

# Performance Metrics for Soil Moisture Retrievals and Application Requirements

DARA ENTEKHABI

*Ralph M. Parsons Laboratory for Environmental Science and Engineering, Massachusetts Institute of Technology,  
Cambridge, Massachusetts*

ROLF H. REICHLER AND RANDAL D. KOSTER

*Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland*

WADE T. CROW

*USDA/ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland*

(Manuscript received 5 October 2009, in final form 24 December 2009)

## ABSTRACT

Quadratic performance metrics such as root-mean-square error (RMSE) and time series correlation are often used to assess the accuracy of geophysical retrievals (satellite measurements) with respect to true fields. These metrics are related; nevertheless, each has advantages and disadvantages. In this study the authors explore the relation between the RMSE and correlation metrics in the presence of biases in the mean as well as in the amplitude of fluctuations (standard deviation) between estimated and true fields. Such biases are common, for example, in satellite retrievals of soil moisture and impose constraints on achievable and meaningful RMSE targets. Last, an approach is introduced for converting a requirement in an application's product into a corresponding requirement for soil moisture accuracy. The approach can help with the formulation of soil moisture measurement requirements. It can also help determine the utility of a given retrieval product for applications.

## 1. Introduction

A variety of performance metrics are used for the validation of geophysical retrievals (estimates based on remotely sensed observations) and for the definition of measurement requirements. The choice of metric is mostly governed by the nature of the geophysical variable (units, range, etc.) and is influenced by the characteristics of the science application and its sensitivity to the retrieved geophysical variable (Stanski et al. 1989). No single metric or statistic can capture all the attributes of environmental variables. Each metric is robust with respect to some attributes and relatively insensitive or incomplete with respect to others.

The appropriateness of various performance metrics has received considerable attention in fields such as rainfall-runoff modeling (see, e.g., Gupta et al. 2009); however,

relatively little work has focused on clarifying choices required for the definition of remote sensing measurement requirements. Here we consider the particular challenge of defining metrics for satellite retrievals of surface (top 5 cm) soil moisture and for data products (including rootzone soil moisture) that are derived from the assimilation of the surface retrievals into a land model. Remote sensing of terrestrial microwave emission and radar backscatter in the L-band spectral range is sensitive to the water content of soils in a 0–5-cm surface layer. Such retrievals will soon be available from the Soil Moisture Ocean Salinity (SMOS) mission (available online at <http://www.esa.int/esaLP/LPsmos.html>) and the Soil Moisture Active and Passive (SMAP) mission (available online at <http://smap.jpl.nasa.gov>).

Soil moisture controls the partitioning of available energy into sensible and latent heat fluxes across regions where the evaporation regime is, at least intermittently, water limited. Since the fluxes of sensible heat and moisture at the base of the atmosphere influence the evolution of weather, soil moisture is often a significant factor in

---

*Corresponding author address:* Dara Entekhabi, 48-216G, MIT, Cambridge, MA 02139.  
E-mail: [darae@mit.edu](mailto:darae@mit.edu)

the performance of atmospheric models used for numerical weather and seasonal climate prediction. In this context, the metric that is used to define soil moisture measurement requirements is largely influenced by the need to capture soil moisture's control over land-atmosphere interactions in atmospheric models—in particular, by the ability of the measurement to distinguish soil moisture levels that lead to different evaporation rates.

Other applications that help define soil moisture measurement requirements are less mediated by the processes of land-atmosphere interaction and more sensitive to the relative saturation of the surface soil layer. Examples include flood and flash-flood prediction and terrain trafficability for defense applications. Floods are typically generated when precipitation exceeds the capacity of the soil to absorb incident rain. Therefore, the deficit of surface soil moisture (or the air space in the soil matrix) is the critical factor. Similarly the soil water content affects the geomechanical properties of land surfaces and thus influences the performance of vehicles, such as off-road heavy military equipment.

Together, these examples illustrate how different applications may require different, application-specific metrics to define soil moisture measurement requirements. For evaporation calculations, quantification of soil moisture variations is much more important in the drier portion of the soil moisture range since, in the near-saturated to saturated portion of the range, evaporation is energy limited and not affected by soil moisture variations. Consequently, a measurement at the wet end may have high error with little impact on the computed surface energy budget. In contrast, for flood forecasting, high accuracy in the quantification of soil moisture levels at the wetter end may be vital.

A number of factors beyond the chosen application complicate the choice of performance metric. For example, volumetric soil moisture content, the mixing ratio of water in the bulk soil matrix, is a variable that is bounded by zero at the lower end and by porosity at the upper end. This constitutes a physical range with very strict bounds. Taken together with the fact that soil moisture rises rapidly with time-intermittent precipitation input and decreases quasi exponentially between precipitation events, the marginal frequency distribution of soil moisture is bounded and skewed (Ryu and Famiglietti, 2005; Teuling and Troch 2005).

Furthermore, soil moisture is highly variable in space since its value is ultimately determined by overlaying spatial patterns of precipitation, vegetation, soil texture, terrain, and solar aspect angle variations. A greater degree of spatial averaging generally implies a more constrained dynamic range for the averaged variable. The actual dynamic range of soil moisture within its maximum

possible physical range (zero to porosity) is correspondingly dependent on the spatial scale under consideration (Rodriguez-Iturbe et al. 1995; Hu et al. 1997; Crow and Wood 1999; Entin et al. 2000; Famiglietti et al. 2008). For reference, in situ measurements from three dense soil moisture monitoring networks in Arizona, Georgia, and Oklahoma (Jackson et al. 2009, manuscript submitted to *IEEE Trans. Geosci. Remote Sens.*) indicate surface soil moisture temporal standard deviations of about 0.02, 0.04, and 0.06  $\text{m}^3 \text{m}^{-3}$ , respectively, for spatial resolutions consistent with satellite footprint scales (10–30 km).

