Technical Notes

Core Ideas

- Ensemble Kalman filter data assimilation was used to predict soil water content.
- Analyzed data assimilation frequencies were 1, 2, 3, 5, 7, 9, 11, and 14 d.
- Assimilation of observed data every 7 d or more yielded better results.

How Critical Is the Assimilation Frequency of Water Content Measurements for Obtaining Soil Hydraulic Parameters with Data Assimilation?

Javier Valdes-Abellan*, Yakov Pachepsky, Gonzalo Martinez, and Concepción Pla

Data assimilation (DA) is a promising alternative to infer soil hydraulic parameters from soil water dynamics data. Frequency of measurements and updates are important controls of DA efficiency; however, no strict guidance exists on determining the optimal frequency. In this study, DA was performed with the ensemble Kalman filter (EnKF) with a state augmentation approach to update both model states and parameters. We analyzed updates every 1, 2, 3, 5, 7, 9, 11, and 14 d. Two soil types (sandy loam and loam) and four climates (hot semiarid [Bwh], cold semiarid [Bsk], humid continental [Dfa], and humid subtropical [Cfa]) were considered. Results demonstrate that DA with high update frequencies does not provide better results than results obtained when using low frequencies. For sandy loam soil, assimilation of data every seven or more days yields better results for whatever climate considered. For loam soil, the same is true after 9 mo of assimilation. The chosen performance metric may affect the results, but the general trend of better results with low assimilation frequencies does not change.

Abbreviations: Bsk, cold semiarid; Bwh, hot semiarid; Cfa, humid subtropical; DA, data assimilation; Dfa, humid continental; EnKF, ensemble Kalman filter; nn, numerical nodes.

Determining soil hydraulic parameters is a fundamental step in many soil research projects or even the final objective of the project itself. Soil hydraulic parameters are required to understand and predict soil water dynamics in soil, which is a key aspect of mass and energy cycles in soils (Martinez et al., 2017). Methods to determine soil hydraulic parameters have been and still present a challenge for soil researchers.

Data assimilation is a relatively new methodology to estimate soil hydraulic parameters. It uses monitoring data as soon as they become available to correct modeling results and model parameters reflecting soil hydraulic parameters. The correction is based on the idea that both measurements and modeling results are uncertain, and comparison of the model and data uncertainties indicates how close should modeling results be brought to the measured values when the former are updated as the latter become available (Pachepsky et al., 2014). Data assimilation in soil water flow modeling has been applied extensively, and comprehensive reviews are available (Chirico et al., 2014; Medina et al., 2014a, 2014b; Zhang et al., 2017).

The EnKF is one of the most widely used among the DA methods (Evensen, 1994,2009). It has been widely applied in hydrology (Reichle, 2008; Reichle and Koster, 2003; Reichle et al., 2002) and more specifically in soil water flow modeling based on the Richards equation (Camporese et al., 2009; Das and Mohanty, 2006; Huang et al., 2008; Vereecken et al., 2008; Vrugt et al., 2005b). An important development in EnKF applications led to the possibility of correcting both states—water contents and soil water potential—and soil hydraulic parameters simultaneously (Li and Ren, 2011; Medina et al., 2014b; Montzka et al., 2011; Moradkhani et al., 2005; Song et al., 2014; Vrugt et al., 2005a, 2005b).

Frequency of measurements and corresponding state and parameter updates is one of the controls of the DA efficiency. The need for the consideration of the impacts of data

J. Valdes-Abellan, Dep. of Civil Engineering, Univ. of Alicante, Alicante, Spain; Y. Pachepsky, USDA–ARS, Environmental Microbial and Food Safety Lab., Beltsville, MD 20705; G. Martinez, Dep. of Applied Physics, Univ. of Córdoba, 14071 Córdoba, Spain; C. Pla, Dep. of Civil Engineering, Univ. of Alicante, Alicante, Spain. *Corresponding author (javier.valdes@ua.es).

Received 23 July 2018.
Accepted 5 Jan. 2019.
Supplemental material online.

