AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2014GL060017

Key Points:
- Filtered surface soil moisture can match information from vertical integration
- Matching degree depends on a complex set of process-level consideration
- Microwave retrievals are less constrained by vertical support than assumed

Correspondence to:
W. T. Crow,
Wade.Crow@ars.usda.gov

Citation:
Qiu, J., W. T. Crow, G. S. Nearing, X. Mo, and S. Liu (2014), The impact of vertical measurement depth on the information content of soil moisture times series data, *Geophys. Res. Lett.*, *41*, 4997–5004, doi:10.1002/2014GL060017.

Received 25 APR 2014
Accepted 10 JUN 2014
Accepted article online 14 JUN 2014
Published online 25 JUL 2014

The impact of vertical measurement depth on the information content of soil moisture times series data

Jianxiu Qiu[1,2], Wade T. Crow[2], Grey S. Nearing[3], Xingguo Mo[1], and Suxia Liu[1]

[1]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, [2]USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA, [3]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Abstract Using a decade of ground-based soil moisture observations acquired from the United States Department of Agriculture's Soil Climate Analysis Network (SCAN), we calculate the mutual information (MI) content between multiple soil moisture variables and near-future vegetation condition to examine the existence of emergent drought information in vertically integrated (surface to 60 cm) soil moisture observations ($\theta_{0\text{–}60\ [cm]}$) not present in either superficial soil moisture observations ($\theta_{5\ [cm]}$) or a simple low-pass transformation of θ_5. Results suggest that while $\theta_{0\text{–}60}$ is indeed more valuable than θ_5 for predicting near-future vegetation anomalies, the enhanced information content in $\theta_{0\text{–}60}$ soil moisture can be effectively duplicated by the low-pass transformation of θ_5. This implies that, for drought monitoring applications, the shallow vertical penetration depth of microwave-based θ_5 retrievals does not represent as large a practical limitation as commonly perceived.

1. Introduction

The shallow vertical penetration depth of microwave surface soil moisture retrievals is commonly assumed to represent a significant limitation for ecosystem and agricultural productivity forecasting applications [*Li and Islam*, 1999; *Albergel et al*., 2008]. While the exact vertical support of these retrievals varies according to retrieval frequency and vertical moisture content [*Escorihuela et al*., 2010; *Jackson et al*., 2010], it is widely assumed that they capture soil moisture conditions in a thin (≤ 5 cm) upper region of the soil column which represents only a small fraction of the total vegetation rooting depth (typically assumed to be somewhere between 30 and 120 cm [*Arora and Boer*, 2003]). As a result, a number of solutions have been proposed to extrapolate information contained in remotely sensed surface retrievals ($\theta_{5\ [cm]}$) to estimate vertically integrated soil moisture ($\theta_{0\text{-}D}$ where $D > 5$ [cm]). Among the most popular are data assimilation approaches which assimilate θ_5 retrievals into a multilayer soil water balance model and update the full soil moisture profile using Kalman filtering concepts [*Reichle et al*., 2007; *Sabater et al*., 2007; *Reichle et al*., 2008; *Kumar et al*., 2009]. In addition, *Wagner et al*. [1999] proposed a simple low-pass filter to transform θ_5 time series into a root-zone soil moisture proxy commonly referred to as the soil water index (SWI). Time series of SWI generated from remotely sensed θ_5 have been shown to correlate well with vertically integrated soil moisture estimates acquired from both in situ observations and water balance modeling [*Ceballos et al*., 2005; *Albergel et al*., 2008; *Brocca et al*., 2011; *Mo et al*., 2011] and vegetation growing conditions on a monthly time scale [*Zribi et al*., 2010].

However, relatively little work has been done to explicitly quantify the effect of vertical integration depth on the information content of soil moisture estimates. Here we attempt to quantify the emergent information in vertically integrated soil moisture not present in superficial soil moisture by calculating the mutual information (MI) content between near-future vegetation index anomalies and anomalies of superficial soil moisture (θ_5), vertically integrated soil moisture ($\theta_{0\text{-}D}$), and a low-pass transformation of θ_5 (SWI) obtained at long-term (>10 years) soil moisture measurements sites within the United States. Differences in MI content (with respect to near-future vegetation conditions) among these three products are then used to clarify the degree to which limited vertical measurement support may reduce the value of microwave-based surface soil moisture retrievals for drought applications.

