ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Information theoretic evaluation of satellite soil moisture retrievals

Sujay V. Kumar[a,*], Paul A. Dirmeyer[b], Christa D. Peters-Lidard[c], Rajat Bindlish[a], John Bolten[a]

[a] Hydrological Sciences Laboratory, NASA GSFC, Greenbelt, MD, United States
[b] George Mason University, Fairfax, VA, United States
[c] Earth Sciences Division, NASA GSFC, Greenbelt, MD, United States

ARTICLE INFO

Keywords:
Soil moisture
Remote sensing
Information theory

ABSTRACT

Microwave radiometry has a long legacy of providing estimates of remotely sensed near surface soil moisture measurements over continental and global scales. A consistent assessment of the errors and uncertainties associated with these retrievals is important for their effective utilization in modeling, data assimilation and end-use application environments. This article presents an evaluation of soil moisture retrieval products from AMSR-E, ASCAT, SMOS, AMSR2 and SMAP instruments using information theory-based metrics. These metrics rely on time series analysis of soil moisture retrievals for estimating the measurement error, level of randomness (entropy) and regularity (complexity) of the data. The results of the study indicate that the measurement errors in the remote sensing retrievals are significantly larger than that of the ground soil moisture measurements. The SMAP retrievals, on the other hand, were found to have reduced errors (comparable to those of in-situ datasets), particularly over areas with moderate vegetation. The SMAP retrievals also demonstrate high information content relative to other retrieval products, with higher levels of complexity and reduced entropy. Finally, a joint evaluation of the entropy and complexity of remotely sensed soil moisture products indicates that the information content of the AMSR-E, ASCAT, SMOS and AMSR2 retrievals is low, whereas SMAP retrievals show better performance. The use of information theoretic assessments is effective in quantifying the required levels of improvements needed in the remote sensing soil moisture retrievals to enhance their utility and information content.

1. Introduction

Soil moisture plays an important role in modulating the exchanges of water and energy at the land atmosphere interface and profoundly influences the spatial and temporal variability of weather and climatic conditions (Koster et al., 2004; Seneviratne et al., 2010). Accurate characterization of soil moisture is, therefore, important for applications such as flood/drought forecasting, weather and climate modeling, agricultural and water resources management. Observations of soil moisture from ground measurements tend to be sparse and are often not sufficient to capture the spatial heterogeneity and variability of soil moisture at larger spatial scales, required for such applications. Spaceborne measurements of soil moisture, primarily from microwave (MW) remote sensing, provide an alternative for developing observations of soil moisture over larger spatial extents (Jackson, 1993; Njoku and Entekhabi, 1995). In the past several decades, near surface soil moisture retrievals have become available from a number of low-frequency (C, X, Ku- and L-band) passive and active microwave sensors (Wagner et al., 2003; Njoku et al., 2003; Wen et al., 2003; Owe et al., 2008; Kerr et al., 2010; Entekhabi et al., 2010).

Microwave soil moisture sensors exploit the fact that the emission of the land surface is affected by variables such as surface temperature, roughness, vegetation and soil moisture. The influence of soil moisture is most prominent at low frequencies ($\sim 10 - 1$ GHz, making it the ideal range of satellite remote sensing (Njoku and Kong, 1977; Jackson et al., 1982; Ulaby et al., 1986). Unlike the visible and infrared sensors, the microwave sensors are not limited by cloud cover and nighttime conditions. The observations can be made at any time of the day and are not dependent on solar illumination (Jackson et al., 1996). Longer wavelengths (L-band; 1–2 GHz) also allow for deeper penetration into the soil and reduce the influence of vegetation in attenuating the soil moisture signal (Jackson et al., 1982). The active instruments can provide measurements at higher spatial resolutions than the passive microwave instruments, though radar systems are more strongly affected by the local topography, roughness and vegetation than passive radiometer systems (Entekhabi et al., 2010; Lakshmi, 2013). However, studies such as Brocca et al. (2011) have suggested that ASCAT can outperform passive microwave based retrievals over areas with

* Corresponding author.
E-mail address: sujay.v.kumar@nasa.gov (S.V. Kumar).

http://dx.doi.org/10.1016/j.rse.2017.10.016
Received 23 April 2017; Received in revised form 1 October 2017; Accepted 13 October 2017
0034-4257/

moderate vegetation. Passive observations on the other hand, are more impacted by spatial heterogeneity and scaling effects because of poor spatial resolution. The spatial resolution of the passive microwave soil moisture observations is typically coarse (~25 to 50 km), with the satellite footprint size increasing with wavelength and altitude. The presence of snow cover, frozen soil and precipitation events also limits the skill of the soil moisture retrievals (Parinussa et al., 2011).