A further complication to the choice of performance metric for soil moisture is related to fundamental limitations in the models that will eventually use the data. Soil moisture is a state variable for models of soil hydrology, including those used in conjunction with numerical atmospheric models. Ideally, geophysical retrievals of soil moisture in the surface layer would be used to adjust a model's state variable toward the observations. However, a model's soil moisture is principally designed to link various fluxes in the model in a consistent manner (Koster and Milly 1997; Albertson and Kiely 2001). The evolution of land surface models has focused on getting these fluxes right rather than on producing soil moisture products that are accurate when compared to direct observations. Models are, in fact, strongly limited in their ability to reproduce observed soil moisture by a lack of critical information on soil hydraulic properties and, more importantly, by the logistical need to represent complex, nonlinear, and nonresolvable processes across large distances in a very simple way. Indeed, there are often considerable mean biases or biases in the dynamic range of the retrieved and modeled soil moisture (Reichle et al. 2004). Because the long-term mean soil moisture typically varies seasonally, it is not surprising that biases also vary with season (Drusch et al. 2005).

At first glance, stationary (though unknown) biases in surface retrievals (associated, for example, with inaccurate estimates of ancillary surface data, such as roughness height) may appear to be a problem that inherently limits the usefulness of satellite data. Even if retrievals were completely unbiased relative to nature, their use in models might seem intractable, given the aforementioned biases in the models' soil moisture state variables. Curiously, though, the proper interpretation of land model soil moisture may to a large extent negate these apparent weaknesses. This is because models require not the absolute magnitude of soil moisture but rather the time variation, *in a percentile sense*, of soil moisture fluctuations—variations that can be scaled directly to corresponding variations (percentiles) in the model's soil moisture variable. In effect, biases in the retrievals themselves (in terms of

both mean and variance) may “scale out” through the percentile-based transformation of the retrievals. As long as the retrievals reproduce the time variability of true soil moisture accurately, they can be biased in their mean and dynamic range and still be useful (Reichle et al. 2007; Koster et al. 2009).

Choosing a sensible metric of soil moisture retrieval accuracy for a given application requires a solid understanding of these issues. It requires that the relative strengths and weaknesses of available metrics, and the connections between them, be fully understood. In this study we compare two common, simple performance metrics: the root-mean-square error (RMSE) and the time series correlation ( $r$ ). The widespread use of these quadratic error metrics continues even though more robust or appropriate methods (for certain applications) may be available. For example, as noted above, energy budget and flood prediction applications might best focus on accuracy in different subsets of the total soil moisture range; the RMSE and  $r$  metrics examined here, however, are inherently simple and cannot make such a distinction.

In addition to analyzing the RMSE and  $r$  metrics and showing how they are related, this paper introduces an approach for linking the measurement requirements of specific applications to the common quadratic performance metrics of soil moisture. Again, the purpose of the paper is not to define a numerical value for the measurement requirement, which in any case depends on the application at hand, but to define a framework for appropriate invocation of these performance metrics for validation or the specification of measurement requirements.

## 2. Sample statistical measures

If the true surface volumetric soil moisture (at a given scale) is defined as  $\theta_{\text{true}}$  and the corresponding estimated retrieval is  $\theta_{\text{est}}$ , then the rms error metric is simply

$$\text{RMSE} = \sqrt{E[(\theta_{\text{est}} - \theta_{\text{true}})^2]}, \quad (1)$$

where  $E[\cdot]$  is the expectation operator. This metric quadratically penalizes deviations of the estimate with respect to the true soil moisture (in units of volumetric soil moisture) and is a compact and easily understood measure of estimation accuracy. This metric, however, is severely compromised if there are biases in either the mean or the amplitude of fluctuations in the retrieval. If it can be estimated reliably, the mean bias  $b = E[\theta_{\text{est}}] - E[\theta_{\text{true}}]$  can easily be removed by defining the unbiased RMSE:

$$\text{ubRMSE} = \sqrt{E\{[(\theta_{\text{est}} - E[\theta_{\text{est}}]) - (\theta_{\text{true}} - E[\theta_{\text{true}}])]^2\}}. \quad (2)$$

The RMSE and the unbiased RMSE are related through

$$\text{RMSE}^2 = \text{ubRMSE}^2 + b^2, \quad (3)$$

which implies  $\text{RMSE} \geq |b|$  and underscores the shortcomings of the RMSE metric in the presence of mean bias. (As noted above, the bias in soil moisture may vary with season. It is straightforward to generalize the relationships discussed here to account for such slowly varying bias. Hereinafter, we assume that ubRMSE reflects the RMSE of soil moisture anomalies that are computed by removing the mean seasonal cycle.)

Furthermore, as argued above, while the amplitude of fluctuations between the retrieval estimate and the true soil moisture may be very different, the retrieval is still potentially useful to applications, particularly those involving land models. Given the percentile-based transformation that is typically applied when assimilating retrievals into a model, an additional metric is needed that is insensitive to any retrieval mean bias and bias in amplitude of fluctuations (as expressed in the statistical variance). The metric examined here is the sample time series correlation, expressed as

$$r = \frac{E[(\theta_{\text{est}} - E[\theta_{\text{est}}])(\theta_{\text{true}} - E[\theta_{\text{true}}])]}{\sigma_{\text{est}}\sigma_{\text{true}}}, \quad (4)$$

where  $\sigma_{\text{est}}^2$  and  $\sigma_{\text{true}}^2$  are the time variances of the estimated (retrieval) and true soil moisture for the remote sensing pixel, respectively. The correlation metric captures the correspondence in phase between retrieval estimates and truth, and therefore the coherent phasing information even if there is mean bias and/or differences in variance. In this sense it provides a different perspective on retrieval performance than RSME.