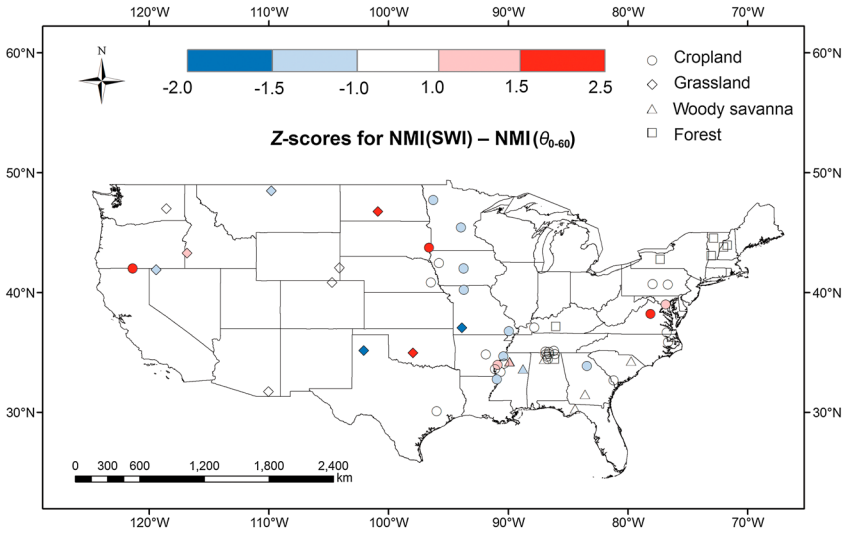

Figure 1. Location of the 61 Soil Climate Analysis Network (SCAN) sites within the contiguous United States used in the analysis. Color shading reflects *Z*-scores for NMI(SWI) minus NMI(θ_{0-60}) differences observed at each site (i.e., the raw difference scaled by its 1σ sampling error) at 16 day lag. Symbols represent different dominant land cover types surrounding each SCAN site (at a 500 m scale).

2. Data and Methods

2.1. Soil Moisture Measurements

Daily soil moisture observations at depths of 5, 10, 20, 50, and 100 cm from 2001 to 2012 were collected from the United States Department of Agriculture's (USDA) Soil Climate Analysis Network (SCAN) (see http://www.wcc.nrcs.usda.gov/scan/). To reduce temporal sampling errors, only the 61 SCAN sites with over 10 years of observations were considered. These sites are located in a variety of climate zones within the contiguous United States. While SCAN sensors are typically installed in 10–50 m^2 plots containing short grass, adjacent landscape-scale land cover properties vary widely between sites. Figure 1 shows a map of all 61 SCAN sites using different symbols (based on the 500 m MODIS MCD12Q1 land cover classification) to represent landscape-scale land cover surrounding each plot-scale site.

Soil moisture measurements at the 5 cm depth (θ_5) were used to represent superficial soil moisture variations obtained from microwave remote sensing. Vertically integrated soil moisture at time t was calculated by linearizing the SCAN soil moisture profile and integrating this linearization between the surface and a 60 cm soil depth:

$$\theta_{0-60,t} = \frac{\theta_{5,t}\cdot d_1 + \frac{\theta_{5,t}+\theta_{10,t}}{2}\cdot(d_2-d_1) + \frac{\theta_{10,t}+\theta_{20,t}}{2}\cdot(d_3-d_2) + \frac{\theta_{20,t}+\theta_{50,t}}{2}\cdot(d_4-d_3) + \frac{\theta_{50,t}+\theta^*_{60,t}}{2}\cdot(60-d_4)}{60} \quad (1)$$

where d_i [cm] is the depth of the ith soil moisture measurement (starting from the surface); θ_d is the SCAN soil moisture measurement acquired at depth d (in units cm); and θ^*_{60} is a linearly interpolated value at 60 cm using θ_{50} and θ_{100}. In cases where θ_{100} measurements are missing, θ^*_{60} is set equal to θ_{50}.

The resulting θ_{0-60} time series reflect mean soil moisture content between the land surface and a 60 cm soil depth. Our fixed choice of a 60 cm integration depth is based on practical considerations (i.e., it represents the deepest depth that can be measured at all SCAN sites) and the fact that it approximates the midpoint of the rooting-depth range commonly assumed for various land cover types [*Arora and Boer*, 2003]. In reality, rooting depths for vegetation in the vicinity of each SCAN site vary widely as a function of climate, species composition, crop growth state, and the vertical composition of soils (among many other factors). A more detailed vertical-integration approach can be applied if site-specific soil texture, root distribution, and vegetation information is available [*Pollacco and Mohanty*, 2012]; however, that is not the case for this particular analysis. As a result, the θ_{0-60} time series obtained in (1) should not be equated with root-zone soil moisture. Instead, it describes only a vertically integrated soil moisture value that (should) provide a better approximation of true root-zone conditions than θ_5.

