Upscaling of soil moisture content from surface to profile: multi-station testing of observation operators

Xiaodong Gao[1,2,3], Xining Zhao[1,2,3], Luca Brocca[4], Ting Lv[3], Gaopeng Huo[3], Pute Wu[1,2,3]

[1]Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi Province, China
[2]Institute of Soil and Water Conservation, CAS and MWR, Yangling, Shaanxi Province, China
[3]National Engineering Research Center for Water Saving Irrigation at Yangling, Yangling, Shaanxi, China.
[4]Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy

Correspondence to: Xining Zhao (xiningz@aliyun.com)

Abstract. The accurate assessment of profile soil moisture for spatial domains is usually difficult due to the associated costs, strong spatial-temporal variability, and nonlinear relationship between surface and profile moisture. Here we attempted to use observation operators built by Cumulative Distribution Frequency (CDF) matching method to directly upscale surface observations to profile soil moisture based on multi-station in situ measurements from the Soil and Climate Analysis Network (SCAN). We first analyzed the effects of temporal resolution (hourly, daily and weekly) and data length (half year in non-growing season, half year in growing season, one year, two years and four years) on the performance of observation operators. The results showed that temporal resolution had a negligible influence on the performance of observation operators. However, data length significantly changed the prediction accuracy of observation operators; prediction errors decreased as data length increased from half year (non-growing season) to two years, but accuracy did not further improve at longer interval. A dataset with a two-year duration was therefore used to test the robustness of observation operators in three primary climates (humid continental, humid subtropical and semiarid) of the continental USA, with the popular exponential filter employed as a reference approach. The results indicated that observation operators generally performed better than exponential filter method in both calibration and validation periods. This suggests that observation operators are a robust statistical tool for upscaling soil moisture from surface to profile. The findings here may be applied in the prediction of profile soil moisture from surface measurements obtained *via* various means, including remote sensing techniques.

1 Introduction

Soil moisture in the root-zone profile is a key variable that influences the agricultural, hydrological, ecological and meteorological systems of the Critical Zone. It determines water availability to crops and participates in the partitioning of available energy into sensible and latent heat and of precipitation into infiltration and runoff (Montaldo and Albertson, 2003). Continuous and accurate measurement of profile soil moisture, however, is difficult because of expensive field measurements, strong spatial-temporal variability, and the nonlinear relationship between surface and profile soil moisture

(Han et al., 2012a; Dumedah et al., 2015). In contrast, the collection of surface soil moisture data is much easier. At fine scales (e.g. field, hillslope, or small watershed), surface soil moisture can be attained from portable soil moisture sensors or derived using cosmic-ray neutron probes and ground-penetrating radar (Penna et al., 2013; Ferrara et al., 2013; Baatz et al., 2014). At larger scales (e.g. regional and global scales), remote sensing techniques can provide surface soil moisture data at

5   fine spatial or temporal resolution (Brocca et al., 2011; Panciera et al., 2014). However, these instruments generally measure soil moisture at relatively shallow layers. For example, remote sensing methods (either active or passive) retrieve near-surface soil moisture that only reaches several centimeters beneath the soil surface (Draper et al., 2009). Therefore, it is necessary to link surface soil moisture to profile soil moisture via robust depth-upscaling approaches.

10   A variety of approaches for predicting profile soil moisture from surface measurements have been proposed, ranging from simple statistical relationships to physically-based retrieval (Wagner et al., 1999). The primary methods used today can generally be classified into three different types: (1) data assimilation methods; (2) analytical methods; and (3) statistical or computational statistical methods. Data assimilation methods refer to techniques which incorporate surface soil moisture measurements (e.g. remote sensing products) into physically-based hydrologic models to obtain an analysis that best

15   represents profile soil moisture. Various data assimilation algorithms, based on direct insertion, Kalman filter, or the extended and ensemble Kalman filter, among others, have already been developed (Evensen, 1994; Walker et al., 2002; Reichle et al., 2002; Heathman et al., 2003). These algorithms have provided robust estimates of root-zone soil moisture by integrating surface soil measurements, either remote sensing retrievals or *in situ* observations from a soil moisture monitoring network, into land surface models (Crow and Wood, 2003; Reichle et al., 2007; Draper et al., 2011; Han et al.,

20   2012a; Parrens et al., 2014; Dumedah et al., 2015). Although data assimilation is considered the most promising approach (Kostov and Jackson, 1993; Dumedah et al., 2015), its application may be constrained by the required model parameters (soil properties, vegetation features and atmospheric forcing), which are difficult to obtain at larger scales, as well as by uncertainties related to the physical description of soil hydrological processes (Albergel et al., 2008; Hu and Si, 2014).

25   The analytical methods require fewer input parameters and are computationally more efficient than data assimilation methods. They are generally mathematically derived from physically-based relationships of water flows that include some simplification assumptions (Arya et al., 1983; Camillo and Schmugge, 1983; Wagner et al., 1999; Manfreda et al., 2014). For example, the exponential filter, introduced by Wagner et al. (1999), assumes that water fluxes between surface and subsurface layers are proportional to the soil moisture difference between these two layers. Currently, the exponential filter

30   is likely the most popular analytical method since it only requires one input parameter, the characteristic time length (T), in addition to surface soil moisture observations. This method has successfully predicted subsurface soil moisture from surface observations for multiple regions that vary in climatic and/or soil conditions (Ceballos, et al., 2005; Albergel et al., 2008; Brocca et al., 2011; Ford et al., 2014; Peterson et al., 2016).

Statistical methods are even simpler approaches for estimating profile soil moisture because they do not account for water flux processes and are completely data driven. The simplest statistical method is the linear regression between surface and profile soil moisture data. However, this method usually produces poor estimations because of the nonlinear relationship

5 between surface and profile soil moisture (Jackson, 1986; Bono and Alvarez, 2012; Shi et al., 2014). Bono and Alvarez (2012) found that multivariate regression and artificial neural network was able to produce reliable profile soil moisture estimations, but required soil moisture measurements from at least two different soil layers as inputs. Time stability analysis, introduced by Vachaud et al. (1985), has been also tested for its application to soil moisture depth upscaling. Hu and Si (2014) indicated that profile soil moisture can be reliably predicted from measurements at the most time stable depth.

10 However, the time stable depth is not necessarily the surface layer, and this would make required measurements more costly and difficult to obtain. Furthermore, Gao et al. (2015) found that measurements of a shallow layer, when used as inputs in time stability analysis, were able to accurately estimate profile soil moisture. However, the depth of the shallow layer was defined to reach down to 40 cm, which is far beyond the scope of satellite sensors. This restricts the application of time stability analysis to profile soil moisture estimation because in many cases only surface measurements (≤ 5 cm) are available.

15

Despite the development of data assimilation and analytical methods, soil moisture upscaling still needs robust statistical methods because of their high computational efficiency and applicability to a wide range of environments (Bono and Alvarez, 2012). The Cumulative Distribution Frequency (CDF) matching method has proven to be both efficient and widely applicable through the construction of observation operators. This method is primarily used to adjust for the systematic

20 differences in soil moisture that arise from different sources such as in situ measurements, modeled outputs and remote sensing retrievals (Reichle and Koster, 2004; Drush et al., 2005; Brocca et al., 2011; Parrens et al., 2014). The method has also been extended to the spatial upscaling of point soil moisture measurements (Han et al., 2012b) and spatial transferring of soil moisture between different areas (Gao et al., 2013). The CDF matching method is expected to be applicable to the presented study because soil moisture at various layers can be regarded as belonging to different spatial domains or sources.

25 In this way, profile soil moisture could be predicted by adjusting for surface soil moisture through the construction of the observation operators for surface and profile data.

The primary objective of the presented research was to test the feasibility and robustness of observation operators built by CDF matching method in upscaling surface soil moisture to profile soil moisture by using multi-station in situ soil moisture

30 observations from the Soil Climate Analysis Network (SCAN) in the United States. The paper is arranged as follows. First, we analyzed how the resolution and length of soil moisture time series affected the performance of observation operators. The results were used to choose an appropriate temporal resolution and length for the construction of observation operators.

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

We then tested the feasibility and robustness of observation operators under different climate regions in the continental United States.

2 Materials and methods

2.1 Soil moisture datasets

5 Here the feasibility and robustness of a CDF matching method in the upscaling of soil moisture from surface to profile was tested using in situ soil moisture measurements from the Soil Climate Analysis Network (SCAN).

SCAN focuses primarily on agricultural areas in the US and consists of over 200 soil moisture monitoring stations across different climate regions, mainly serving to monitor drought and climate change. Soil moisture content at depths of 5, 10, 20,

10 50, and 100 cm was measured at each station with HydraProbe soil moisture sensors (Stevens Water Monitoring Systems Inc., Portland, OR). All SCAN soil moisture data, at both daily and hourly resolution, are available at the National Water and Climate Center website (www.wcc.nrcs.usda.gov/scan). A total of 31 stations were chosen for analyses according to the objectives of this study. General information about these 31 stations is presented in Table 1. Although these data have been corrected by data managers, we were able to detect outliers in the datasets. To identify outliers in one given layer, soil

15 moisture at this depth was linked to values at adjacent depth(s) and rainfall events. Generally, an increase in soil moisture is closely associated with a precipitation event. When fluctuations in soil moisture at a given layer were not related with rainfall events and independent of the fluctuations of adjacent layers, these soil moisture values were identified as outliers and excluded from the analyses.

2.2 Observation operators

20 **2.2.1 The CDF matching method**

The Cumulative Distribution Frequency (CDF) matching method was used to create observation operators. The observation operators were then used to predict profile soil moisture content from surface measurements by adjusting for the systematic differences between these two layers. It is important to note that the CDFs used in this study were built from soil moisture time series and assume that all soil moisture values are equally probable (Pachepsky and Hill, 2016).