Nonetheless, the two metrics are related. Murphy (1995), Barnston (1992), and others show that expansion of the grouped products in (1) and (4) yields a relation between RSME and  $r$ :

$$\text{RMSE} = \sqrt{\sigma_{\text{est}}^2 + \sigma_{\text{true}}^2 - 2r\sigma_{\text{est}}\sigma_{\text{true}} + b^2}. \quad (5)$$

Using (3), this equation can easily be reformulated to express the relationship between unbiased RMSE and correlation:

$$\text{ubRMSE} = \sqrt{\sigma_{\text{est}}^2 + \sigma_{\text{true}}^2 - 2 \cdot r \cdot \sigma_{\text{est}} \cdot \sigma_{\text{true}}}. \quad (6)$$

A plot of this relationship provides insight into the two statistical metrics (ubRMSE and  $r$ ). Figure 1 presents, for a fixed value of  $\text{ubRMSE} = 0.04 \text{ m}^3 \text{ m}^{-3}$  (chosen arbitrarily here), the corresponding correlation coefficient

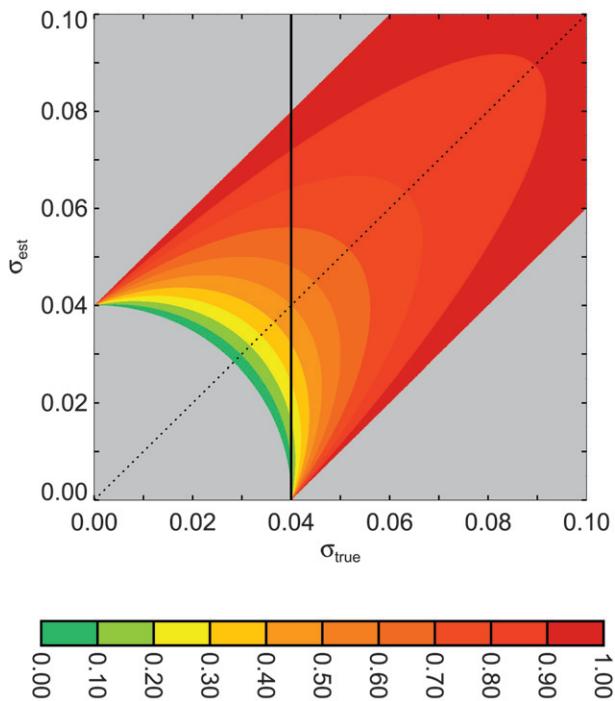


FIG. 1. Temporal correlation coefficient  $r$  as a function of true variability  $\sigma_{\text{true}}$  and variability in a retrieval estimate  $\sigma_{\text{est}}$  for a nominal ubRMSE =  $0.04 \text{ m}^3 \text{ m}^{-3}$ .

as a function of  $\sigma_{\text{est}}$  (the retrieval standard deviation) and  $\sigma_{\text{true}}$  (the true standard deviation). For example, if  $\sigma_{\text{est}} = 0.03 \text{ m}^3 \text{ m}^{-3}$  and  $\sigma_{\text{true}}$  is twice that value, then  $\text{ubRMSE} = 0.04 \text{ m}^3 \text{ m}^{-3}$  implies a time correlation between the retrieval and truth values of about 0.8. As should be expected, when the variance of the retrieval is not biased relative to truth (i.e., when  $\sigma_{\text{est}} = \sigma_{\text{true}}$ ), an increase in soil moisture variance implies an increase in the corresponding correlation between the truth and the retrievals to maintain the fixed  $\text{ubRMSE} = 0.04 \text{ m}^3 \text{ m}^{-3}$  ( $r$  increases as one moves up the 1:1 line in the plot). More generally, Fig. 1 clearly illustrates that for a given value of ubRMSE, the corresponding correlation coefficient can lie anywhere between 0 and 1.

Perhaps the most interesting features of the contour plot are the gray areas, which represent regions in the  $[\sigma_{\text{est}}, \sigma_{\text{true}}]$  parameter space for which  $\text{ubRMSE} = 0.04 \text{ m}^3 \text{ m}^{-3}$  cannot be realized for any sensible value of  $r$  (between 0 and 1). The gray areas in the upper left and lower right corners correspond to the condition  $|\sigma_{\text{est}} - \sigma_{\text{true}}| > 0.04 \text{ m}^3 \text{ m}^{-3}$ ; more generally, manipulation of (6) reveals that ubRMSE must satisfy

$$|\sigma_{\text{est}} - \sigma_{\text{true}}| \leq \text{ubRMSE}; \tag{7}$$

that is, the unbiased RMSE cannot be less than the bias in the standard deviation. This is a powerful constraint,

akin to that imposed by the mean bias on the RMSE ( $\text{RMSE} \geq |b|$ ).

The gray region in the lower left of the plot corresponds to a regime in the parameter space for which the ubRMSE always lies below the  $0.04 \text{ m}^3 \text{ m}^{-3}$  value, provided  $r$  is between 0 and 1. (Note that allowing negative  $r$  values—anticorrelations between true and retrieved soil moistures—can fill in part of this area, but this interpretation of  $r$  is not sensible and thus not considered here.) In fact, a look at (6) shows that, in general (for  $r \geq 0$ ),

$$\text{ubRMSE} \leq \sqrt{\sigma_{\text{est}}^2 + \sigma_{\text{true}}^2}. \tag{8}$$

In other words, ubRMSE cannot exceed a given value that depends on the variability of the estimated and true soil moisture.