2.2. NDVI Data

Vegetation condition was assessed using Normalized Difference Vegetation Index (NDVI) retrievals. Our choice of NDVI over leaf area index (LAI) or fraction of photosynthetically active radiation (FPAR) as a target vegetation index was based on the successful application of NDVI in previous soil moisture/vegetation coupling studies [*Bolten and Crow*, 2012] and known issues with MODIS LAI/FPAR product accuracy [*De Kauwe et al*., 2011]. To calculate NDVI, 8 day composites of atmospherically corrected surface reflectance (*SR*) data for the 620–670 nm (b1) and 841–876 nm (b2) bands were obtained from the 250 m Moderate Resolution Imaging Spectroradiometer (MODIS) MOD09Q1 product (version 5) and used to construct an 8 day composite NDVI time series following *Zhou et al*. [2005]:

$$NDVI = \frac{SR_{b2} - SR_{b1}}{SR_{b2} + SR_{b1}}. \tag{2}$$

SR_{b1} and SR_{b2} reflectance products were quality checked by the Reflectance Band Quality field in the MOD09Q1 product, and only retrievals categorized as "highest quality" in both bands were included in the analysis. NDVI was calculated by applying (2) to patterns of 1, 5, and 9 MOD09Q1 pixels centered on each SCAN site. At most SCAN sites, little difference was seen in NDVI values averaged over all three pattern types (not shown). Given this result and known limitations in MODIS geo-location accuracy (50 m at nadir view) [*Wolfe et al*., 2002], final NDVI time series were based on the average of a 9 pixel pattern centered on individual SCAN sites; however, results for a smaller 1 pixel pattern will also be discussed below. Dixon's Q-test [*Dean and Dixon*, 1951] was performed on the final NDVI time series to filter outliers.

2.3. SWI and Soil Moisture Processing

Following *Wagner et al*. [1999], a semi-empirical exponential filter expressed in (3) was used to convert θ_5 data acquired between [*t-M*, *t*] into a θ_{0-60} proxy SWI at time *t*:

$$SWI_t = \frac{\sum_{t_i=t-M}^{t} \theta_{5,t_i} \exp\left(-\frac{t-t_i}{T}\right)}{\sum_{t_i=t-M}^{t} \exp\left(-\frac{t-t_i}{T}\right)} \tag{3}$$

where *T* is a response time scale (parameter), and *M* is the measurement length over which past θ_5 information is integrated. Unless otherwise stated, $M = 3T$. In general, *T* is affected by inter-annual hydroclimatic variability, soil wetness seasonality, and soil hydrological properties which control the vertical hydraulic connectivity between various soil layers [*Albergel et al*., 2008]. Specific strategies for defining *T* are described below.

Following the application of (1) and (3), daily time series of θ_5, θ_{0-60}, and SWI were individually averaged in time to match the 8 day compositing periods of the NDVI product. To remove the seasonal cycle and focus solely on inter-annual climatological anomalies, climatological averages for θ_5, θ_{0-60}, SWI, and NDVI were calculated for each 8 day period of the annual cycle (within the time period 2001 and 2012) and subtracted from each raw 8 day time series. In order to focus on the period with maximum vegetation growth sensitivity to water stress, only observations acquired during the boreal growing season (May to September) were considered [*Bolten and Crow*, 2012].

2.4. Information Measures

Our fundamental approach is based on calculating the MI content [*Cover and Thomas*, 1991] between current θ_5, θ_{0-60}, SWI, and near-future NDVI. Such an approach is based on the rationale that the MI content between various soil moisture products and lagged NDVI can be used to quantify the value of each product for forecasting near-future variations in vegetation productivity (the typical goal of an agricultural drought monitoring system). MI is a nonparametric measure of correlation between two random variables, in this case soil moisture and vegetation. Note that correlation here is defined strictly as the lack of independence between two random variables, and it differs from the commonly used metrics as Pearson-product moment correlation coefficient, which is a parametric approximation of normalized MI (NMI). In addition, NMI is a rigorous measure compared to commonly used metrics such as rank correlation.