25

As shown in Figure 1, the CDF matching method rescales the CDF of one dataset (surface soil moisture time series in this study) to match that of another dataset (profile soil moisture time series in this study). The technical procedure of this method progressed as follows. First, the in situ measured surface and profile soil moisture datasets were ranked. Next, the differences (Δ) in soil moisture between corresponding elements in the surface and profile datasets were calculated. A

30 polynomial fit was then used to quantify the relationship between surface soil moisture and Δ, which enabled the creation of

4

Hydrology and
Earth System
Sciences
Discussions

observation operators. Finally, Profile soil moisture could then be estimated by using the observation operators to rescale surface measurements. Previous studies (Reichle and Koster, 2004; Drush et al., 2005; Han et al., 2012b) have usually used a third-order polynomial fit to establish observation operators. This study, on the other hand, employed pre-experiments to identify the optimal order (results not shown here), and a fifth-order polynomial was finally used when considering the accuracy of fitting and the principle of parsimony.

2.2.2 Time series resolution effect

Three different temporal resolutions (hourly, daily, and weekly) were used to probe how time series resolution affects the performance of observation operators. To this end, three stations from SCAN with varying soil moisture values and surface-profile soil moisture dynamics were selected for analysis, including Shagbark Hills in Iowa, Perdido Riv Farm in Alabama, and Sevilleta in New Mexico. These stations were chosen because of the continuity and completeness of soil moisture datasets at both daily and hourly resolutions.

The testing procedure progressed as follows. First, hourly data within a given period were used to build observation operators. Surface moisture at daily (weekly) resolution during the same period was then incorporated into these observation operators to predict the corresponding daily (weekly) profile soil moisture. Next, observation operators derived from daily data were used to predict hourly (weekly) profile soil moisture in corresponding periods. Finally, observation operators derived from weekly data were used to predict hourly (daily) profile soil moisture in corresponding periods. Statistical metrics, including determination coefficient (R^2), root mean square error (RMSE) and Nash-Sutcliffe coefficient (NSC; Nash and Sutcliffe, 1970), were used to judge whether the developed observation operators were transferable between different temporal resolutions. The RMSE and NSC are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\theta_{obs,i}-\theta_{est,i})^2} \tag{1}$$

and

$$\text{NSC} = 1 - \frac{\sum_{i=1}^{N}(\theta_{obs,i}-\theta_{est,i})^2}{\sum_{i=1}^{N}(\theta_{obs,i}-\overline{\theta}_{obs})^2} \tag{2}$$

where $\theta_{obs,i}$ and $\theta_{est,i}$ represent the observed and estimated profile soil moisture content, respectively, and N is the number of soil moisture values in the corresponding time series.

2.2.3 Time series length effect

Three stations (Blue Creek, BC; Green River, GR; and Little Red Fox, LRF) in Utah with varying surface and profile soil moisture time series from 2010 to 2015, as well as continuous and complete datasets, were used to probe how time series length affects the performance of observation operators. Five different data lengths were chosen: a half-year growing season

5   (Apr. 1 to Sept. 30; DL1); a half-year non-growing season (Oct. 1 to Mar. 31; DL2); one calendar year (DL3); two calendar years (DL4); and four calendar years (DL5). Four replicates were conducted for each data length. Specifically, soil moisture from the years of 2010, 2011, 2014, and 2015 was used to establish the observation operators for data lengths DL1, DL2 and DL3 (calibration), and data from 2012 and 2013 were used for validation. For data length DL4, soil moisture from four different two-year combinations (2010&2011, 2014&2015, 2010&2014 and 2011&2015) was used for calibration, and data

10   from 2012 and 2013 were used for validation. As data length DL5 covers four years, four different four-year combinations (2010&2011&2012&2013, 2011&2012&2013&2014, 2012&2013&2014&2015, and 2010&2011&2014&2015) were used for calibration. The two years that were not included in each of the specified combinations were used for validation. Statistical metrics (R^2, RMSE, and NSC) were used to test the feasibility and robustness of observation operators built from soil moisture datasets that varied in the length of both calibration and validation periods. These analyses served to identify an

15   appropriate data length for the construction of robust observation operators. Note that the surface measurements should also be ranked first during validation, after which they can act as inputs for the observation operators that derive ranked profile soil moisture.

2.2.4 Testing procedures in applications

The constructed observation operators were then applied to different climate regions to test how robust the predictions are

20   for areas that vary in soil moisture. Three primary climate regions in the continental USA were chosen (humid continental, humid subtropical, and semiarid) and three stations were selected for each climate region based on the continuity and completeness of soil moisture datasets. Detailed information regarding these nine stations is presented in Table 1.

2.3 Exponential filter

A popular analytical method, the exponential filter, the robustness of which has been validated in various climates (Wagner

25   et al., 1999; Ceballos et al., 2005; Albergel et al., 2008; Brocca et al., 2011; Ford et al., 2014; Gao et al., 2014; Peterson et al., 2016), served as a reference method to judge the performance of observation operators in different climate regions.

According to Wagner et al. (1999), a soil profile can be divided into the surface layer and a second (subsurface) layer. The exponential filter was introduced to predict second-layer soil moisture (w_2) from surface measurements (w_1). This method

30   assumes that the water flux between surface and subsurface layers is proportional to the difference in soil moisture between these two layers. The surface and subsurface soil moisture can be linked as follows:

$$L_2 \frac{dw_2(t)}{dt} = C\left[w_1(t) - w_2(t)\right] \tag{3}$$

Assuming C is constant and $T=L/C$, the solution of Eq. (3), is derived by integration as:

$$w_2(t) = \frac{1}{T} \int_{-\infty}^{t} w_g(\tau) \exp(-\frac{t-\tau}{T}) d\tau \tag{4}$$

where L_2 is the depth of the second layer, C is the area-representative pseudo-diffusivity constant, and T represents the

5 characteristic time length. Generally, T is considered to be an integrative parameter that explains all the hydrologic, pedologic and ecological processes that influence soil moisture variations with depth (Wagner et al., 1999; Albergel et al., 2008). To simplify the computation, Albergel et al. (2008) gave the recursive formulation of Eq. (4) as follows:

$$\mathrm{SWI}_{2(t)} = \mathrm{SWI}_{2(t-1)} + K_t[\mathrm{SWI}_{1(t)} - \mathrm{SWI}_{2(t-1)}] \tag{5}$$

and

10 $$K_t = \frac{K_{t-1}}{K_{t-1} + \exp(\frac{\Delta t}{T})} \tag{6}$$

where SWI_1 and SWI_2 represent the soil water index (SWI) of the surface and second layer, respectively, and K_t is the gain of the exponential filter. The dimensionless SWI represents the scaled soil moisture content, which ranges from 0 to 1 based on the minimum and maximum values of each time series.

15 It is important to note that predicted second-layer soil moisture (SWI_2) *via* Eq. (5) provides scaled values. To compare these values with results from observation operators, SWI_2 must be rescaled by using the maximum ($w_{2,\max}$) and minimum ($w_{2,\min}$) values of the corresponding original time series of the second layer as follows:

$$\hat{w}_2(t) = \mathrm{SWI}_{2(t)}(w_{2,\max} - w_{2,\min}) + w_{2,\min} \tag{7}$$

where $\hat{w}_2(t)$ is the rescaled value of predicted soil moisture at the second layer, $\mathrm{m}^3\,\mathrm{m}^{-3}$. The rescaled value $\hat{w}_2(t)$ and the

20 surface measurements $w_1(t)$ can then be coupled to obtain the profile soil moisture as:

$$\hat{w}_p(t) = \frac{\hat{w}_2(t) \times L_2 + w_1(t) \times L_1}{L_2 + L_1} \tag{8}$$

where $\hat{w}_p(t)$ is the predicted profile soil moisture at time t, and L_1 is the depth of surface layer.

3 Results & discussion

3.1 The effect of time series resolution

The surface and profile soil moisture time series at different temporal resolutions for the Shagbark Hills, Perdido Riv Farms and Sevilleta stations are shown in Figure 2. Soil moisture is not shown at an hourly resolution because it generally

5   overlapped with daily data. In general, the surface and profile soil moisture at different resolutions did not show different temporal trends. However, as the resolution became coarser, more information was lost. For instance, daily resolution data cannot reflect the hourly variation of soil moisture within a day and weekly resolution data may not include information about the daily variation of soil moisture, potentially failing to capture extreme values following rainfall events and strong evapotranspiration (Figure 2).

10

The performance metrics for the application of observation operators built from a dataset of a certain resolution to another dataset with a different resolution are presented in Table 2. Generally, observation operators that had been calibrated using datasets of a different resolution performed well, with RMSE less than 5.39×10^{-3} m^3 m^{-3} and R^2 and NSC higher than 0.99. Furthermore, there was little variation between the three statistical metrics from all stations when the operators were

15   transferred from daily to hourly datasets. However, a clear increase in RMSE was observed when the observation operators built from either daily or hourly data were applied to weekly data. In contrast, the observation operators built from weekly data showed no decrease in RMSE when they were applied to either hourly or daily data. This suggests that observation operators built from datasets with finer temporal resolutions can lose a certain degree of accuracy when they are applied to coarser resolution data. However, observation operators derived from coarser resolution datasets seem to be less prone to

20   losing accuracy when applied to finer resolution data. Overall, the temporal resolution of soil moisture time series had a relatively weak effect on the performance of observation operators, and hence can be ignored during the construction of observation operators. In this study, observation operators were built from soil moisture data at a daily resolution in order to reach a compromise between computing efficiency and operator robustness.

3.2 The effect of time series length

25   Soil moisture data for the three stations (LRF, GR and BC) over a six-year interval are presented in Figure 3. Overall, the soil moisture time series varied between stations. At the LRF station, surface moisture content was generally lower than profile moisture content. Differently, soil moisture content was generally higher at the surface than in the profile at the GR site, except during very dry conditions. Data from the BC station showed that surface soil moisture was generally higher than profile soil moisture during wet conditions, but lower than profile soil moisture during dry conditions.