Due to the differences in the spatial and temporal span of different MW instruments and due to the limited availability of reliable ground measurements, a consistent evaluation of soil moisture remote sensing datasets is difficult. Land surface model climatology has often been used the reference to address the climatological differences between different retrievals when developing multi-sensor products (Liu et al., 2011b) and for consistent evaluations of multiple products. In a recent study, Kumar et al. (2015) has shown that such approaches lead to the loss of valuable signals and cause the statistical properties of the retrieval products to be similar to that of the reference datasets. Therefore, performance measures not reliant on the availability of ancillary soil moisture data can be useful for characterizing and assessing the quality of the soil moisture retrieval datasets. As a result, studies have used indirect approaches such as triple collocation (TC; Stoffelen, 1998; Dorigo et al., 2010) and spectral fitting (SF; Su et al., 2014) to assess the relative quality of global soil moisture retrievals. TC comparisons involve three different soil moisture products (often a mix of satellite soil moisture retrievals and land surface model estimates), with assumptions of linearity (between the true soil moisture and observations), signal and error stationarity, error orthogonality and independence of errors in the constituent datasets (Gruber et al., 2016b). Recent studies have examined the applicability of these assumptions for soil moisture datasets (Yilmaz and Crow, 2014) and have proposed enhancements to address the limitations imposed by these assumptions, making it a powerful method for global soil moisture evaluation (Zwieback et al., 2013; Gruber et al., 2016a,b). The SF error estimator, based on the method developed by Su et al. (2013) for de-noising satellite soil moisture datasets, estimates the stochastic random errors by comparing the spectral properties of a given soil moisture time series and a linearized water balance model. This method also does not require ancillary datasets and was shown to provide error estimates comparable to those from TC.

Similar to these stand-alone assessment methods, here we present the use of information theoretic and autoregressive analysis of time series data for quantifying errors and information content of remote sensing retrieval datasets from a number of recent soil moisture missions. Information theory measures, originally proposed by Shannon (1948), consider the stochasticity in time series data as sources of information. A key information theoretic measure is entropy, which quantifies the information content or randomness associated with the probability distribution of the data. Similarly, temporal measures of complexity rooted in information theory can be used to discriminate datasets based on time series complexity. Entropy and complexity provide separate measures of information by characterizing the randomness and state changes within a given time series of the data. Entropy is a measure of uncertainty, which is low for periodic sequences and high for random processes. On the other hand, complexity is a measure that is low for both periodic and random sequences, but high for sequences that are not easy to describe with a minimal set of parameters (Lange, 1999). Such measures have been employed for comparing model outputs of soil moisture (Pachepsky et al., 2006), space-borne soil moisture retrievals (Nearing et al., 2017), runoff and precipitation measurements from different catchment systems (Lange, 1999; Hauhs and Lange, 2008) and ecological systems (Parrott, 2010). A key advantage of information theoretic methods is that they enable the quantification of hidden patterns and structures of the data without requiring ancillary or independent data.

In addition to the use of information theoretic measures, we also employ time series red noise spectrum analysis to develop estimates of accuracy. Vinnikov et al. (1996) employed a first-order Markov process model framework to evaluate observational soil moisture data, which was extended by Dirmeyer et al. (2016) in a recent study to compare measurement errors from different in-situ soil moisture observational networks. Here we apply this method for comparing measurement errors associated with remote sensing soil moisture retrievals. Similar to the information theoretic measures, a key advantage of this approach is that it does not require specific validation or independent reference data. The simultaneous development of information theoretic and measurement error estimates allows the comparison of associated tradeoffs in accuracy, uncertainty and complexity.

The article is organized as follows: Section 2 presents the details of the datasets and the evaluation approaches. The application of the information theory methods to the remote sensing soil moisture retrievals is described in Section 3. Section 4 provides a summary and discussion of the major conclusions of this study.

2. Approach

2.1. Methods

The information theoretic measures are developed by treating the time series data as a symbol sequence with a finite number of states. The standard approach is to categorize the time series data into a binary string ("symbols") (Lange, 1999; Pachepsky et al., 2006), by encoding values above and below the median (for time series at each grid point), as 1 and 0, respectively. The entropy and complexity measures are then computed based on the probabilities of observing patterns of states/words (a group of L consecutive symbols) within the sequence. In this article, we use three symbol states ($L = 3$), consistent with prior studies (Pachepsky et al., 2006; Pan et al., 2011). These include the probability of occurrence of a given state i ($p_{L,i}$) as well as the second order probability ($p_{L,ij}$) of observing state i next to j. For binary symbol sequences, there are 2^L possible words of length L. (For example, if an encoded symbol string starts as '0011', then the first word is '001', which transitions to the second word '011' and so on.)

Shannon entropy is the expected value of the information contained in a symbol sequence. The metric entropy is specified as the normalized measure of Shannon entropy for states of size L and is defined as:

$$H(L) = -\frac{\sum_{i=1}^{2^L} p_{L,i} \log_2 p_{L,i}}{L} \quad (1)$$

$H(L)$ ranges between 0 (for constant sequences) and 1 (for uniformly distributed random sequences).

The fluctuation complexity (Bates and Shephard, 1993), which measures the spread between information within a symbol string between consecutive states is expressed as:

$$C(L) = \sum_{i,j}^{2^L} p_{L,ij} \left(\log_2 \frac{p_{L,i}}{p_{L,j}} \right)^2 \quad (2)$$

$C(L)$ can be thought of as a measure of the ordering of states within a symbol sequence, with high and low values associated with complex and simple orderings, respectively. The fluctuation complexity, therefore, is a measure of the extent of the changes in information gain or loss in a time series and it approaches zero for signals with limited probable states (Pan et al., 2011).

Note that both the choice of the classification and the length of the words have an impact on the metrics that are computed. The use of a finer classifications (rather than wet and dry) and the use of larger number of words enables a more granular detection of the entropy and complexity measures, but requires longer and consistent time series. Though the use of the three-symbol states in this study limits the granularity of the soil moisture changes detected by the information theory measures, they are helpful in examining the general trends