The amount of information contained in the realization of a random variable ζ distributed over an event space $\zeta \in \Omega_\zeta$ according to distribution p_ζ is $-\ln(p_\zeta(\zeta))$ [*Shannon*, 1948]. The entropy of the distribution p_ζ is the expected amount of information from a sample of p_ζ:

$$\mathrm{H}(p_\zeta) = \mathrm{E}_\zeta\left[-\ln\left(p_\zeta(\zeta)\right)\right]. \tag{4}$$

Entropy can be interpreted as a measure of uncertainty about ζ according to the distribution p_ζ. Given that a random variable is distributed according to $p_\zeta(\zeta)$, approximating p_ζ by p'_ζ results in information loss which can be measured as the divergence from p_ζ to p'_ζ [*Kullback and Leibler*, 1951]:

$$D(p_\zeta \,\|\, p'_\zeta) = E_\zeta\left[\ln\left(p_\zeta(\zeta)\right) - ln(p'_\zeta(\zeta))\right]. \tag{5}$$

The expected amount of information about one random variable ζ contained in a realization of another random variable ψ is called the MI between ζ and ψ and is measured by the expected divergence, over $\psi \in \Omega_\psi$, between the conditional and marginal distributions over ζ:

$$\mathrm{MI}(\zeta;\psi) = \mathrm{E}_\psi\left[D\left(p_{\zeta|\psi} \,\|\, p_\zeta\right)\right]. \tag{6}$$

The integrations necessary for measuring MI in (5) and (6) are performed by discretizing the soil moisture and NDVI observation spaces separately. This not only allows tractability but also ensures that MI is non-negative [*Cover and Thomas*, 1991]. The observation space is discretized using a histogram bin width w given by *Scott* [2004]:

$$w = 3.73\hat{\sigma}k^{-\frac{1}{3}} \tag{7}$$

where $\hat{\sigma}$ refers to the standard deviation of soil moisture or NDVI in observation space, and k is the sample size for soil moisture and NDVI pair. Integrations are approximated as summations over the empirical probability distribution function bins in observation space.

At each SCAN site, we estimated the MI between NDVI and three different soil moisture time series (θ_5, $\theta_{0\text{–}60}$, and SWI). NMI values, denoted as NMI(θ_5), NMI($\theta_{0\text{–}60}$), and NMI(SWI), were obtained by dividing MI values by the entropy of the coincident NDVI time series. This ratio, also known as Theil index [*Theil*, 1967], describes the fractional reduction in uncertainty due to Bayesian conditioning [*Nearing et al*., 2013]. A nonparametric bootstrapping approach with 1000 replicates was used to quantify the standard error of NMI differences between various soil moisture products (e.g., NMI(θ_5) − NMI($\theta_{0\text{–}60}$) at a particular site). Error bars for pooled average differences (across all 61 SCAN sites) were calculated assuming spatially independent sampling errors at individual sites.

2.5. Parameterization of *T*

We first optimized T (T_{opt}) for each site so that SWI has its highest NMI content with NDVI at a 16 day lag. Next, to provide a more general parameterization (and guard against the possibility of over-fitting T_{opt} results), we calculated a single linear regression relationship between T_{opt} and site-specific hydro-climatic variables, including long-term (2001–2012) annual mean NDVI, annual mean precipitation ([mm], also collected from each SCAN site), and the aridity index. A stepwise regression analysis demonstrated that among the three candidate predictors, only long-term climatological NDVI is an effective regressor of T_{opt}. Consequently, a single global linear relationship was established between the long-term mean NDVI at each site ($<$NDVI$>$) and T_{opt} ($T_{\mathrm{opt}} = -75.263 \times <\mathrm{NDVI}> + 68.171$, Pearson correlation coefficient $R = 0.50$, $p < 0.01$). This single relationship was then used to estimate T at all 61 SCAN sites. An upper bound of 60 [days] was imposed on T.

Unless otherwise stated, all subsequent SWI time series results are based on (3) and values of T derived in this manner. Note that, our strategy for calculating T is kept intentionally simplistic to test the hypothesis that even a parsimonious parameterization is capable of providing skillful SWI predictions. Naturally, more complex parameterizations (based, e.g., on exact species composition, growth stage, and pedologic considerations) would likely lead to better SWI results and are therefore recommended for other applications of (3).

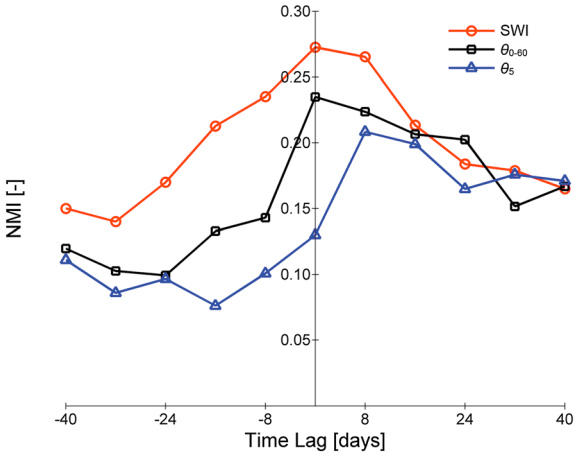

Figure 2. Example NMI(SWI), NMI(θ_{0-60}), and NMI(θ_5) results for a single SCAN site in Southeastern Arizona. Positive lags reflect the use of soil moisture as the leading variable.