30

A cross correlation analysis was used to characterize surface-profile soil moisture relationships. This is because subsurface soil moisture shows a delayed response to atmospheric variables (e.g. precipitation and evapotranspiration) and cross correlation analysis is able to characterize lagged relations (Georgakakos et al., 1995; Mahmood et al., 2012; Ford et al., 2014). As shown in Figure 4, cross correlation coefficients varied greatly over stations and time. At the LRF station, the
5   cross correlations were relatively weak in 2010 and 2012, with a maximum coefficient of less than 0.04; in 2010 surface and profile soil moisture were negatively correlated at all lag times (< 50 days). This indicated that, in 2010, profile soil moisture was weakly influenced by surface wetness and may have been largely controlled by groundwater. When compared to the other years, these two years can be regarded as anomalous years from the perspective of surface-profile soil moisture relationships. Over the other years, maximum cross correlation at LRF ranged from 0.42 to 0.80, occurring with no lag or at
10   one day of lag, except for 2013, when the coefficient peaked at a lag time of 11 days. The large variation in cross correlation over different years at the LRF station means that observation operators built by using data from one period may fail when applied to other periods. At the GR station, the maximum cross correlation ranged from 0.26 to 0.61 over the analyzed years and peaked at no lag or one-day lag, except for 2010, when maximum correlation occurred at a 10 day lag. The BC station showed relatively stable and large maximum cross-correlation coefficients of 0.62 to 0.77 over the years, peaking at a lag
15   time from 1 day to 20 days. This means that surface and profile soil moisture at this station are well coupled across time periods and the observation operators built in one period could be successfully transferred to other periods.

The effects of soil moisture time series data length on the performance of observation operators in both calibration and validation periods are shown in Figure 5. As expected, observation operators clearly performed better in calibration than in
20   validation periods, especially when the RMSE and NSC metrics are studied. However, the performance in validation periods can serve as a better reference for judging data length effects because independent datasets are employed. Except for the LRF station, the performance of observation operators, based on both calibration and validation periods, improved as data length increased from half year (non-growing season; DL1) to two years (DL4); this was shown by an increase in predicted error (RMSE) and a decrease in correlated strength (R^2 and NSC). As shown in Figure 3, the non-growing season generally
25   had moderate soil moisture contents and did not include high and low values. As a result, observation operators built from non-growing season data would fail to represent the surface-profile soil moisture relationships across the annual level, thus resulting in large predicted errors. As data length increased, the observation operators became more representative, producing more robust predictions of profile soil moisture. However, observation operators did not perform better according to all three statistical metrics when data length further increased to four years (DL5), and even performed slightly worse
30   during validation periods than DL4-based operators in light of RMSE and R^2 values. This suggests that larger datasets do not produce greater gains when data length exceeds two years. Therefore, the observation operators in this study were built based on a data length of two years.

Special attention was given to the LRF station, which showed large prediction errors and low correlation between predictions and observations during both calibration and validation periods when compared with the other stations. This was not surprising based on the findings from Figures 3 and 4. These two figures demonstrate that the surface-profile soil moisture relationships in 2010 and 2012 varied from other years. In this study, soil moisture from 2010 was used as part of the

5 calibration data and soil moisture from 2012 was used as part of the validation data. Therefore, observation operators built using datasets that included soil moisture values from 2010 would differ greatly from those not based on these soil moisture data. Validation of these observation operators in a normal year (i.e. 2013) or in a different anomalous year (i.e. 2012) would result in large prediction errors. This suggests that datasets describing years with anomalous surface-profile soil moisture relationships should be excluded in order to build robust observation operators.

10 **3.3 Testing of observation operators across climate regions**

The feasibility and robustness of observation operators was tested using stations distributed across three primary climate regions in the continental USA. As stated above, the exponential filter method was employed as a reference to further verify the robustness of observation operators.

15 The first step of the exponential filter method is the calculation of the optimal value for the T parameter (T_{opt}), which is obtained by maximizing the Nash-Sutcliffe coefficient (NSC); the T value at which NSC peaks is regarded as the T_{opt} for a given dataset (Wagner et al., 1999). The relationships between the T parameter and NSC at various stations in calibration period are shown in Figure 6. The T_{opt} varied from 1 to 23 days for different stations, except for the Holden station, which had negative NSC values over the whole range of T values under consideration (0-50 days) and did not even peak when T

20 was set to 100 days (data not shown here). In-depth analysis showed that the T_{opt} value was highly dependent on the type of climate. The average T_{opt} values for the humid continental, humid subtropical and semiarid climates were 14.0, 1.3 and 17.5 days, respectively. This is partly consistent with the findings of Albergel et al. (2008), who also found that climate impacts T_{opt}, yet reported a relatively weak effect. According to Albergel et al. (2008), the T_{opt} reflects the response of second-layer soil moisture (w_2) to surface moisture (w_1), with a higher T_{opt} representing a slower response. In the humid subtropical

25 climate, a relatively high soil moisture content over the whole profile (see Figure 9) reflects large soil hydraulic conductivity and a relatively fast response of w_2 to w_1. In the semiarid climate, however, both surface and subsurface soil moistures are relatively low, which results in low soil hydraulic conductivity and a slower second-layer response to surface infiltration.

Prior to testing, each soil moisture time series was checked to avoid anomalous data. Soil moisture datasets from 2014 and

30 2015 were used as calibration data for both methods at all stations except for Perdido Riv Farms, where datasets from 2013 and 2014 were employed due to considerable missing values in 2015. Data from either 2013 or 2016 were used for validation depending on the completeness of dataset. Figures 7 to 9 show the predicted profile soil moisture time series from

observation operators and exponential filters in humid continental, humid subtropical and semiarid climates, respectively. Statistical metrics of the predicted results for all three climates are shown in Figure 10. The observation operators clearly performed better than the exponential filter method during the calibration; the predicted profile soil moisture almost completely matched the observed values at different stations, with R^2 and NSC higher than 0.97 and RMSE less than

5   6.74×10^{-3} m^3 m^{-3}. This can be attributed to the perfect adjusting of the cumulative distribution frequencies of surface moisture and profile data by observation operators (the left graphs in Figure 7-9). The exponential filter method generally reproduced the observed time series well, although apparent deviations existed in some stations, such as the Molly Caren station in the humid continental climate, for which the RMSE reached 5.83×10^{-2} m^3 m^{-3} and R^2 and NSC were only 0.35 and 0.25, respectively. As shown in Figure 11, however, the prediction error (RMSE) of observation operators clearly increased,

10   ranging from 8.16×10^{-3} to 2.34×10^{-2} m^3 m^{-3}, and the matching degree (NSC) of predicted and observed values clearly decreased, ranging from -0.04 to 0.99, when they were applied to validation periods. In contrast, the exponential filter method maintained relatively stable performance from calibration to validation periods, except for the NSC metric in the semiarid climate, for which a very low value was obtained at the Holden station. Nonetheless, the statistical metrics of observation operators remained superior to those of the exponential filter method also in the validation period. These results

15   indicate that observation operators built from a soil moisture dataset with a length of two years, and which excludes anomalous values, were robust in upscaling surface soil moisture to obtain profile predictions.

Figure 10 shows that the climate region affects, to different extents, the performance of both observation operators and the exponential filter method. For the observation operators, a semiarid climate showed significantly ($P<0.05$) lower RMSE than

20   the other climates in calibration. However, the observation operators produced relatively larger estimation errors and lower values of R^2 and NSC in the semiarid climate during validation periods, indicating a relatively weak correlation between predicted and observed profile soil moisture. Similarly, the semiarid climate also produced lower values of R^2 and NSC than the other climates for the exponential filter method. However, the semiarid climate showed relatively lower prediction error than other climates, especially in the calibration period, probable due to the lower soil moisture values there. The lower NSC

25   values calculated for the semiarid climate could be the consequence of the lower correlation between surface and profile soil moisture content in this climate. The maximum cross correlations in the semiarid climate during calibration and validation periods were 0.592 and 0.666, respectively. They were lower than the corresponding values in humid continental (0.736 in calibration, 0.738 in validation) and humid subtropical (0.932 in calibration, 0.882 in validation) climates. This indicates that a relatively poor coupling between surface and profile soil moisture exists in semiarid climates.

4. Conclusions

In this study, we tested the feasibility and robustness of observation operators derived by Cumulative Distribution Frequency (CDF) matching for the upscaling of soil moisture from surface to profile by using datasets of multiple stations in the Soil and Climate Analysis Network (SCAN). In general, the time series resolution (hourly, daily and weekly) had negligible effects on the performance of observation operators and daily resolution provided a good compromise between computing efficiency and operator robustness. In contrast, time series length was shown to significantly affect the prediction accuracy of observation operators. Our analyses showed that soil moisture data from a two-year interval produced optimal predictions, and that a longer data period did not improve prediction accuracy. Therefore, soil moisture measurements over a two-year interval can be used as the standard dataset from which to build observation operators.

By using the popular exponential filter method as a reference approach we were able to show that the application of observation operators to three primary climates (humid continental, humid subtropical and semiarid) in the continental USA can successfully upscale soil moisture from surface to profile. The findings described here may be applicable to the prediction of profile soil moisture from surface measurements obtained through various techniques, including remote sensing.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: 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, 2008.

Arya, L.M., and Richter, J.C.: Estimating profile water storage from surface zone soil moisture measurements under bare field conditions, Water Resour. Res., 19(2), 403-412, 1983.

Baatz, R., Bogena, H.R., Franssen, H.J.H., Huisman, J.A., Qu, W., Montzka, C., and Vereecken H.: Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods, J. Hydrol., 516, 231-244, 2014.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., and Bittelli, M.: Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe, Remote Sens. Environ., 115, 3390–3408, 2011.

Bono, A., and Alvarez, R.: Use of surface soil moisture to estimate profile water storage by polynomial regression and artificial neural networks, Agron. J., 104, 934–938, 2012.

Camillo, P., and Schmugge, T. J.: Estimating soil moisture storage in the root zone from surface measurements, Soil Science, 135(4), 245-264, 1983.

Ceballos, A., Scipal, K., Wagner, W., and Martinez-Fernandez, J.: Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain, Hydrol. Process., 19, 1549-1566, 2005.

Chen, F., Crow, W., Starks, P. J., and Moriasi, D. N.: Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, Adv. Water Resour., 34, 526-536, 2011.