In fact, according to (2), trivially constant “retrievals” of the form  $\theta_{\text{est}} \equiv E[\theta_{\text{est}}]$  (so that  $\sigma_{\text{est}} = 0$ ) produce a ubRMSE value of precisely  $\sigma_{\text{true}}$ . This means that from the RMSE perspective, retrievals with a ubRMSE that exceeds  $\sigma_{\text{true}}$  should be abandoned in favor of a constant value, set equal to  $E[\theta_{\text{est}}]$ . In the context of Fig. 1, while any (nongray) point in the parameter space to the left of the vertical line at  $\sigma_{\text{true}} = 0.04 \text{ m}^3 \text{ m}^{-3}$  does produce a ubRMSE of  $0.04 \text{ m}^3 \text{ m}^{-3}$  for the plotted value of  $r$ , an even smaller ubRMSE could be achieved by setting  $\theta_{\text{est}} \equiv E[\theta_{\text{est}}]$ . Of course, from the correlation perspective, such constant “retrievals” make no sense because a constant contains no information at all ( $r = 0$ ). From the correlation perspective, any positive  $r$  value, even to the left of the line, implies that retrievals do contain real and potentially useful information.

For a satellite mission, the above discussion has important implications for the formulation of accuracy requirements in terms of the RMSE metric ( $\text{RMSE}_{\text{target}}$ ). The following constraints apply:

- (i) Even if the temporal correlation between the retrieval and true soil moisture were perfect ( $r = 1$ ),  $\text{RMSE}_{\text{target}}$  cannot be achieved if the mean bias exceeds it (i.e., if  $|b| > \text{RMSE}_{\text{target}}$ ).
- (ii) Even with perfect temporal correlations and zero mean bias ( $r = 1$  and  $b = 0$ ),  $\text{RMSE}_{\text{target}}$  cannot be achieved if the bias in the standard deviation  $|\sigma_{\text{est}} - \sigma_{\text{true}}|$  exceeds it (i.e., if  $|\sigma_{\text{est}} - \sigma_{\text{true}}| > \text{RMSE}_{\text{target}}$ ).
- (iii) A sensible  $\text{RMSE}_{\text{target}}$  must consider the variability of soil moisture because, from an RMSE perspective, simply using constant retrievals guarantees that we can always achieve an RMSE of at most  $\sigma_{\text{true}}$  (for  $b = 0$ ).

The three constraints are useful in practice only if we can estimate the mean and variability of soil moisture with

sufficient accuracy. (One could indeed argue that a fundamental objective of a soil moisture satellite mission is the measurement of this climatology, which is still poorly known.) Furthermore, constraints (i) and (ii) assume that  $r = 1$ . If the temporal correlation is not perfect ( $r < 1$ ), as it likely will not be, the ability to achieve  $\text{RMSE}_{\text{target}}$  is that much more difficult. Constraint (iii) is of a different nature. Because  $\sigma_{\text{true}}$  differs from place to place, constraint (iii) implies that a single global  $\text{RMSE}_{\text{target}}$  may not be appropriate.

In contrast, the accuracy requirement for a mission could be formulated in terms of correlation ( $r_{\text{target}}$ ). In this case, the only theoretical constraint would be to require a positive correlation (i.e.,  $r_{\text{target}} > 0$ ), for again, any positive correlation implies that the retrieval contains potentially useful information. An  $r$  requirement, however, ignores potentially important biases in mean or variability that may need to be evaluated or constrained.

To summarize, the dimensionless correlation metric  $r$  captures the coherence in phasing of estimate and truth regardless of biases in mean and variance. The RMSE metric captures the closeness of estimate and truth with a quadratic penalty for error outliers. It has units of the original variable but it is hampered by biases in the mean and in the amplitude of fluctuations. If these biases are large enough (see above), a mission RMSE target will be unachievable even with perfect temporal correlation between the estimates and truth. Conversely, if the true soil moisture standard deviation at a location is less than the RMSE target and if the true mean is known with sufficient accuracy, the RMSE requirement at that location can be met trivially without ever putting a satellite in orbit. The two performance metrics are related and—as expected—the relation is mediated by the bias in the mean and the bias in the amplitude of variation around the mean. It is important to recognize that the correction of the mean bias and of any amplitude differences requires knowledge of the true soil moisture climatology, that is, of  $E[\theta_{\text{true}}]$  and  $\sigma_{\text{true}}^2$ .

A remaining fundamental problem is the lack of soil moisture observations that can be used for the validation of global soil moisture data products. The reliable estimation of any soil moisture statistic from point-scale in situ soil moisture sensors is difficult globally—especially when statistics are required to match the coarse (typically  $>10$  km) spatial support of satellite retrievals. However, recent work has seen the development of a number of independent strategies to address this upscaling challenge. They include the extraction of  $r$  information from simple data assimilation systems (Crow 2007), the application of time stability approaches to maximize the coarse-scale representativeness of point observations

in the estimation of biases (see, e.g., Mohanty and Skaggs 2001; Martínez-Fernández and Ceballos 2003; Cosh et al. 2006), and the estimation of RMSE based on the comparison of three or more soil moisture products with independent errors (Scipal et al. 2008).