Citation: Valdes-Abellan, J., Y. Pachepsky, G. Martinez, and C. Pla. 2019. How critical is the assimilation frequency of water content measurements for obtaining soil hydraulic parameters with data assimilation? Vadose Zone J. 18:180142. doi:10.2136/vzj2018.07.0142

frequency, duration, and coverage on DA results was noted, for instance, when observations were combined with biogeochemical modeling (Kaufman et al., 2018), when the data were assimilated for real-time flood forecasts (Mazzoleni et al., 2017), and when the land DA system was used to couple the microwave remote sensing and land surface model and then improve the accuracy of land surface fluxes and status estimation (Lu et al., 2016). Results of using geophysical and remote sensing DA to estimate soil water contents depended on the assimilation frequency (Cosenza, 2016; Rosolem et al., 2014).

Effects of the soil sensor DA frequency on the modeling of soil water contents in soil profile have been noted previously (Chu et al., 2015; Wu and Margulis, 2013). De Lannoy et al. (2007) studied the effect of the assimilation frequency with soil moisture data from different depths on model states; they used roughly 2 yr of observations in sandy soils in Maryland and came to the conclusion that the optimum update frequency was about 1 to 2 wk. To our knowledge, no research has been done on the effect of the update frequency on the efficiency of soil moisture DA in the determination of both model states and model parameters from sets on sensors placed at different depths.

The objective of this study was to explore the effect of different DA frequencies on soil hydraulic parameters estimation with DA from sensors placed at several depths under four different climates and two different soil textures.

Materials and Methods

The HYDRUS-1D software (Šimůnek et al., 2009) was the engine used to solve the variably saturated soil water flow model according to the Richards equation (Richards, 1931):

$$\frac{\partial \theta}{\partial t}=\frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right]-S(h) \qquad [1]$$

where θ is volumetric water content [$L^3\ L^{-3}$], t is time [T], z is the vertical coordinate [L], $K(\theta)$ is the soil unsaturated hydraulic conductivity [$L\ T^{-1}$], h is the soil water pressure head [L], and $S(h)$ is the sink term that represents water uptake by plants [$L^3\ L^{-3}\ T^{-1}$]. The van Genuchten–Mualem constitutive relationships (Mualem, 1976; van Genuchten, 1980) were adopted to define the functional relation between the soil hydraulic properties:

$$S_e=\frac{\theta(h)-\theta_r}{\theta_s-\theta_r}=\begin{cases}\left(1+|\alpha h|^n\right)^{-(1-1/n)} & h<0\\ 1 \ \leftrightarrow\ \theta(h)=\theta_s & h\geq 0\end{cases} \qquad [2]$$

$$K(h)=K_s S_e^l\left\{1-\left[1-S_e^{n/(n-1)}\right]^{1-1/n}\right\}^2 \qquad [3]$$

where S_e is the effective saturation, θ_s is the saturated water content [$L^3\ L^{-3}$], θ_r is the residual water content [$L^3\ L^{-3}$], K_s is the saturated hydraulic conductivity [$L\ T^{-1}$], and α [L^{-1}], n (dimensionless), and l (dimensionless) are empirical coefficients that determine the shape of the hydraulic functions. The present study did not optimize the l parameter, and its value was set constant at 0.5.

Observation Data

Synthetic soil water content datasets were obtained and presented earlier in the work of Valdes-Abellan et al. (2018a). A single-layer soil profile of 110 cm (arbitrarily chosen) was discretized into 111 numerical nodes (nn = 111). The top and bottom boundary conditions were chosen as a variable atmospheric condition and free drainage, respectively. Four different climates were selected from the Köppen classification to generate the top boundary conditions: Bwh, Bsk, Dfa, and Cfa climates. The CLIGEN weather simulator (Nicks et al., 1995) was used to generate time series of daily rainfall, maximum and minimum temperature, and solar radiation. Input data for these climates were taken from Tucson, AZ (32.25 N, 110.83 W, 771 m asl), Moscow, ID (46.73 N, 117.00 W, 801 m asl), College Station, TX (30.58 N, 93.35 W, 94 m asl), and Indianapolis, IN (39.73 N, 86.27 W, 240 m asl), respectively. Evapotranspiration was computed according to the modified Hargreaves model (Martínez et al., 2014; Williams et al., 2008).

Two different soil textures were considered: coarse (sandy loam) and medium texture (loam). Soil water retention parameters (θ_r, θ_s, α, n) were obtained from Wösten et al. (1999) while K_s values were obtained from the RAWLS database (Schaap and Leij, 1998). These parameters are listed in Table 1. They constitute the "correct" values.