3. Results

As described in section 2, our analysis focuses on SWI time series derived via (3) and an estimated value of T obtained from <NDVI>. Figure 2 shows sample NMI results at a single SCAN site in Southeastern Arizona. For positive lags (reflecting the forecasting of near-future NDVI using current soil moisture conditions) less than 24 days, NMI(θ_{0-60}) is somewhat larger than NMI(θ_5). However, NMI(SWI) is higher than NMI(θ_{0-60}) for nearly all positive lag values. This demonstrates that, at least at this particular site, SWI obtained from θ_5 and (3) has more NDVI forecasting information than θ_{0-60}.

The successful application of exponential filter to simulate vertically integrated soil moisture from superficial observations is based on the assumption that vertical hydraulic communication between soil layers is within a reasonable range. In certain circumstances, this assumption may be invalidated by the physical decoupling of the surface and sub-surface [*Albergel et al.*, 2008]. Therefore, it is important to examine NMI(SWI) minus NMI(θ_{0-60}) differences over as wide a range of land surface, soil, and climate conditions as possible. Figure 3a shows the mean and 2σ sampling error of NMI differences (at various positive lags) averaged across all 61 SCAN sites in Figure 1. Consistent with earlier site-specific results in Figure 2, NMI(θ_5) is significantly lower than NMI(θ_{0-60}) for all positive lags—indicating that superficial θ_5 observations are not as informative as vertically integrated θ_{0-60} measurements for NDVI forecasting. However, when averaged across all the 61 SCAN sites, no significant difference is observed between NMI(SWI) and NMI(θ_{0-60}). This implies that the relative advantage of θ_{0-60} measurements for drought monitoring applications is effectively reproduced via the filtering of the θ_5 via (3) and the simple parameterization of T as a fixed linear function of <NDVI>.

It should be stressed that the entropy and MI estimates in this study are maximum likelihood estimators (MLEs) for discrete random variables, and the bias of an MLE estimator is determined by both the number of observations and, generally, the magnitude of MI [*Paninski*, 2003]. In this study, the bias issue is mitigated as the sampling sizes of NMI(θ_5), NMI(θ_{0-60}), and NMI(SWI) are identical and the magnitudes of the three NMI values are similar. Nevertheless, caution should be taken when interpreting the results; for example, the NMI difference gives an estimate of the relative order of the NMI values, rather than the absolute quantity. In addition, the data processing inequality states that MI cannot be increased by transforming the soil moisture data with a model [*Cover and Thomas*, 1991]. In this case, the filtered θ_5 data have higher MI with NDVI than the raw θ_5 data because the low-pass filter aggregates information contained in lagged θ_5 data (through the time period M). This aggregation is what serves to increase MI with NDVI, and the data processing inequality is not violated. To verify this, we examined NMI(SWI) at a 16 day lag using various M in (3). These results (not shown) indicate that NMI (SWI) steadily increases as M is

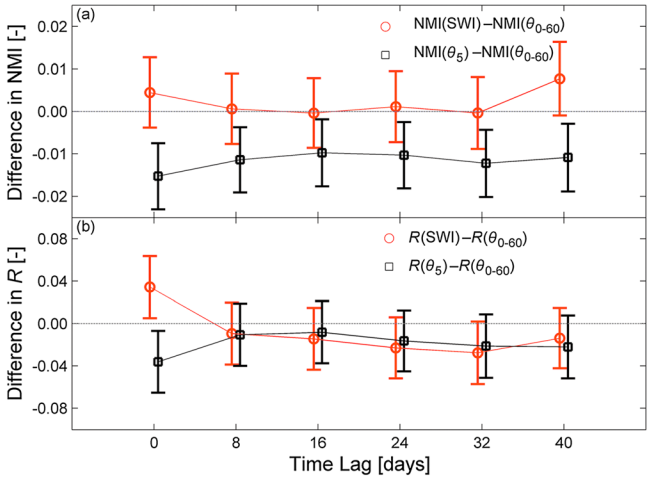

Figure 3. For all 61 SCAN sites, mean and 2σ sampling error for (a) NMI differences between both SWI and θ_{0-60}, and θ_5 and θ_{0-60}. (b) Same except for the use of R (instead of NMI) as the skill metric.