### 3. Application requirements and metrics of performance

Because there are diverse applications for soil moisture and because each application has a different sensitivity to errors in soil moisture (section 1), we seek a new approach to examining soil moisture accuracy in the context of a given application, thereby allowing the meaningful determination of an improved soil moisture accuracy requirement. Key to the construction of such a metric is an understanding of the underlying relationship between soil moisture and the application's quantity. In our discussion below, we will assume that this relationship is known. If  $\mu$  represents an application's quantity of interest, we assume we can compute

$$\mu_{\text{est}} = f(\theta_{\text{est}}, \text{other quantities}), \quad (9)$$

where  $\theta_{\text{est}}$  is the soil moisture measurement. The quantity  $\mu$ , for example, could be the degree of crop wilting due to water stress, the evaporative fraction (EF) (the ratio of time-averaged latent heat flux to time-averaged available energy), the rate of soil carbon respiration, or the degree to which a unit of heavy rolling machinery might sink into the soil. A user interested in the quantity  $\mu$  will naturally have some specific requirements in mind for the accuracy of  $\mu$ . The idea here is to translate, through various manipulations involving (9), these user-defined requirements into a required accuracy for the soil moisture measurement,  $\theta_{\text{est}}$ , in terms of the traditional  $r$  or uBRMSE metrics typically referenced in accuracy requirements for satellite data products.

We present such a methodology here through example. Consider a user interested in EF, a key element of the surface energy budget. The user is assumed to have some accuracy requirements for EF; we translate these requirements into accuracy requirements for soil moisture using, for (9), the functional relationship shown in heavy black lines in Fig. 2a: EF rises with soil moisture up to a certain point, after which it remains constant, insensitive to soil moisture variations. Note that, while the general form used here for this relationship is well supported in the literature (Budyko 1974; Eagleson 1978), the precise details of its structure in nature are largely unknown; for our example, which is meant for illustrative purposes only, the transition points and the plateau value of EF shown in the figure are chosen arbitrarily. Naturally

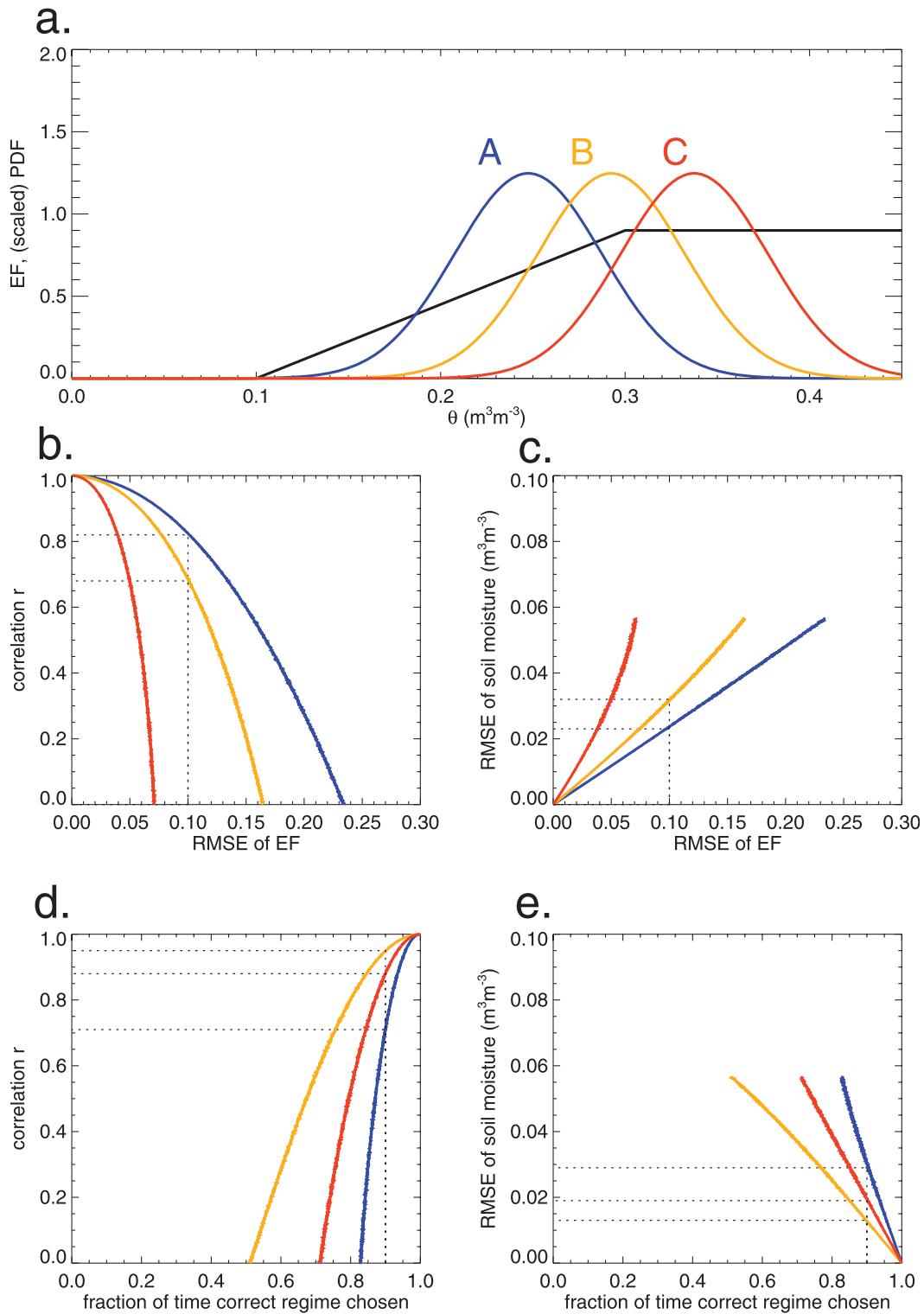


FIG. 2. (a) Solid black lines: assumed relationship between EF and soil moisture. Colored curves: assumed soil moisture PDFs. (b) Derived (color coded) relationships between the RMSE of EF and  $r$  between retrieval soil moisture estimates and truth. (c) Derived (color coded) relationships between the RMSE of EF and the RMSE of soil moisture, assuming no bias in the mean or standard deviation. (d) As in (b) but for a different EF metric: the fraction of time the correct evaporative regime is determined. (e) As in (c) but for the alternative EF metric.

the success of any translation of application requirements to soil moisture requirements depends on the accuracy of the equation used for (9), but such issues indeed underlie *all soil moisture* applications work, not just the framework presented here.