A 4-yr period was simulated to obtain the synthetic data. The first year was used as a spin-up period to obtain a transient soil profile less dependent on the initial condition, which was set at field capacity as defined in Twarakavi et al. (2009). Three observation depths were considered (0.15, 0.55, and 0.75 m) that corresponded to often found depths of major soil horizons A and B.

The modeled soil water content at those depths were later perturbed by using a conditional multivariate normal distribution (Or and Hanks, 1992). For each depth, 20 replications were created to determine the variance and covariance in the observation. This step requires a covariance matrix which was adopted from Jacques (2000) and Pachepsky et al. (2005). Consideration of a complete covariance matrix is a crucial factor when assimilating data using

Table 1. Soil hydraulic parameters† used to generate the synthetic data and used as the starting point in the searching process.

Soil texture	Purpose of use‡	θ_r	θ_s	α	n	K_s
				cm^{-1}		$cm\ d^{-1}$
Sandy loam	correct	0.0250	0.4030	0.0383	1.3774	150
	initial	0.0426	0.3846	0.0349	1.4271	45.67
Loam	correct	0.0100	0.4390	0.0314	1.1804	16
	initial	0.0627	0.4063	0.0097	1.4966	9.94

† θ_s, saturated volumetric water contents; θ_r, residual volumetric water contents; α and n, empirical shape parameters; K_s, saturated hydraulic conductivity.

‡ *Correct* indicates values used to generate synthetic data; *initial* indicates initial soil hydraulic parameters used in the search process (from Rosetta).

the EnKF method because, roughly speaking, the ability of the method to correct both model parameter and states is based on the relative difference between the variance in the ensemble of simulations and the variance in the ensemble of observations. The initial soil moisture profile used during the updating period was obtained considering a linear dependence of pressure head between the values inferred for the three observation depths.

Data Assimilation with the Ensemble Kalman Filter

A detailed description of the EnKF lies beyond the scope of this contribution; we address to Evensen (2009), Burgers et al. (1998), Franssen and Kinzelbach (2009), or Zhou et al. (2011) among others, to get further information. In brief, the EnKF method with a state augmentation approach or AEnKF (Crosman and Horel, 2010; Chen and Zhang, 2006; Chen et al., 2015; Monsivais-Huertero et al., 2010) was selected. At each update time, the nn values of soil water content of the N units were collected in matrix $\mathbf{X}$ (nn,N); the p parameters of the N units were collected in matrix $\mathbf{Y}$ (p,N) and both matrixes were combined in a single augmented state matrix $\mathbf{Z}$ (p+nn,N). The gain Kalman matrix, $\mathbf{K_t}$ (p+nn,m), which relates the variability in the model ensemble and in the data, is obtained as

$$\mathbf{K}_t = \mathbf{C}_{\mathrm{YX\text{-}XX}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{C}_{\mathrm{XX}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \quad [4]$$

where the covariance matrix of the ensemble is $\mathbf{C}_{\mathrm{XX}}$ (nn,nn) and the covariance matrix of data is $\mathbf{R}$ (m,m). To link the state parameters (soil hydraulic parameters) with the model states (soil water content), the matrix $\mathbf{C}_{\mathrm{YX-XX}}$ (p+nn,nn) is computed at each update time. Matrix $\mathbf{R}$ (m,m) contains covariance for experimental data and it was obtained as described in Pan et al. (2012) to correct biases in the measurement in different locations at the same depth and to prevent inflated Type I errors (Quinn and Keough, 2002; Wigley et al., 1984).

The updating is achieved by applying Eq. [5], and a new augmented state matrix, $\mathbf{Z_t}^{+}$, is obtained from the prior one, $\mathbf{Z_t}^{-}$:

$$\mathbf{Z}_t^{+} = \mathbf{Z}_t^{-} + \mathbf{K}_t\left(\mathbf{D}_t - \mathbf{H}\mathbf{X}_t\right) \quad [5]$$

where the matrix $\mathbf{D_t}$ (m,N) contains the observed values plus a white noise based on the data variances and covariances. The matrix $\mathbf{H}$ (m,nn) relates observations and simulated states; it is the identity matrix when model states and observations are the same and all model states are updated. The latter filter is applied to the updated soil hydraulic parameters values (not to the states) assigning realistic values if the update leads to physically unjustifiable values of these parameters, in a way that has received the name of constrained EnKF in previous studies (Wang et al., 2009). To define those realistic boundary values, a study using the Rosetta pedotransfer function was developed: all potential combinations from percentages of sand, silt, and clay were obtained and the distribution functions of the inferred parameters from those combinations were studied. Then maximum and minimum values of soil hydraulic parameters were set as boundary values at 95% of the probability distributions across all textural compositions of soils.