Table 1. Normalized Mutual Information (NMI) Differences at Various Time Lag τ [Days] Broken Down by Land Cover Types[a]

		NMI(SWI) − NMI(θ_{0-60})					NMI(θ_5) − NMI(θ_{0-60})				
Land Cover	#Site	$\tau=8$	$\tau=16$	$\tau=24$	$\tau=32$	$\tau=40$	$\tau=8$	$\tau=16$	$\tau=24$	$\tau=32$	$\tau=40$
Cropland	34	−0.0054	−0.0032	−0.0020	0.0028	0.0075	−0.0061	−0.0075	−0.0075	−0.0093	−0.0076
Grassland	13	**0.0215**	0.0060	−0.0062	−0.0063	0.0074	**−0.0409**	**−0.0369**	**−0.0369**	**−0.0345**	**−0.0256**
Woody savanna	7	0.0021	−0.0045	0.0183	0.0122	0.0133	0.0058	0.0144	0.0102	0.0074	−0.0043
Forest	7	−0.0105	0.0057	0.0123	−0.0174	0.0033	0.0004	0.0051	0.0047	−0.0052	−0.006

[a]Bold font indicates that 95% confidence interval of NMI difference does not include zero.

incrementally increased from zero to about 3*T*. However, increasing *M* beyond 3*T* has little subsequent impact on NMI(SWI).

In addition to NMI results in Figure 3a, we include results in Figure 3b based on applying Pearson's *R* to describe the strength of the linear relationship between soil moisture and near-future NDVI anomalies. The *R* between NDVI and three soil moisture time series (θ_5, θ_{0-60}, and SWI) anomalies are denoted as $R(\theta_5)$, $R(\theta_{0-60})$, and *R*(SWI). To ensure consistency with NMI-based results in Figure 3a, *T* is obtained in the same way as in NMI results except that T_{opt} is now determined by maximizing the correlation *R* (as opposed to NMI) between SWI and NDVI anomalies at a 16 day lag. Consistent with earlier NMI results, sampled values of $R(\theta_5)$ in Figure 3b are lower than $R(\theta_{0-60})$ (albeit marginally). However, in contrast to NMI results in Figure 3a, the transformation of θ_5 into SWI slightly *degrades* the strength of the linear relationship between soil moisture and near-future NDVI anomalies. The discrepancy between NMI and *R* based results in Figure 3 is likely due to the ability of NMI, unlike *R*, to properly account for potential nonlinearities in the relationship between soil moisture and NDVI. In fact, a nonlinear relationship between soil moisture and NDVI anomalies can be observed at a number of SCAN sites.

In order to maximize the statistical power of our sampling, results in Figure 3a are based on NMI calculated across a wide range of sites. However, NMI results can also be broken down by land cover characteristics. For example, symbol colors in Figure 1 captures the 16 day lag NMI(SWI) minus NMI(θ_{0-60}) differences (scaled by their corresponding 1σ sampling errors) for each of the individual 61 SCAN sites. While sites with significantly negative differences of NMI(SWI) minus NMI(θ_{0-60}) exist, no clear climate or land cover patterns emerge. In fact, explicitly breaking down results by land cover type (over a range of time lag-values) reveals no significant land cover-based variations in sampled NMI(SWI) minus NMI(θ_{0-60}) results (Table 1).

4. Conclusion

The upcoming October 2014 launch of NASA's Soil Moisture Active/Passive satellite mission is expected to represent a major advance in our ability to retrieve surface soil moisture using microwave remote sensing [*Entekhabi et al*., 2010]. However, the shallow vertical representative depth of microwave-based retrievals is generally considered to be a major limitation for their use in drought monitoring. Here, we examine the loss of drought monitoring information when substituting (zero to 60 cm) vertically integrated soil moisture observations ($\theta_{0-60\ [cm]}$) with superficial observations obtained at 5 cm ($\theta_{5\ [cm]}$) or a simple low-pass transformation of $\theta_{5\ [cm]}$ (SWI). Our evaluation is based on the calculation of normalized mutual information (NMI) between anomalies in each of these three soil moisture products and near-future NDVI anomalies.

Overall, results demonstrate that while θ_{0-60} soil moisture exhibits greater NMI (in relation to near-future NDVI) than θ_5 (see Figures 2 and 3a), no statistically significant difference is found between NMI(SWI) and NMI (θ_{0-60}) when results are averaged across all 61 SCAN sites (Figure 3a). This result, obtained despite the use of simplified (and highly general) parameterization of *T*, implies that the shallow vertical measurement support of satellite-derived θ_5 does not represent as large a limitation as commonly perceived. While some statistically significant NMI(SWI) minus NMI(θ_{0-60}) differences are observed at individual SCAN sites, these differences are equally distributed between positive and negatives ones and seem to demonstrate no coherent spatial pattern (Figure 1). It should be noted that reasonable vertical connectively between soil layers is required for the success of (3) in reproducing actual variations in vertically integrated soil moisture. Therefore, the lack of geographical clustering (and/or clear relationship with land cover type) in NMI(SWI) minus NMI(θ_{0-60}) differences may reflect a strong dependence on local soil hydraulic properties.