The relationship in Fig. 2a is nonlinear and, as a result, measurement requirements will differ under different moisture regimes; soil moisture measurements will certainly need to be more accurate if  $\theta$  in a given climate varies between 0.2 and 0.3  $\text{m}^3 \text{m}^{-3}$  than if it varies between 0.3 and 0.4  $\text{m}^3 \text{m}^{-3}$  because EF is constant in the latter range. To illustrate quantitatively the impacts of such nonlinearity, we consider here three climatic regimes, characterized by the three distinct soil moisture probability density functions (PDFs) also shown in Fig. 2a: the first for a drier climate (A; mean = 0.25  $\text{m}^3 \text{m}^{-3}$ ), the second for an intermediate climate (B; mean = 0.29  $\text{m}^3 \text{m}^{-3}$ ), and the third for a wetter climate (C; mean = 0.34  $\text{m}^3 \text{m}^{-3}$ ). The imposed standard deviation for each is 0.04  $\text{m}^3 \text{m}^{-3}$ , a value consistent with the aforementioned ground-based surface soil moisture time series (Jackson et al. 2009, manuscript submitted to *IEEE Trans. Geosci. Remote Sens.*). We use Gaussian PDFs for simplicity here, despite the fact that soil moisture PDFs in nature are, especially for certain conditions, non-Gaussian (e.g., Famiglietti et al. 1999; Ryu and Famiglietti 2005). Regardless of the shapes used for the PDFs, the concepts discussed here are valid and applicable.

Different users, of course, will have different application requirements. To demonstrate that the approach presented here is flexible enough to deal with a variety of possible requirement specifications, we consider two users: X and Y. User X is interested in the average rms error of the EF estimate,  $\text{RMSE}_{\text{EF}}$ , and wants soil moisture measured with an accuracy that ensures  $\text{RMSE}_{\text{EF}} < 0.1$ . User Y, on the other hand, is interested solely in the distinction between evaporative regimes, wanting to know only if a given evaporation rate is in the soil-moisture-controlled regime ( $0.1 \text{ m}^3 \text{m}^{-3} < \theta < 0.3 \text{ m}^3 \text{m}^{-3}$  in Fig. 2a) or in the energy-controlled regime ( $0.3 \text{ m}^3 \text{m}^{-3} < \theta < 0.45 \text{ m}^3 \text{m}^{-3}$  in Fig. 2a). User Y will consider a set of soil moisture measurements to be accurate if the measurements can properly distinguish between these regimes with a success rate of 90%.

We consider user X first. For each PDF in Fig. 2a, we use Monte Carlo techniques to construct a joint set of suitably lengthy soil moisture time series: a “truth” time series  $\theta_{\text{true}}(t)$  and an “estimated” or retrieval time series  $\theta_{\text{est}}(t)$ . Both  $\theta_{\text{true}}(t)$  and  $\theta_{\text{est}}(t)$  are sampled from the PDF in question, and the two are forced to be temporally correlated with  $r$ . The time series  $\theta_{\text{true}}(t)$  and  $\theta_{\text{est}}(t)$  are converted, using the relationship in Fig. 2a, into

corresponding time series  $\text{EF}_{\text{true}}(t)$  and  $\text{EF}_{\text{est}}(t)$ , from which a value of  $\text{RMSE}_{\text{EF}}$  is directly derived—a value of  $\text{RMSE}_{\text{EF}}$  that is a distinct function of the PDF considered and the prescribed time series correlation  $r$ .

The procedure is repeated numerous times for each PDF, using a wide range of  $r$  values. Figure 2b summarizes the results, with the derived  $\text{RMSE}_{\text{EF}}$  on the  $x$  axis. Each soil moisture PDF provides a unique relationship between  $\text{RMSE}_{\text{EF}}$  and the prescribed  $r$ . When the soil moisture measurement is perfect (i.e.,  $r = 1$ ),  $\text{RMSE}_{\text{EF}}$  for each PDF is accordingly zero. A decrease in soil moisture retrieval information (i.e., a decrease in  $r$ ) is associated with an increase in  $\text{RMSE}_{\text{EF}}$ . As expected from the discussion above, the wetter climate (PDF C) leads to lower values of  $\text{RMSE}_{\text{EF}}$  for any given value of  $r$ .

Figure 2c shows the equivalent results in terms of the RMSE of soil moisture. Here, the soil moisture RMSE is computed directly from  $r$  using (6), assuming no biases in mean or variance. As expected, an increase in  $\text{RMSE}_{\text{EF}}$  implies an increase in the RMSE of soil moisture, with the wettest climate (PDF C) allowing the greatest (most forgiving) soil moisture error for a given  $\text{RMSE}_{\text{EF}}$ .