Crow, W. T., and Wood, E. F.: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97, Adv. Water Resour., 26, 137–149, 2003.

Dorigo, W. A., WQagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Qevelen, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrol. Earth Syst. Sci., 15, 1675-1698, 2011.

Draper, C. S., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France, Hydrol. Earth Syst. Sci., 15, 3829–3841, 2011.

Drusch, M., Wood, E.F., and Gao, H.: Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. Geophys. Res. Lett. 32 (15), 2005.

Dumedah, G., Walker, J.P., and Merlin, O.: Root-zone soil moisture estimation from assimilation of downscaled Soil Moisture and Ocean Salinity data, Adv. Water Resour., 84, 14-22, 2015.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res. Ocean, 99: 10143–10162, 1994.

Ferrara, C., Barone, P.M., Steelman, C.M., Pettinelli, E., and Endres, A.L.: Monitoring Shallow Soil Water Content Under Natural Field Conditions Using the Early-Time GPR Signal Technique, Vadose Zone J., 12(4), 2013.

Ford, T. W., Harris, E., and Quiring, S. M.: Estimating root zone soil moisture using near-surface observations from SMOS, Hydrol. Earth Syst. Sci., 18, 139–154, 2014.

Heathman, G.C., Starks, P.J., Ahuja, L.R., and Jackson, T.J.: Assimilation of surface soil moisture to estimate profile soil water content, J. Hydrol., 279, 1-17, 2003.

Gao, X., Wu, P., Zhao, X., Zhou, X., Zhang, B., Shi, Y., and Wang, J.: Estimating soil moisture in gullies from adjacent upland measurements through different observation operators, J. Hydrol., 486, 420-429, 2013.

Gao, X.D., Zhao, X.N., Zhang, B.Q., and Wu, P.T.: Estimation of root-zone soil moisture over gullies using an exponential filter, Adv. Water Sci., 25(5), 684-694, 2014. (in Chinese with an English abstract)

Hydrology and
Earth System
Sciences
Open Access
Discussions

EGU

Gao, L., Lv, Y.J., Wang, D.D., Tahir, M., and Peng, X.H.: Can shallow-layer measurements at a single location be used to predict deep soil water storage at the slope scale, J. Hydrol., 531, 534-542, 2015.

Georgakakos, K.P., Bae, D.G., and Cayan, D.R: Hydroclimatology of continental watersheds. 1. Temporal analyses. Water Resour. Res., 31, 655–675,1995.

Han, E., Merwade, V., and Heathman, G.: Application of data assimilation with the Root Zone Water Quality Model for soil moisture profile estimation in the upper Cedar Creek, Indiana, Hydrol. Process., 26, 1707-1719, 2012a.

Han, E., Heathman, G.C., Merwade, V., and Cosh, M.H.: Application of observation operators for field scale soil moisture averages and variances in agricultural landscapes, J. Hydrol., 444-445, 34-50, 2012b.

Hu, W., and Si, B.C.: Can soil water measurements at a certain depth be used to estimate mean soil water content of a soil profile at a point or at a hillslope scale, J. Hydrol., 516, 67–75, 2014.

Jackson, T.J.: Soil water modeling and remote sensing, IEEE Trans. Geosci. Rem. Sens. GE-24:37–46, 1986.

Kostov, K.G., and Jackson, T.J.: Estimating profile soil moisture from surface layer measurements—a review, SPIE 1941, 125–136, 1993.

Manfreda, S., Brocca, L., Moramarco, T., Melone, F., and Sheffield, J.: A physically based approach for the estimation of root-zone soil moisture from surface measurements, Hydrol. Earth Syst. Sci., 18, 1199-1212, 2014.

Montaldo, N., and Albertson, J.D.: Multi-scale assimilation of surface soil moisture data for robust root zone moisture predictions, Adv. Water Resour., 26, 33-44, 2003.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models, part II – a discussion and principles, J. Hydrol., 10, 282-290, 1970.

Pachepsky, Y., and Hill, R.L.: Scale and scaling in soils, Geoderma, doi: 10.1016/j.geoderma.2016.08.017, 2016.

Panciera, R., Walker, J.P., Jackson, T.J., Gary, D.A., Tanase, M.A., Ryu, D., Monerris, A., Yardley, H., Rudiger, C., and Wu, X.L.: The Soil Moisture Active Passive Experiments (SMAPEx): Toward Soil Moisture Retrieval From the SMAP Mission, IEEE Trans. Geosci. Remote Sens., 52(1), 490-507, 2014.

Parrens, M., Mahfouf, J.F., Barbu, A.L., and Calvet, J.C.: Assimilation of surface soil moisture into a multilayer soil model: design and evaluation at local scale, Hydrol. Earth Syst. Sci., 18, 673–689, 2014.

Peterson, A. M., Hlgason, W. D., and Ireson, A. M.: Estimating field-scale root zone soil moisture using the cosmic-ray neutron probe, Hydrol. Earth Syst. Sci., 20, 133-1385, 2016.

Reichle, R.H., Walker, J.P., Koster, R.D., and Houser, P.R.: Extended versus Ensemble Kalman Filtering for land data assimilation. J. Hydrometeorol., 3, 728–740, 2002.

Reichle, R.H., and Koster, R.D.: 2004. Bias reduction in short records of satellite soil moisture, Geophys. Res. Lett., 31 (19).

Reichle, R.H., Koster, R.D., Liu, P., Mahanama, S. P., Njoku, E. G., Owe, M.: 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. Atmos., 112, 2007.

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

Shi, Y.G., Wu, P.T., Zhao, N.X., Li, H.C., Wang, J.W., and Zhang, B.Q.: Statistical analyses and controls of root-zone soil moisture in a large gully of the Loess Plateau, Environ. Earth Sci. 71, 4801–4809, 2014.

Vachaud, G., Passerat De Silans, A., Balabanis, P., and Vauclin, M.: Temporal stability of spatially measured soil water probability density function, Soil Sci. Soc. Am. J. 49, 822–828, 1985.

5 Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sens. Environ., 70, 191–207, 1999.

Walker, J., Willgoose, G., and Kalma, J.: Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: simplified kalman filter covariance forecasting and field application, Water Resour. Res., 38(12), 1301, 2002.Aman, A. A., Cman, C., and Bman, B. B.: More test articles, J. Adv. Res., 35, 13–28, doi:10.2345/67890, 2014.

10

Hydrology and
Earth System
Sciences
Discussions

Table 1. The geographical information of SCAN sites used in this study.

Site name	Site ID	State*	Climate region**	Latitude (degree)	Longitude (degree)	Soil texture*** 5 cm	10 cm	20 cm	50 cm	100 cm	Elevation (m)
Alkali Mesa	2138	UT	Mediterranean continental	37.67 N	109.37 W	L	L	L	L	CL	1966
Blue Creek	2135	UT	Mediterranean continental	41.93 N	112.43 W	SIL	SIL	SIL	SICL	SIL	1582
Cache Junction	2136	UT	Mediterranean continental	41.82 N	111.98 W	SIC	SIC	SIC	SICL	SIC	1351
Cave Valley	2161	UT	Mediterranean continental	37.35 N	113.12 W	S	S	S	S	S	1918
Centralia Lake	2094	KS	Humid continental	39.07 N	96.17 W	SIC	SIC	SIC	SIC	SICL	397
Chicken Ridge	2150	UT	Mediterranean continental	41.33 N	111.30 W	SL	SL	FSL	FSL	L	2334
Eastland	2130	UT	Mediterranean continental	37.78 N	109.17 W	L	L	L	L	L	2086
Enterprise	2128	UT	Semiarid	37.63 N	113.65 W	SCL	CL	CL	CL	FSL	1600
Ephraim	2126	UT	Mediterranean continental	39.42 N	111.57 W	CL	CL	CL	L	SIC	1678
Green River	2131	UT	Semiarid	39.02 N	110.17 W	FSL	FSL	FSL	L	FSL	1252
Grouse Creek	2160	UT	Mediterranean continental	41.78 N	113.82 W	L	L	L	CL	SCL	1782
Hals Canyon	2164	UT	Semiarid	38.60 N	113.75 W	COSL	COSL	SL	COSL	LCOS	1600
Holden	2127	UT	Semiarid	39.20 N	112.40 W	FSL	FSL	FSL	L	SL	1445
Little Red Fox	2155	UT	Semiarid	40.18 N	110.30 W	CL	CL	CL	CL	CL	1645
Manderfield	2156	UT	Mediterranean continental	38.37 N	112.65 W	SICL	SIL	SIL	L	COSL	1882
McCracken Mesa	2140	UT	Semiarid	37.45 N	109.33 W	VFSL	VFSL	VFSL	L	VFSL	1621
Milford	2129	UT	Semiarid	38.35 N	113.02 W	SIL	SICL	SICL	SIL	SIL	1523
Molly Caren #1	2014	OH	Humid continental	39.95 N	83.45 W	SIL	SIL	SIL	C	L	323
Mountain Home	2134	UT	Mediterranean continental	40.37 N	110.60 W	SCL	SCL	SCL	CL	CL	2118
Panguitch	2157	UT	Mediterranean continental	37.87 N	112.43 W	L	SCL	SCL	L	L	1998
Perdido Riv Farms	2181	AL	Humid subtropical	31.12 N	87.33 W	/	/	/	/	/	91
Price	2132	UT	Mediterranean continental	39.53 N	110.48 W	L	L	CL	SIL	L	1721
Sevilleta	2171	NM	Semiarid	34.35 N	106.68 W	/	/	/	/	/	1595
Shagbark Hills	2068	IA	Humid continental	42.43 N	95.77 W	SICL	SICL	SICL	SICL	SICL	427
Silver City	2086	MS	Humid subtropical	33.08 N	90.52 W	SIL	SIL	SIL	SICL	SICL	35
Split Mountain	2154	UT	Semiarid	40.38 N	109.35 W	CL	CL	CL	L	L	1475

Hydrology and
Earth System
Sciences
Discussions

Spooky	2158	UT	Semiarid	37.52 N	111.27 W	FS	FS	FS	FSL	VFSL	1627
Tule Valley	2163	UT	Semiarid	39.23 N	113.47 W	L	L	L	SL	SL	1397
West Summit	2139	UT	Mediterranean	38.02 N	109.13 W	L	L	L	L	L	2135
Willow Wells	2108	NM	Semiarid	33.53 N	103.63 W	FS	FS	FS	SCL	COSL	1383
Youmans Farm	2038	SC	Humid subtropical	32.67 N	81.20 W	/	/	/	/	/	23

*Abbreviations in state: AL: Alabama, IA: Iowa, KS: Kansas, MS: Mississippi, NM: New Mexico, OH: Ohio, SC: South Carolina, UT: Utah

**Köppen climate classification

***Abbreviations in soil texture: C: clay, CL: clay loam, COSL: coarse sandy loam, FS: fine sand, FSL: fine sandy loam, L: loam, LCOS: loamy coarse sand, LFS: loamy fine sand, S: sand; SCL: sandy clay loam, SIC: silty clay, SICL: silty clay loam, SIL: silt loam, VFSL: very fine sandy loam.

