M. Cosh, and T. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *J. Atmosp. Ocean. Technol.*, vol. 24, no. 12, pp. 2073–2077, Dec. 2007.

[38] T. Jackson, R. Bindlish, M. Cosh, T. Zhao, P. Starks, D. Bosch, M. Seyfried, S. Moran, Y. Kerr, and D. Leroux, "Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1530–1543, May 2012.

[39] D. Leroux, Y. Kerr, A. A. Bitar, C. Gruhier, R. Bindlish, T. Jackson, B. Berthelot, and G. Portet, "Comparison between SMOS, VUA, ASCAT and ECMWF soil moisture products over four watersheds in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, submitted for publication. AQ27

[40] J. Sabatier, A. Fouilloux, and P. de Rosnay, "Technical implementation of SMOS data in the ECMWF integrated forecast system," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 252–256, Mar. 2012.

[41] [Online]. Available: ftp://n4ftl01u.ecs.nasa.gov/SAN/AMSA/

[42] C. Gruhier, P. de Rosnay, Y. Kerr, E. Mougin, E. Ceschia, J. Calvet, and P. Richaume, "Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions," *Geophys. Res. Lett.*, vol. 35, no. 10, p. L10 405, May 2008.

[43] C. Rudiger, J. C. Calvet, C. Gruhier, T. Holmes, R. de Jeu, and W. Wagner, "An intercomparison of ERS-SCAT and AMSR-E soil moisture observations with model simulations over France," *Amer. Meteorol. Soc.*, vol. 10, no. 2, pp. 431–447, Apr. 2009.

[44] C. Draper, J. Walker, P. Steinle, R. de Jeu, and T. Holmes, "An evaluation of AMSR-E derived soil moisture over Australia," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 703–710, Apr. 2009.

[45] C. Gruhier, P. de Rosnay, S. Hasenauer, T. Holmes, R. de Jeu, Y. Kerr, E. Mougin, E. Njoku, F. Timouk, W. Wagner, and M. Zribi, "Soil moisture active and passive microwave products: Intercomparison and evaluation over a Sahelian site," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 1, pp. 141–156, Jan. 2010.

[46] S. Chaurasia, D. Tung, P. Thapliyal, and P. Joshi, "Assessment of AMSR-E soil moisture product over India," *Int. J. Remote Sens.*, vol. 32, no. 23, pp. 7955–7970, Dec. 2011.

[47] T. Jackson, M. Cosh, R. Bindlish, P. Starks, D. Bosch, M. Seyfried, D. Goodrich, M. Moran, and J. Du, "Validation of advanced microwave scanning radiometer soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4256–4272, Dec. 2010.

[48] B. Schweizer and E. Wolff, "On nonparametric measures of dependence for random variables," *Ann. Stat.*, vol. 9, no. 4, pp. 879–885, Jul. 1981.

[49] Y. Wang and H. Pham, "Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 13–22, Mar. 2012.

[50] C. Genest and L.-P. Rivest, "Statistical inference procedures for bivariate Archimedean copulas," *J. Amer. Stat. Assoc.*, vol. 88, no. 423, pp. 1034–1043, Sep. 1993.

[51] J. Fermanian, "Goodness of fit tests for copulas," *J. Multivariate Anal.*, vol. 95, no. 1, pp. 119–152, Jul. 2005.

[52] C. Genest, J.-F. Quessy, and B. Rémillard, "Goodness-of-fit procedures for copula models based on the probability integral transform," *Scand. J. Stat.*, vol. 33, no. 2, pp. 337–366, Jun. 2006.

[53] C. Genest and B. Rémillard, "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models," *Annales Henri Poincaré*, vol. 44, no. 6, pp. 1096–1127, 2008.

[54] D. Berg, "Copula goodness-of-fit testing: An overview and power comparison," *Eur. J. Finance*, vol. 15, no. 7/8, pp. 675–701, 2009.