All results here are based on cross comparisons between point-scale soil moisture acquired at SCAN sites and (750 m scale) variations in MODIS-derived NDVI. This scale contrast will likely reduce the size of sampled NMI values. However, it is difficult to define any mechanism by which this scale contrast would preferably bias either NMI(SWI) or NMI(θ_{0-60}) and thus systematically impact NMI(SWI) minus NMI(θ_{0-60}) differences at the core of the analysis. In addition, NMI is a non-parametric measure of skill and thus robust to potential scale invariance in the linearity of the soil moisture and NDVI relationship. Finally, modifying MODIS NDVI calculations to be based on a single 250 m pixel (as opposed to the 750 m mean obtained within a centered 9 pixel pattern) had almost no qualitative impact on results (not shown). Nevertheless, the recent development of more dense sampling soil moisture ground networks and higher-resolution NDVI data products will provide an opportunity for future work to verify these results across a wider range of measurement scales.

The observed success of (3) in duplicating information present in vertically integrated soil moisture content using only superficial observations is likely due to a complex set of process-level considerations involving the relative importance of various vertically distributed soil water balance processes and the (multi-species) response of landscape-scale vegetation productivity to vertical variations in soil water availability. Even a partial accounting of these considerations is beyond the scope of this paper. In addition, the SWI approach presented here is obviously simplified and therefore not meant to represent a viable substitute for more complex land data assimilation systems (which vertically extrapolate superficial soil moisture observations by integrating them into a multilayer land surface model). However, given the well-known model dependencies and parameterization challenges faced by these systems [*Crow and Van Loon*, 2006; *Kumar et al*., 2009], it is critically important to understand how the information content of soil moisture observations is impacted by depth.

Acknowledgments

The soil moisture data set for this paper is available at USDA SCAN website (see http://www.wcc.nrcs.usda.gov/scan/), and MODIS surface reflectance and land cover data are available at http://modis.gsfc.nasa.gov. Thanks to the Natural Science Foundation of China grant (31171451), the Key Project for the Strategic Science Plan in IGSNRR, CAS (2012ZD003), and the Chinese Scholarship Council for supporting the first author to conduct research at the USDA-ARS Hydrology and Remote Sensing Laboratory. This work was also partially supported by a grant from the NASA Terrestrial Ecology Program (NNH09ZDA001N) to W.T. Crow.

The Editor thanks two anonymous reviewers for assistance in evaluating this manuscript.

References

Albergel, C., C. Rüdiger, T. Pellarin, J. C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin (2008), From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, *12*, 1323–1337, doi:10.5194/hess-12-1323-2008.

Arora, V. K., and G. J. Boer (2003), A representation of variable root distribution in dynamic vegetation models, *Earth Interact.*, *7*(6), 1–19, doi:10.1175/1087-3562(2003)007<0001:AROVRD>2.0.CO;2.

Bolten, J. D., and W. T. Crow (2012), Improved prediction of quasi-global vegetation conditions using remotely-sensed surface soil moisture, *Geophys. Res. Lett.*, *39*, L19406, doi:10.1029/2012GL053470.

Brocca, L., et al. (2011), Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, *Remote Sens. Environ.*, *115*(12), 3390–3408, doi:10.1016/j.rse.2011.08.003.

Ceballos, A., K. Scipal, W. Wagner, and J. Martínez-Fernández (2005), Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain, *Hydrol. Process.*, *19*(8), 1549–1566, doi:10.1002/hyp.5585.

Cover, T. M., and J. A. Thomas (1991), *Elements of Information Theory*, John Wiley, New York.

Crow, W. T., and E. Van Loon (2006), Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, *J. Hydrometeorol.*, *7*(3), 421–432, doi:10.1175/JHM499.1.

De Kauwe, M., M. Disney, T. Quaife, P. Lewis, and M. Williams (2011), An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, *Remote Sens. Environ.*, *115*(2), 767–780, doi:10.1016/j.rse.2010.11.004.

Dean, R. B., and W. J. Dixon (1951), Simplified statistics for small numbers of observations, *Anal. Chem.*, *23*(4), 636–638, doi:10.1021/ac60052a025.