Table 2. Statistics of the performance of observation operators built by dataset of one resolution and their applications in datasets with other resolutions.

Statistics	Shagbark Hills			Perdido Riv Farms			Sevilleta		
	RMSE	R^2	NSC	RMSE	R^2	NSC	RMSE	R^2	NSC
Calibrated by hourly data	4.07E-3	0.995	0.995	9.30E-4	0.999	0.999	8.84E-4	0.995	0.995
Applied in daily data	4.12E-3	0.995	0.995	1.35E-3	0.999	0.998	1.26E-3	0.991	0.990
Applied in weekly data	6.03E-3	0.990	0.990	2.27E-3	0.995	0.995	1.57E-3	0.989	0.989
Calibrated by daily data	3.98E-3	0.996	0.995	9.64E-4	0.999	1.000	8.11E-4	0.996	0.996
Applied in hourly data	4.34E-3	0.995	0.995	1.32E-3	0.999	0.998	2.25E-3	0.992	0.969
Applied in weekly data	6.09E-3	0.990	0.990	2.07E-3	0.995	0.995	1.44E-3	0.989	0.989
Calibrated by weekly data	5.39E-3	0.992	0.992	2.00E-3	0.996	0.996	1.26E-3	0.990	0.990
Applied in hourly data	4.91E-3	0.994	0.994	1.41E-3	0.998	0.998	1.83E-3	0.981	0.980
Applied in daily data	5.68E-2	0.991	0.991	1.11E-3	0.999	0.999	1.16E-3	0.993	0.993

5

Hydrology and
Earth System
Sciences

Discussions

EGU
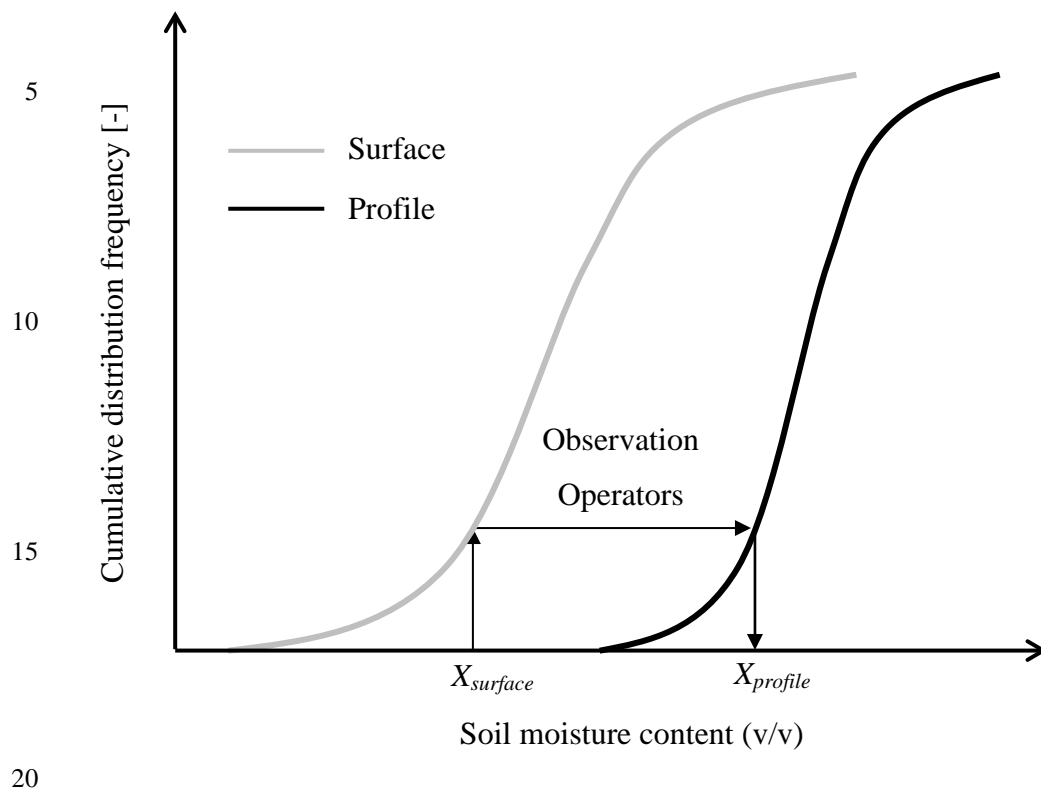
Figure 1. A graph shows how the cumulative distribution frequency (CDF) of surface soil moisture is adjusted into that of profile soil moisture by observation operators.
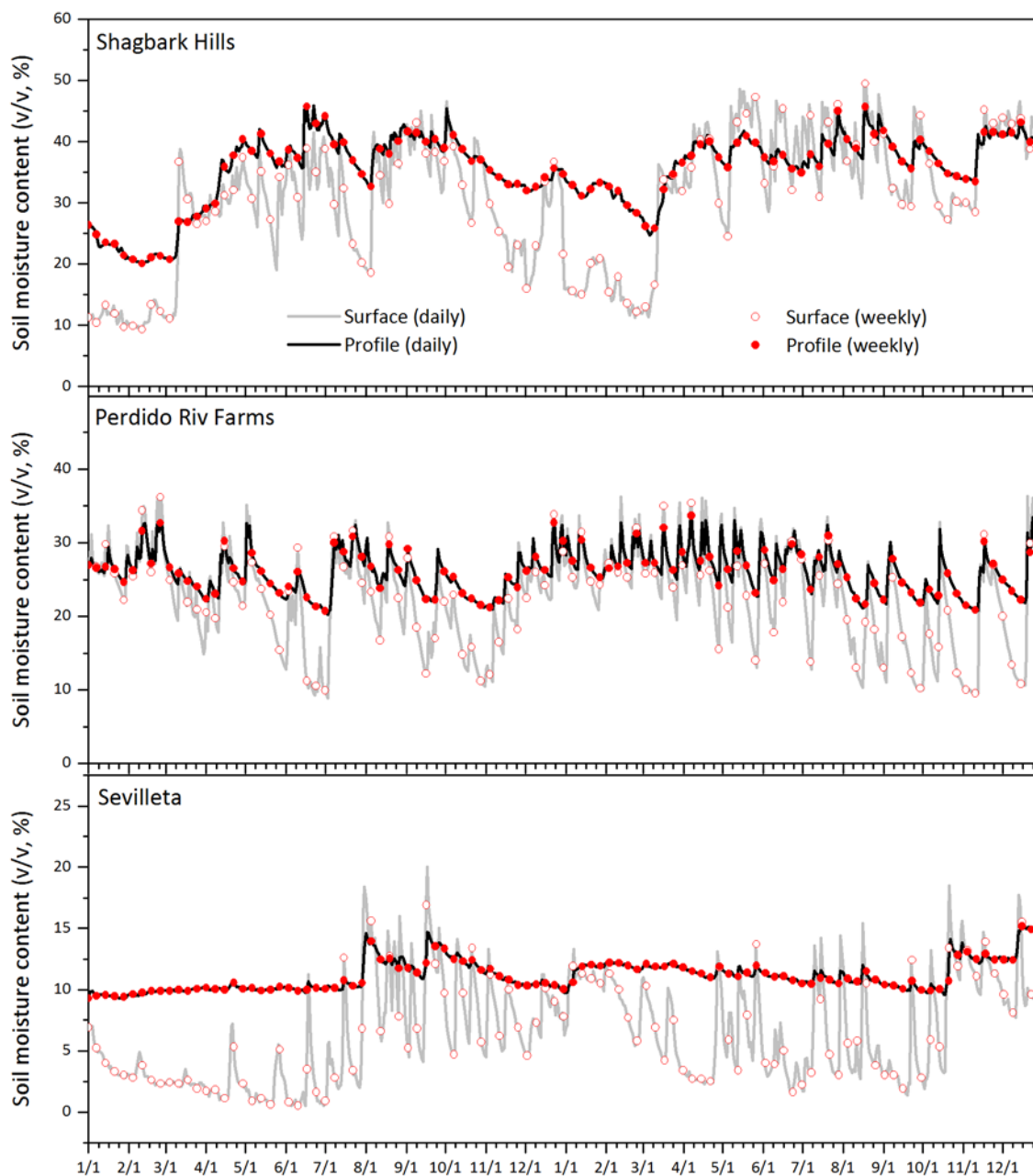
Figure 2. Surface and profile soil moisture time series at daily and weekly temporal resolutions in the sites of Shagbark Hills, Perdido Riv Farms and Sevilleta.
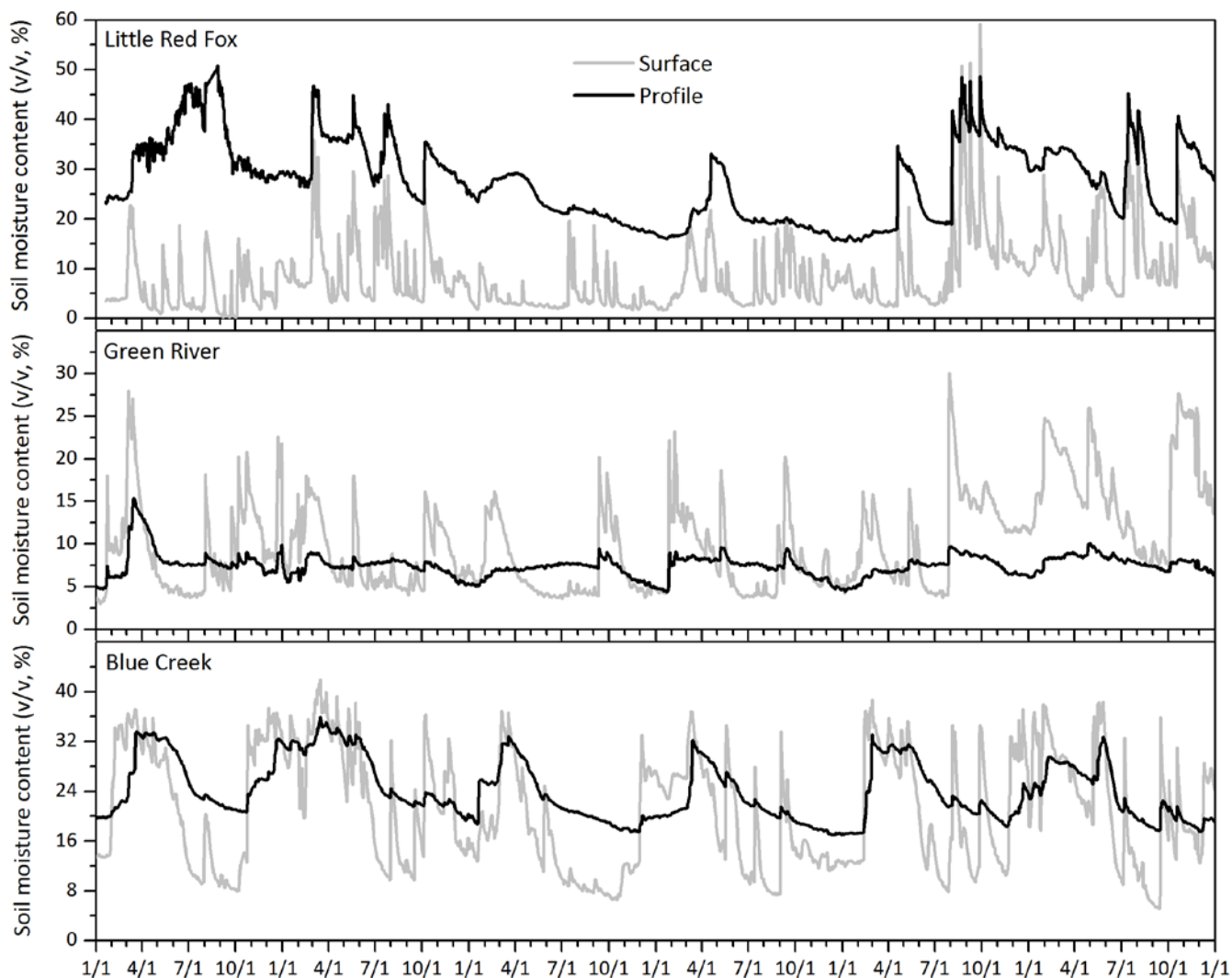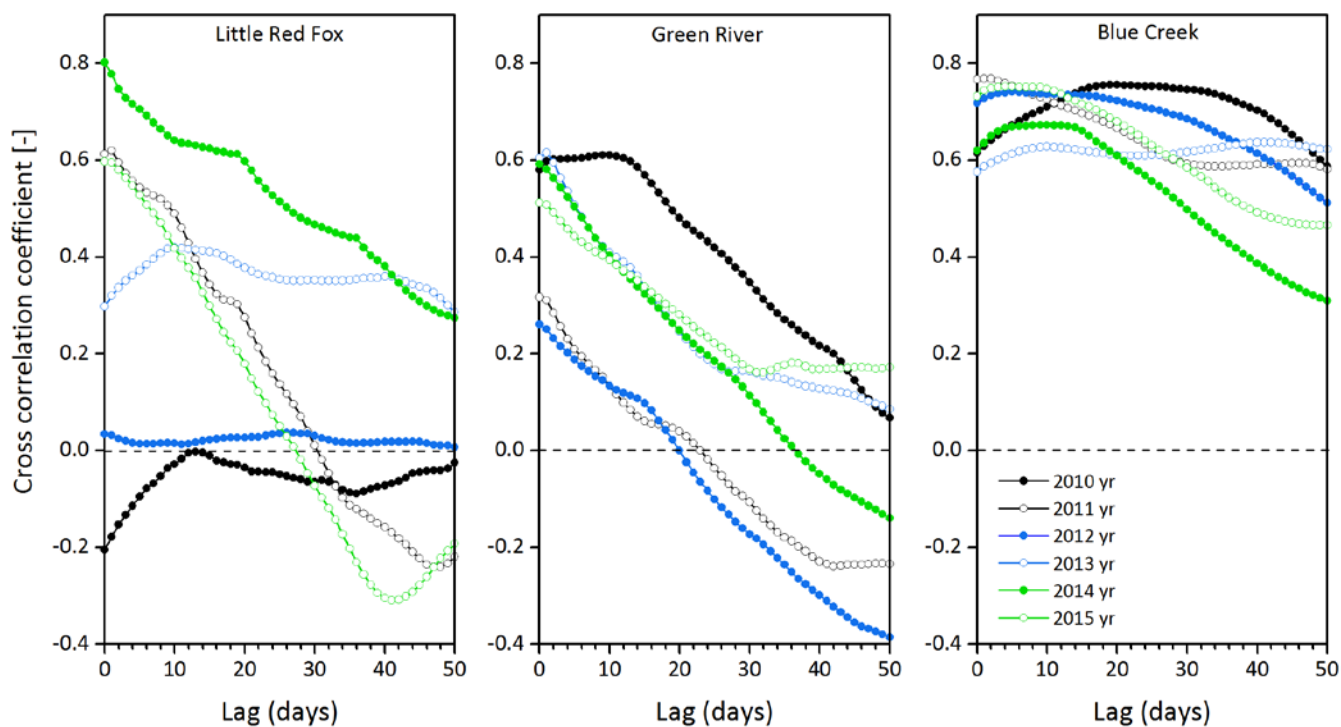
5