[55] C. Genest, B. Rémillard, and D. Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Math. Econom.*, vol. 44, no. 2, pp. 199–213, Apr. 2009.

[56] D. Huard, G. Evin, and A. Favre, "Bayesian copula selection," *Comput. Stat. Data Anal.*, vol. 51, no. 2, pp. 809–822, Nov. 2006.

Delphine J. Leroux received the M.S. degree in applied mathematics from Institut National des Sciences Appliquees, Toulouse, France, in 2009 and the Ph.D. degree in spatial hydrology from Universite Paul Sabatier, Toulouse, in 2012.

From 2009 to 2012, she was with the Centre d'Etudes Spatiales de la Biosphere, Toulouse, where she worked on the validation of the Soil Moisture and Ocean Salinity soil moisture product at the local and global scales by using statistical and physical methods. She is currently with the Jet Propulsion Laboratory (JPL), Pasadena, CA, USA, on the Soil Moisture Active Passive mission. AQ28

Yann H. Kerr (M'88–SM'01–F'13) received the engineering degree from Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, the M.Sc. degree in electronics and electrical engineering from Glasgow University, Glasgow, U.K., and the Ph.D. degree in astrophysique geophysique et techniques spatiales from Université Paul Sabatier, Toulouse. AQ29

He is currently the Director of the Centre d'Etudes Spatiales de la Biosphere, Toulouse. His fields of interest are in the theory and techniques for microwave and thermal infrared remote sensing of the Earth, with emphasis on hydrology and water resource management. He was an EOS Principal AQ30 AQ31 Investigator (interdisciplinary investigations) and PI and precursor of the use AQ32 of the SCAT over land. In 1990, he started to work on the interferometric AQ33 concept applied to passive microwave Earth observation and was subsequently the Science Lead on the MIRAS project for ESA. In 1997, he proposed the AQ34 AQ35 Soil Moisture and Ocean Salinity (SMOS) mission, the natural outcome of the previous MIRAS work. He is currently involved in the exploitation of SMOS data, in the Cal Val activities and related level 2 soil moisture and level 3 and 4 AQ36 developments. He is also working on the SMOS next concept.

Dr. Kerr received the World Meteorological Organization 1st prize (Norbert AQ37 Gerbier), the USDA Secretary's team award for excellence (Salsa Program), and the GRSS certificate of recognition for leadership in the development of AQ38 the first synthetic aperture microwave radiometer in space and success of the SMOS mission and is a Distinguished Lecturer for GRSS. AQ39

Eric F. Wood received the B.A.Sc. degree in civil engineering from the University of British Columbia, Vancouver, BC, Canada, in 1970 and the S.M., C.E., and Sc.D. degrees in civil engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1972, 1973, and 1974, respectively.

He is currently a Professor in the Department of Civil and Environmental Engineering, Princeton University, NJ, USA. His recent contributions include the macroscale hydrologic prediction of the coupled water and energy balances of the land surface, remote sensing as an integral tool for observation and modeling of the hydrologic cycle, and developing and evaluating seasonal hydrological forecasts based on coupled seasonal climate forecasts.

Dr. Wood is a fellow of the American Geophysical Union and the American Meteorological Society. He received, among other honors, the Robert E. Horton Award from the Hydrology Section of the American Geophysical Union (1977), the John Dalton Medal from the European Geophysical Union (2007), and the Jules G. Charney Award from the American Meteorological Society (2010).

Alok K. Sahoo received the Ph.D. degree in computational sciences from George Mason University, Fairfax, VA, USA, in 2008.

After finishing his Ph.D., he worked as a Research Associate at the Institute of Global Environment and Society, Beltsville, MD, USA, from 2008 to 2009. He is currently an Associate Research Hydrologist with the Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA. His research interests involve the application of microwave remote sensing in hydrologic applications. His current research projects include soil moisture estimation from microwave sensors and checking consistency among hydrologic cycle variables for drought monitoring.