Recall that user X had a particular numerical requirement for  $\text{RMSE}_{\text{EF}}$ . Figures 2b and 2c can be used to convert this requirement directly into corresponding requirements for soil moisture temporal correlation and RMSE. As indicated by the dotted lines in Fig. 2b, the  $\text{RMSE}_{\text{EF}} < 0.1$  requirement implies that soil moisture must be measured with a temporal correlation  $r$  greater than about 0.82 for PDF A and 0.68 for PDF B. For PDF C, the EF requirement is always met, regardless of the accuracy of the soil moisture measurement (i.e., even for  $r = 0$ ). Equivalently, again assuming no biases, Fig. 2c shows that the requirement  $\text{RMSE}_{\text{EF}} < 0.1$  implies that the soil moisture RMSE should be less than about 0.023  $\text{m}^3 \text{m}^{-3}$  for PDF A and 0.032  $\text{m}^3 \text{m}^{-3}$  for PDF B, and it implies that no skill at all is needed for PDF C.

The example of user X shows how a “traditional” RMSE metric for evaporative fraction can be transformed into a traditional  $r$  or RMSE metric for soil moisture accuracy. We now show that the same approach can be used to transform a nontraditional EF metric, such as that employed by user Y, into a traditional soil moisture metric. From the Monte Carlo time series of  $\text{EF}_{\text{true}}(t)$  and  $\text{EF}_{\text{est}}(t)$  described earlier, we can also determine, as a function of soil moisture PDF and prescribed  $r$ , the probability that  $\text{EF}_{\text{true}}(t)$  and  $\text{EF}_{\text{est}}(t)$  lie in the same evaporative regime (we simply count the number of times they do lie in the same regime and divide by the length of the time series). In analogy to Figs. 2b and 2c, Figs. 2d and 2e show the relationships between the probability of choosing the correct evaporative regime and soil moisture  $r$  and RMSE, respectively. A 90% success rate

translates through this approach to  $r$  values of 0.71, 0.95, and 0.88 for PDFs A, B, and C, respectively. It correspondingly translates to soil moisture RMSE values of about 0.029, 0.013, and 0.019  $\text{m}^3 \text{m}^{-3}$  for PDFs A, B, and C, respectively. Notice that for user Y, PDF A provides the most forgiving metrics, whereas for user X, PDF C does—soil moisture metrics are indeed user specific.

It is important to keep in mind that the above examples are provided strictly for illustration purposes. The EF requirements for user X and user Y were arbitrary; a full discussion of the best way to define an application-specific metric is beyond the scope of this paper. The numbers provided in the examples should be considered secondary to the description of the process itself and the demonstration of its ability to convert even nontraditional user requirements into traditional soil moisture metrics.

The examples assume the measurement is unbiased in mean and variance (i.e.,  $b = 0$ ,  $\sigma_{\text{est}} = \sigma_{\text{true}}$ ) so that the same PDF is used for  $\theta_{\text{true}}(t)$  and  $\theta_{\text{est}}(t)$ . Note that a mean or variance bias can be worked with ease into a Monte Carlo analysis; however, if these biases are known, it would make more sense to remove these biases from the retrievals immediately before processing them. Still another issue is our use of normal distributions. As noted earlier, PDFs for soil moisture may be strongly nonnormal. In concept, alternative distributions can be utilized directly in such a procedure as long as pairs of time series can be sampled from the distributions with prescribed time correlations. We avoid all of these issues in our examples because our goal here is solely to present an overall framework for generating a soil moisture error metric that reflects the needs of the application's user.

The chief practical limitation of the approach lies in the difficulty of knowing a priori the soil moisture PDF for a region in question. Optimally, this would be achieved through the analysis of historical soil moisture information; such information, however, is very difficult to obtain. In situ measurements are highly localized and nonexistent in most parts of the world. Existing space-based soil moisture retrievals are limited to the top few millimeters of soil, do not exist below dense vegetation, and suffer from their own bias issues (Reichle et al. 2007). Nevertheless, some indication of a region's soil moisture PDF can be obtained through multidecadal land model integrations driven with observations-based meteorological forcing (e.g., Sheffield et al. 2006) or through existing analytical solutions to the soil water balance equation forced by stochastic rainfall (Rodriguez-Iturbe et al. 1999). Data assimilation systems that combine such model products with satellite retrievals can further refine our quantification of soil moisture variability in the region (Reichle et al. 2007).

The practical difficulties associated with the approach—the construction of the soil moisture PDF, the determination of the functional form in (9), and so on—are real and may indeed limit its widespread application without additional research and analysis. Even so, at least conceptually, defining a soil moisture error metric in terms of an application's requirements is arguably more attractive than defining a soil moisture accuracy requirement (in terms of RMSE or correlation) based on, say, past conventional wisdom. Specifying, for example, a 0.04  $\text{m}^3 \text{m}^{-3}$  RMSE target for a desert that is mostly dry makes little sense, since arbitrarily choosing a low and constant soil moisture value (the climatological mean) would be just as effective under this metric as the most accurate measurement instrument. Specifying a 0.04  $\text{m}^3 \text{m}^{-3}$  RMSE target for a region that undergoes significant soil moisture variability may be overly harsh, given that it may translate to an overly precise estimation (given the needs of the user) of an application's quantity. The approach introduced here avoids such issues and thus has, in this sense, a strong conceptual advantage.

#### 4. Summary and conclusions

RMSE and correlation are two commonly used quadratic metrics that capture the degree of mismatch between retrievals and the true values of the measured variables. The RMSE metric is highly sensitive to biases in both mean and amplitude of fluctuations (such as a bias in standard deviation). In contrast, the correlation measure is indifferent to any bias in mean or amplitude of variations.