Hydrology and
Earth System
Sciences

Discussions

Open Access

EGU

Figure 3. Surface and profile soil moisture time series from 2010 to 2015 in the sites of Little Red Fox, Green River and
5 Blue Creek.

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

Figure 4. Cross correlation coefficient between surface and profile soil moisture in different single year for the sites of Little Red Fox, Green River and Blue Creek.
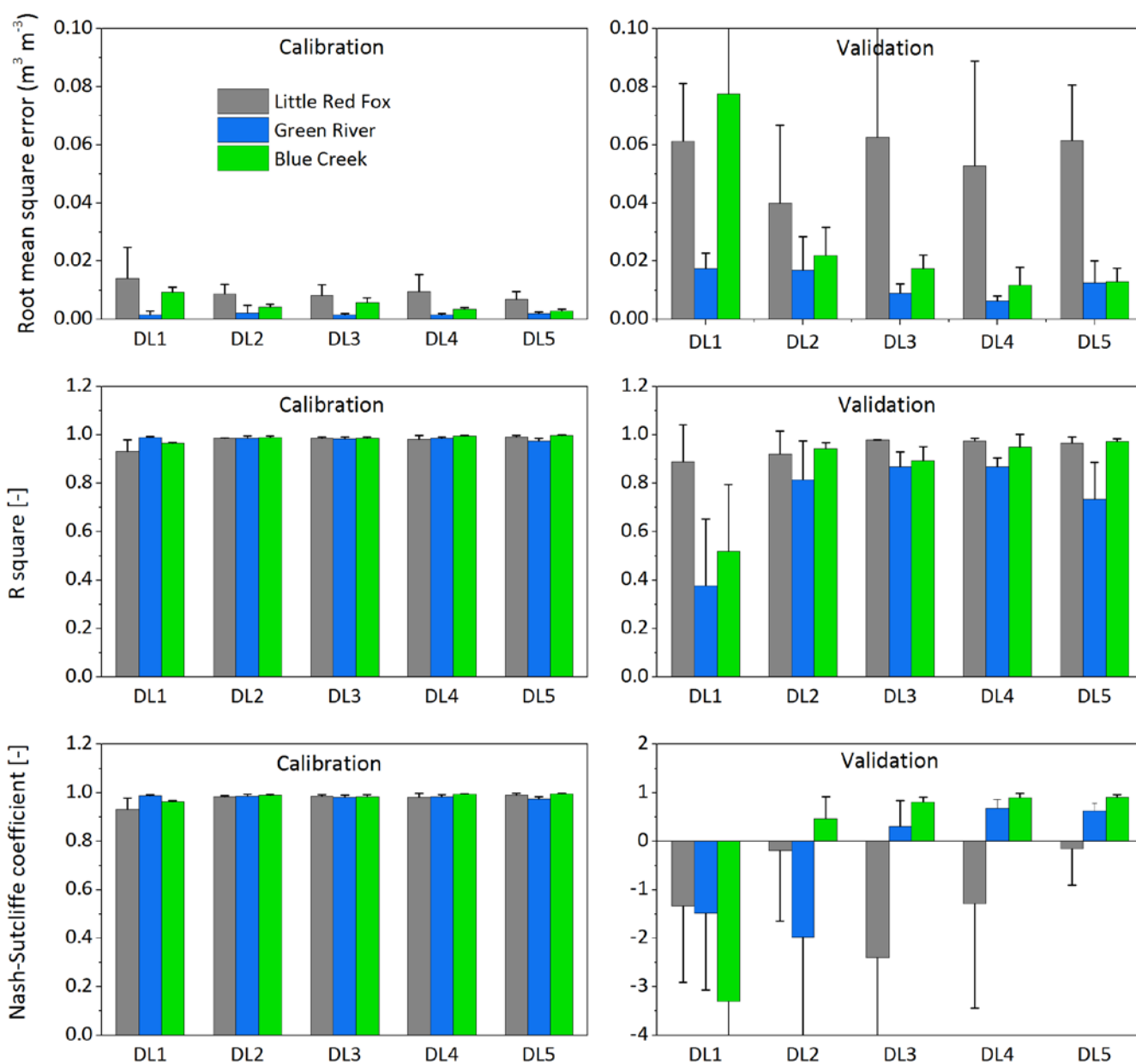
Figure 5. Statistics of the RMSE, R^2 and NSC for data length of soil moisture time series on the performance of observation operators in both calibration and validation periods. The error bar represents one standard deviation.