Rajat Bindlish (S'98–AM'99–M'03–SM'05) received the B.S. degree in civil engineering from the Indian Institute of Technology, Bombay in 1993 and AQ40 the M.S. and Ph.D. degrees in civil engineering from The Pennsylvania State University in 1996 and 2000, AQ41 respectively.

He is currently with SSAI, working at USDA AQ42 Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. His research interests involve the application of microwave remote sensing in hydrology. He is currently working on soil moisture estimation from microwave sensors and their subsequent application in land surface hydrology.

Thomas J.Jackson (SM'96–F'02) received the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 1976.

He is a Research Hydrologist with the U.S. Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. His research involves the application and development of remote sensing technology in hydrology and agriculture, primarily the microwave measurement of soil moisture.

Dr. Jackson is or has been a member of the science and validation teams of the Aqua, ADEOS-II, Radarsat, Oceansat-1, Envisat, AQ43 ALOS, Soil Moisture and Ocean Salinity, Aquarius, GCOM-W, and Soil AQ44 AQ45 Moisture Active Passive remote sensing satellites. He is a fellow of the Society of Photo-Optical Instrumentation Engineers, the American Meteorological Society, and the American Geophysical Union. In 2003, he received the William T. Pecora Award (NASA and Department of Interior) for outstanding contributions toward understanding the Earth by means of remote sensing and the AGU Hydrologic Sciences Award for outstanding contributions to the AQ46 science of hydrology. He received the IEEE Geoscience and Remote Sensing Society Distinguished Achievement Award in 2011.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Please be aware that the authors are required to pay overlength page charges ($200 per page) if the paper is longer than 6 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

AQ1 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ2 = Please provide the expanded form of “USDA.”
AQ3 = Please provide the expanded form of “ARS.”
AQ4 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ5 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ6 = “AMSR-E” is defined as “Advanced Microwave Scanning Radiometer-Earth Observing System” for consistency. Please check if appropriate, and correct if necessary.
AQ7 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ8 = “Picard *et al.*” was changed to “Picard and Fily.” Please check if appropriate, and correct if necessary."
AQ9 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ10 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ11 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ12 = The sentence was restructured for clarity. Please check if the original thought was retained, and correct if necessary.
AQ13 = “rmse” is defined as “root-mean-square error.” Please check if appropriate, and correct if necessary."
AQ14 = The sentence was rephrased. Please check if the original thought was retained, and correct if necessary."
AQ15 = Please provide the expanded form of “NASA.”
AQ16 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ17 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ18 = The caption was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ19 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ20 = “CDF” is defined as “cumulative density function” for consistency. Please check if appropriate, and correct if necessary.
AQ21 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.

AQ22 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ23 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ24 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ25 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ26 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ27 = Please provide publication update in Ref. [39].
AQ28 = The author's current affiliation indicated in the footnote did not correspond to the current affiliation provided in the curriculum vitae. Please check.
AQ29 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ30 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ31 = Please provide the expanded form of "EOS."
AQ32 = Please provide the expanded form of "PI."
AQ33 = Please provide the expanded form of "SCAT."
AQ34 = Please provide the expanded form of "MIRAS."
AQ35 = Please provide the expanded form of "ESA."
AQ36 = Please provide the expanded form of "Cal Val."
AQ37 = "Dr." was inserted as the title for author Yann H. Kerr. Please check if appropriate, and correct if necessary.
AQ38 = Please provide the expanded form of "GRSS."
AQ39 = The sentence was rephrased for clarity. Please check if the original thought was retained, and correct if necessary.
AQ40 = Please provide the specific location of the Indian Institute of Technology, Bombay.
AQ41 = Please provide the specific location of The Pennsylvania State University.
AQ42 = Please provide the expanded form of "SSAI."
AQ43 = Please provide the expanded form of "ADEOS-II."
AQ44 = Please provide the expanded form of "ALOS."
AQ45 = Please provide the expanded form of "GCOM-W."
AQ46 = Please provide the expanded form of "AGU."

END OF ALL QUERIES