Entekhabi, D., et al. (2010), The Soil Moisture Active and Passive (SMAP) mission, *IEEE Proc.*, *98*(5), 704–716, doi:10.1109/JPROC.2010.2043918.

Escorihuela, M. J., A. Chanzy, J. P. Wigneron, and Y. H. Kerr (2010), Effective soil moisture sampling depth of L-band radiometry: A case study, *Remote Sens. Environ.*, *114*(5), 995–1001, doi:10.1016/j.rse.2009.12.011.

Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. Seyfried, D. C. Goodrich, M. S. Moran, and J. Du (2010), Validation of advanced microwave scanning radiometer soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, *48*(12), 4256–4272, doi:10.1109/TGRS.2010.2051035.

Kullback, S., and R. A. Leibler (1951), On information and sufficiency, *Ann. Math. Stat.*, *22*(1), 79–86, doi:10.1214/aoms/1177729694.

Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard (2009), Role of subsurface physics in the assimilation of surface soil moisture observations, *J. Hydrometeorol.*, *10*(6), 1534–1547, doi:10.1175/2009JHM1134.1.

Li, J., and S. Islam (1999), On the estimation of soil moisture profile and surface fluxes partitioning from sequential assimilation of surface layer soil moisture, *J. Hydrol.*, *220*(1–2), 86–103, doi:10.1016/S0022-1694(99)00066-9.

Mo, X., J. Qiu, S. Liu, and V. Naeimi (2011), Estimating root-layer soil moisture for north China from multiple data sources, in *IAHS Red Book for XXV IUGG Symposia JH01: GRACE, Remote Sensing and Ground-based Methods in Multi-Scale Hydrology*, edited by M. Hafeez et al., pp. 118–124, IAHS, Wallingford, Oxon., U. K.

Nearing, G. S., H. V. Gupta, W. T. Crow, and W. Gong (2013), An approach to quantifying the efficiency of a Bayesian filter, *Water Resour. Res.*, *49*, 2164–2173, doi:10.1002/wrcr.20177.

Paninski, L. (2003), Estimation of entropy and mutual information, *Neural Comput.*, *15*(6), 1191–1253, doi:10.1162/089976603321780272.

Pollacco, J. A. P., and B. P. Mohanty (2012), Uncertainties of water fluxes in soil–vegetation–atmosphere transfer models: Inverting surface soil moisture and evapotranspiration retrieved from remote sensing, *Vadose Zone J.*, *11*(3), doi:10.2136/vzj2011.0167.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.

Reichle, R. H., W. T. Crow, R. D. Koster, H. O. Sharif, and S. P. P. Mahanama (2008), Contribution of soil moisture retrievals to land data assimilation products, *Geophys. Res. Lett.*, *35*, L01404, doi:10.1029/2007GL031986.

Sabater, J. M., L. Jarlan, J. C. Calvet, F. Bouyssel, and P. de Rosnay (2007), From near-surface to root-zone soil moisture using different assimilation techniques, *J. Hydrometeorol.*, *8*(2), 194–206, doi:10.1175/JHM571.1.

Scott, D. W. (2004), Multivariate density estimation and visualization, in *Handbook of Computational Statistics: Concepts and Methods*, edited by J. E. Gentle, W. Haerdle, and Y. Mori, pp. 517–538, Springer, New York.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell Tech. J.*, *27*(4), 379–423, doi:10.1002/j.1538-7305.1948.tb00917.x.

Theil, H. (1967), *Economics and Information Theory*, Rand McNally, Chicago, Ill.

Wagner, W., G. Lemoine, and H. Rott (1999), A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sens. Environ.*, *70*(2), 191–207, doi:10.1016/S0034-4257(99)00036-X.

Wolfe, R. E., M. Nishihama, A. J. Fleig, J. A. Kuyper, D. P. Roy, J. C. Storey, and F. S. Patt (2002), Achieving sub-pixel geolocation accuracy in support of MODIS land science, *Remote Sens. Environ.*, *83*(1–2), 31–49, doi:10.1016/S0034-4257(02)00085-8.

Zhou, X., H. Xie, and J. M. Hendrickx (2005), Statistical evaluation of remotely sensed snow-cover products with constraints from streamflow and SNOTEL measurements, *Remote Sens. Environ.*, *94*(2), 214–231, doi:10.1016/j.rse.2004.10.007.

Zribi, M., T. P. Anguela, B. Duchemin, Z. Lili, W. Wagner, S. Hasenauer, and A. Chehbouni (2010), Relationship between soil moisture and vegetation in the Kairouan plain region of Tunisia using low spatial resolution satellite data. *Water Resour. Res.*, *46*, W06508, doi:10.1029/2009WR008196.