5
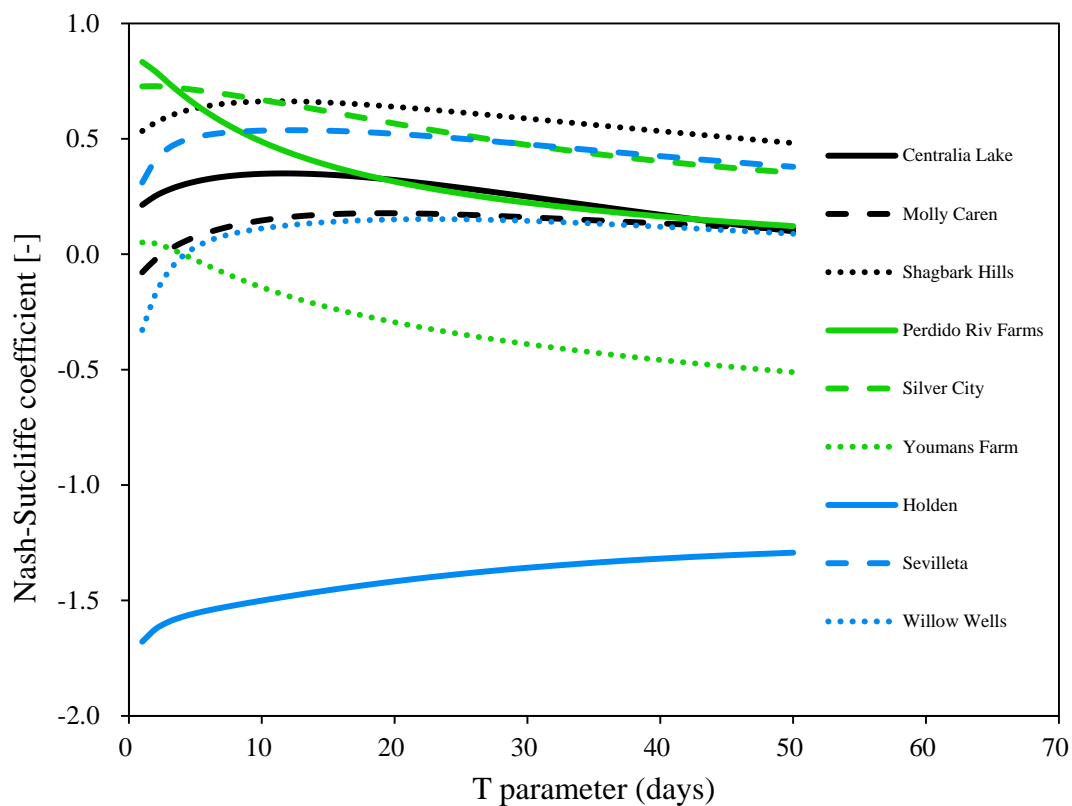
Figure 6. A graph shows how the cumulative distribution frequency (CDF) of surface soil moisture is adjusted into that of
profile soil moisture by observation operators.
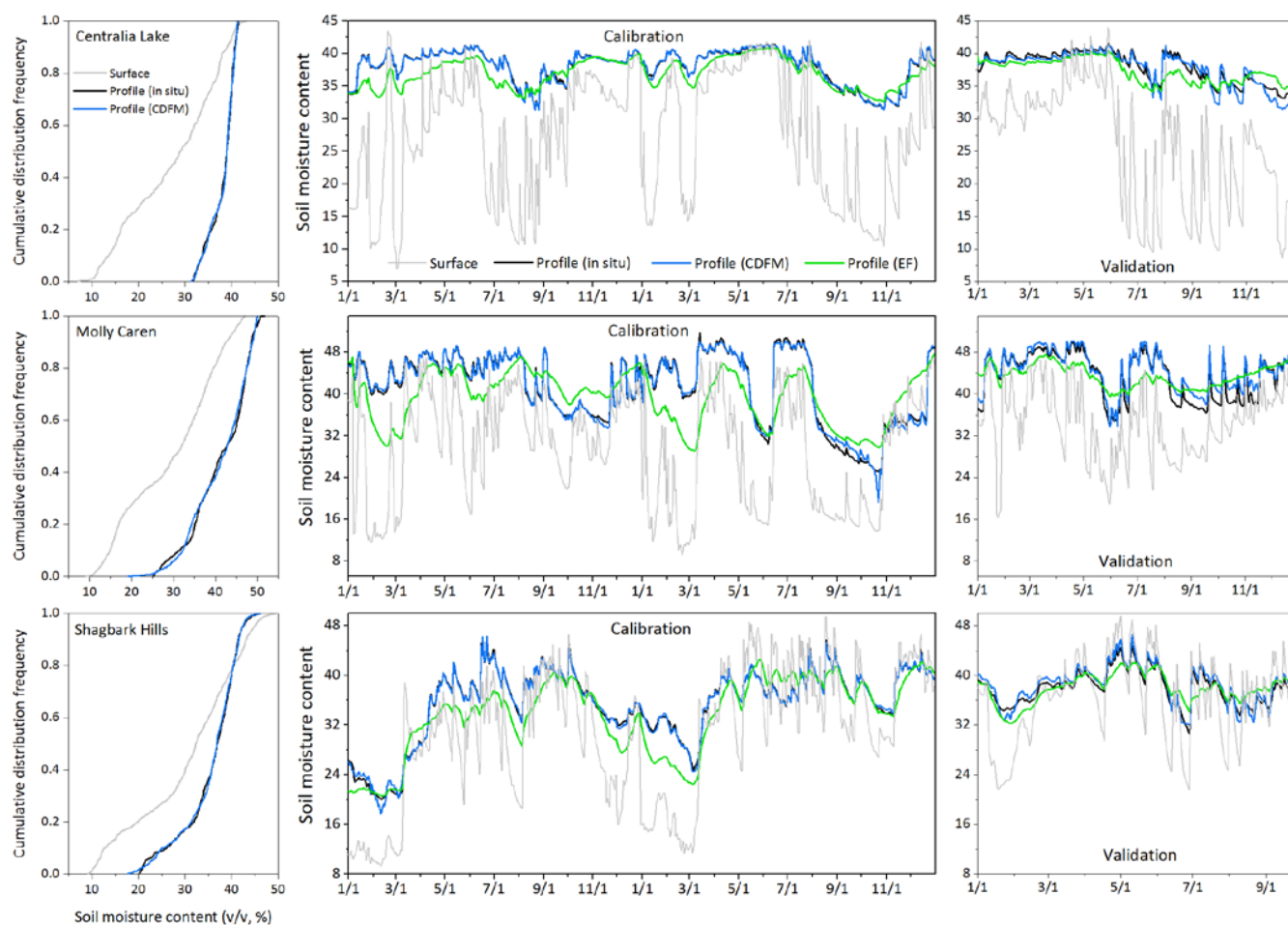
Figure 7. The predicted profile soil moisture by cumulative distribution frequency (CDF) matching (Profile (CDFM)) and
exponential filter (Profile (EF)) during calibration and validation periods for the three sites in humid continental climate. The
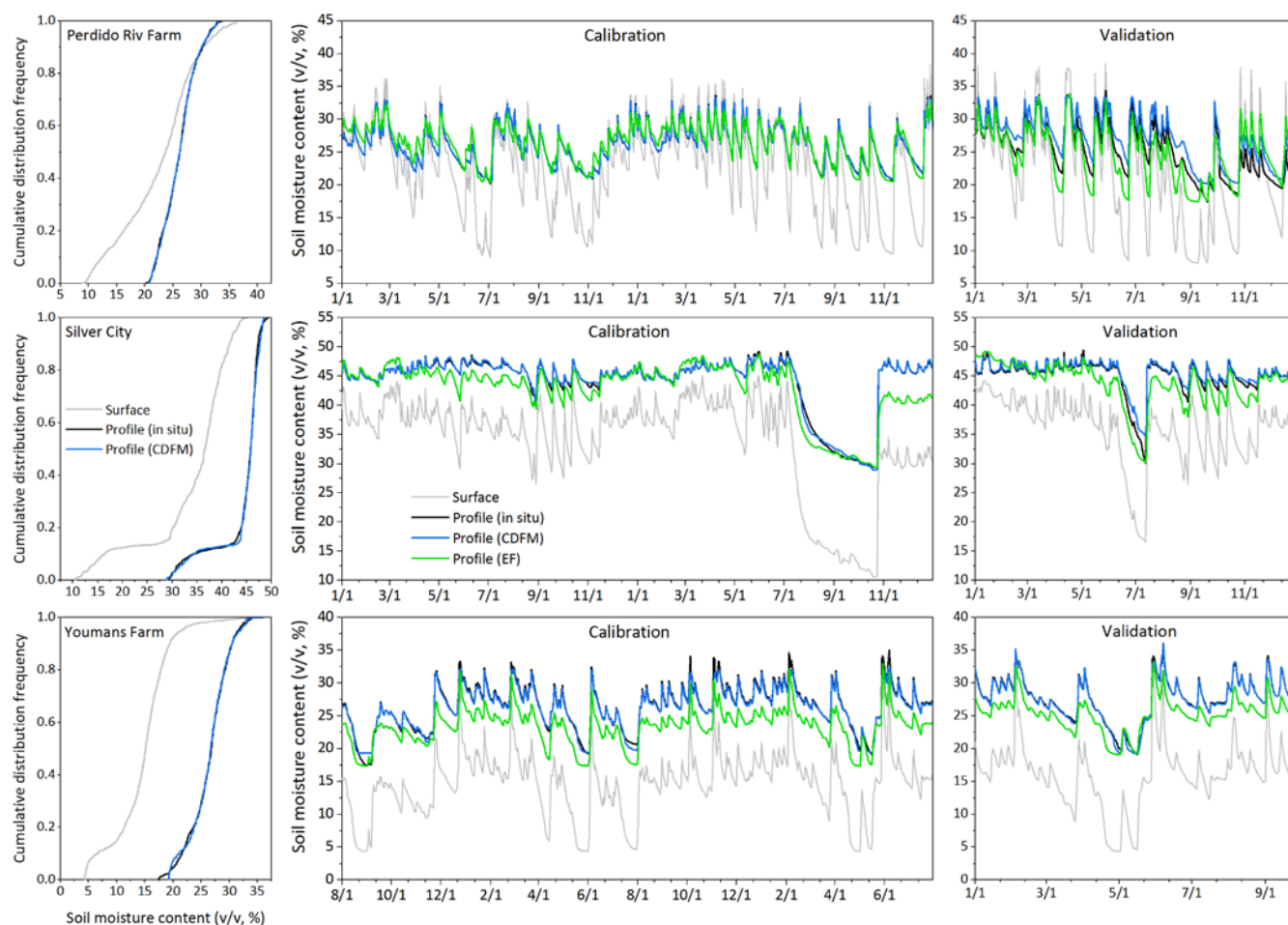three graphs in the left shows the CDFs of surface and measured and predicted profile soil moisture.