The correspondence of the retrieval estimates and the true values are evaluated differently by the RMSE and correlation statistics. Analysis of (5) and (6) shows that a target RMSE value ( $\text{RMSE}_{\text{target}}$ ) cannot be achieved if its magnitude lies below that of the bias in either the mean or the standard deviation. Furthermore, the assignment of climatology will trivially satisfy the accuracy target (in RMSE terms) if the standard deviation of the soil moisture being measured ( $\sigma_{\text{true}}$ ) lies below  $\text{RMSE}_{\text{target}}$ , provided the true mean is known. In contrast, regardless of the value of  $\sigma_{\text{true}}$ , a small but positive correlation would imply that the retrievals do contain potentially useful information in the context of change and/or anomaly detection.

It is assumed in much of our discussion that bias in the mean is known and removed from the RMSE calculations (i.e., we often focus on ubRMSE rather than RMSE itself). This may or may not be possible; much depends on the quality and quantity of available calibration and validation data. In any case, this problem relates to the RMSE metric only; any existing long-term biases in the mean or variance do not affect the correlation ( $r$ ) metric.

In this study we also introduce a nontraditional approach to defining soil moisture requirements. It is built around the idea that a user can define specific requirements for a given application's quantity and that these requirements, when combined with knowledge of the relationship between the applications quantity and soil moisture and with knowledge of the soil moisture PDF, can be transformed into the more traditional RMSE and  $r$  soil moisture metrics when defining specific mission validation requirements. Practical problems persist, in particular the need for accurate estimates of the soil moisture PDF and of the relationship between soil moisture and the quantity of relevance for the application. While these problems may limit the immediate application of the approach, they are not insurmountable and are left for future research. We envision that a further development of the framework can facilitate the interpretation and/or specification of measurement and validation requirements for SMAP and other future soil moisture satellite missions.

*Acknowledgments.* This study by members of the SMAP Science Definition Team was supported by NASA funding for the SMAP project.

#### REFERENCES

- Albertson, J. D., and G. Kiely, 2001: On the structure of soil moisture time series in the context of land surface models. *J. Hydrol.*, **243**, 101–119.
- Barnston, A. G., 1992: Correspondence among the correlation, RMSE, and Heidke forecast verification measures: Refinement of the Heidke score. *Wea. Forecasting*, **7**, 699–709.
- Budyko, M. I., 1974: *Climate and Life*. Academic Press, 508 pp.
- Cosh, M. H., T. J. Jackson, P. J. Starks, and G. Heathman, 2006: Temporal stability of surface soil moisture in the Little Washita River Watershed and its applications in satellite soil moisture product validation. *J. Hydrol.*, **323**, 168–177.
- Crow, W. T., 2007: A novel method for quantifying value in spaceborne soil moisture retrievals. *J. Hydrometeorol.*, **8**, 56–67.
- , and E. F. Wood, 1999: Multi-scale dynamics of soil moisture variability observed during SGP'97. *Geophys. Res. Lett.*, **26**, 3485–3488.
- Drusch, M., E. F. Wood, and H. Gao, 2005: Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophys. Res. Lett.*, **32**, L15403, doi:10.1029/2005GL023623.
- Eagleson, P. S., 1978: Climate, soil and vegetation 4. The expected value of annual evapotranspiration. *Water Resour. Res.*, **14**, 731–739.
- Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhay, 2000: Temporal and spatial scales of observed soil moisture variations in the extratropics. *J. Geophys. Res.*, **105** (D9), 11 865–11 877.
- Famiglietti, J. S., and Coauthors, 1999: Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) hydrology experiment. *Water Resour. Res.*, **35**, 1839–1851.
- , D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, 2008: Field observations of soil moisture variability across scales. *Water Resour. Res.*, **44**, W01423, doi:10.1029/2006WR005804.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, doi:10.1016/j.jhydrol.2009.08.003.
- Hu, Z., S. Islam, and Y. Cheng, 1997: Statistical characterization of remotely sensed soil moisture images. *Remote Sens. Environ.*, **61**, 310–318.
- Koster, R. D., and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, **10**, 1578–1591.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335.
- Martínez-Fernández, J., and A. Ceballos, 2003: Temporal stability of soil moisture in a large-field experiment in Spain. *Soil Sci. Soc. Amer. J.*, **67**, 1647–1656.
- Mohanty, B. P., and T. H. Skaggs, 2001: Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv. Water Resour.*, **24**, 1051–1067.
- Murphy, A. H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting*, **10**, 681–688.
- Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg, 2004: Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeorol.*, **5**, 430–442.
- , —, 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.
- Rodriguez-Iturbe, I., G. K. Vogel, R. Rigon, D. Entekhabi, F. Castelli, and A. Rinaldo, 1995: On the spatial organization of soil moisture fields. *Geophys. Res. Lett.*, **22**, 2757–2760.
- , A. Porporato, L. Ridolfi, V. Isham, and D. R. Cox, 1999: Probabilistic modelling of water balance at a point: The role of climate, soil and vegetation. *Proc. Roy. Soc. London*, **455A**, 3789–3805.
- Ryu, D., and J. S. Famiglietti, 2005: Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment. *Water Resour. Res.*, **41**, W12433, doi:10.1029/2004WR003835.
- Scipal, K., T. Holmes, R. A. M. de Jeu, V. Naeimi, and W. Wagner, 2008: A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophys. Res. Lett.*, **35**, L24403, doi:10.1029/2008GL035599.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. WMO World Weather Watch Tech. Rep. 8, WMO/TD 358, 114 pp.
- Teuling, A. J., and P. A. Troch, 2005: Improved understanding of soil moisture variability dynamics. *Geophys. Res. Lett.*, **32**, L05404, doi:10.1029/2004GL021935.