Figure 8. The predicted profile soil moisture by cumulative distribution frequency (CDF) matching (Profile (CDFM)) and exponential filter (Profile (EF)) during calibration and validation periods for the three sites in humid subtropical climate. The three graphs in the left shows the CDFs of surface and measured and predicted profile soil moisture.
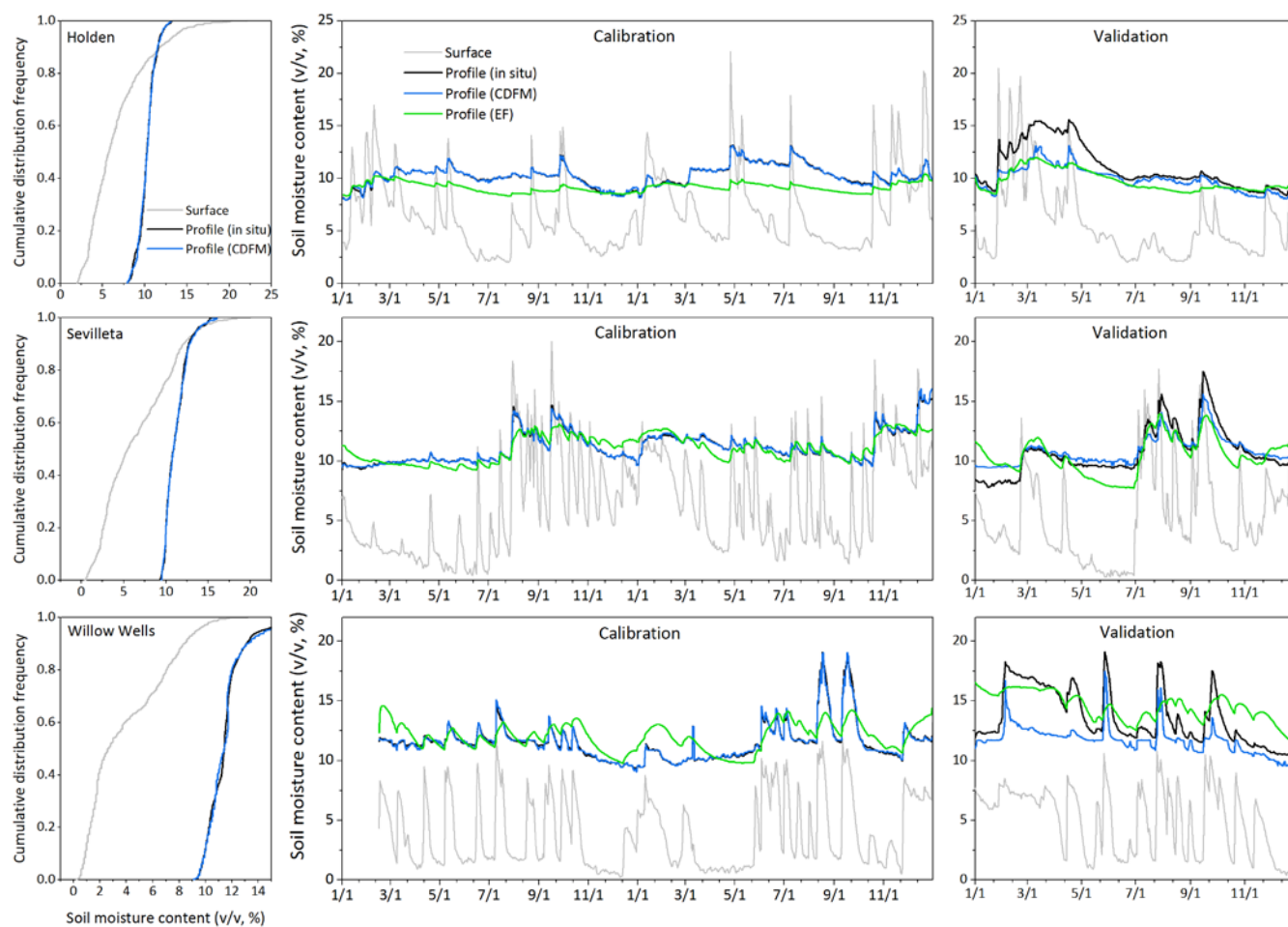
Figure 9. The predicted profile soil moisture by cumulative distribution frequency (CDF) matching (Profile (CDFM)) and exponential filter (Profile (EF)) during calibration and validation periods for the three sites in semiarid climate. The three graphs in the left shows the CDFs of surface and measured and predicted profile soil moisture.
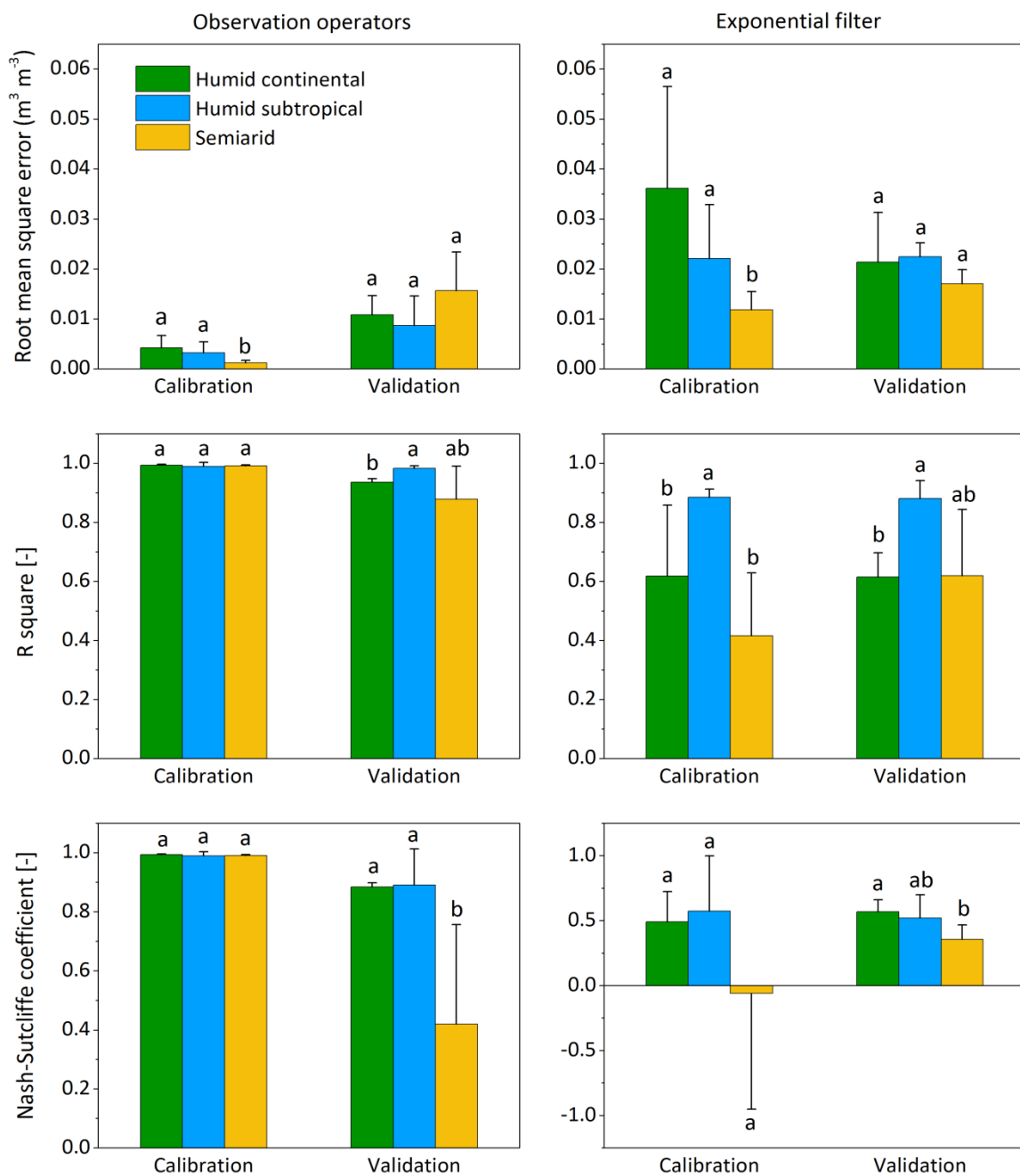
Figure 10. Statistics of the RMSE, R^2 and NSC for the performance of observation operators and exponential filter methods in the different sites in humid continental, humid subtropical and semiarid climates. Different lowercase letters above bars indicate significant ($P<0.05$) differences between climates in either calibration or validation period.