The DA was applied as in Valdes-Abellan et al. (2018a). The practically same algorithm and software can be found in Valdes-Abellan et al. (2018b). In brief, an ensemble of 100 models (N = 100) was used to update five parameters (p = 5), that is, θ_r, θ_s, α, n, and K_s and the soil water content in the complete profile. Model parameters were the unique force perturbed in this exercise. Synthetic measurements from the three observation depths (m = 3) were used as the input to update the model predictions. All models of the ensemble were moved forward independently from the beginning to the first update time and then from one to the following update time. At that time, both soil water content and parameters were updated in all ensemble units and a new step forward is performed until the following update time. The update was accomplished with a state augmentation approach as mentioned above, updating states and parameters in one single step. Eight different updating frequencies—every 1, 2, 3, 5, 7, 9, 11, and 14 d—were considered to evaluate the performance of the DA procedure. Frequencies higher than once a day were not evaluated because (i) we consider that it implies expensive monitoring equipment that not all research teams can afford, (ii) we consider that intra daily variation of soil moisture is not high enough to show an impact in the DA procedure, and (iii) the higher the frequency, the longer the computational time required to run the code. So with such high frequencies, the attractiveness of the DA methodology declines rapidly.

The Rosetta software (Schaap et al., 2001) provided the initial values of the soil parameters based only on the textural class (Table 1). This methodology leads to different distances between the correct and the initial parameters for each different soil. N sets of soil parameters, one for each model of the ensemble, were required; a multivariate normal distribution (centered in the values provided by Rosetta) was used and random values were obtained for all parameters. The covariance matrix for the soil hydraulic parameters was obtained from Faulkner et al. (2003) and is provided in the Supplemental Materials.

Statistical Analysis

Soil hydraulic parameters obtained at each update time were used to evaluate the accuracy of the model for the entire period from the beginning of assimilations (excluding the spin-up period) to the update time. The performance of the model was evaluated using the RMSE between the average observed and average simulated soil water content.

For all soil–climate combinations, results were classified into two groups: high frequencies (i.e., update every 1, 2, 3, and 5 d) and low frequencies (i.e., update every 7, 9, 11, and 14 d). Statistical analysis was applied to detect statistically significant (p = 0.05) differences between the aforementioned groups. This comparison was done collecting results from 3-mo periods, so the entire DA period of 3 yr was split into 12 periods for comparison purposes. The best group was defined as the one with the lowest average RMSE along each 3-mo period when focusing on this statistic; and the group with the smallest difference between the average predicted and the correct values or the soil hydraulic parameters.

Results and Discussion

Time series of estimated soil hydraulic parameters are shown for sandy loam soil and loamy soil in Fig. 1 and 2, respectively. All the update frequencies considered, from once a day to every 14 d, provided very similar results for the sandy loam soil (Fig. 1). Only results from Bwh climate showed substantial differences among the different assimilation frequencies. Within-groups variability was relatively high. Therefore, statistically significant differences in average values between the groups of high and low frequencies in α, n, and K_s parameters were rare. This variability in data might be a consequence of the inherent characteristics of Bwh climate, with long dry periods and scarce and high-intensity rainfalls. These features produce eventual sharp increments in soil water content, which provoke significant shifts in the updated soil hydraulic parameters. Overall for Bwh climate, none of the considered update frequencies provided good results; important differences were found among the predicted parameters with different frequencies and also between the predicted and correct soil hydraulic parameter values.

In general, results of the statistical analysis confirmed the existence of relatively consistent significant differences between estimated parameters from the two groups. For example, the parameter θ_r was always better captured by the high-frequency group (frequencies shorter than a week). With sandy loam soil, parameters θ_r and n were overestimated for practically all frequencies and climates.

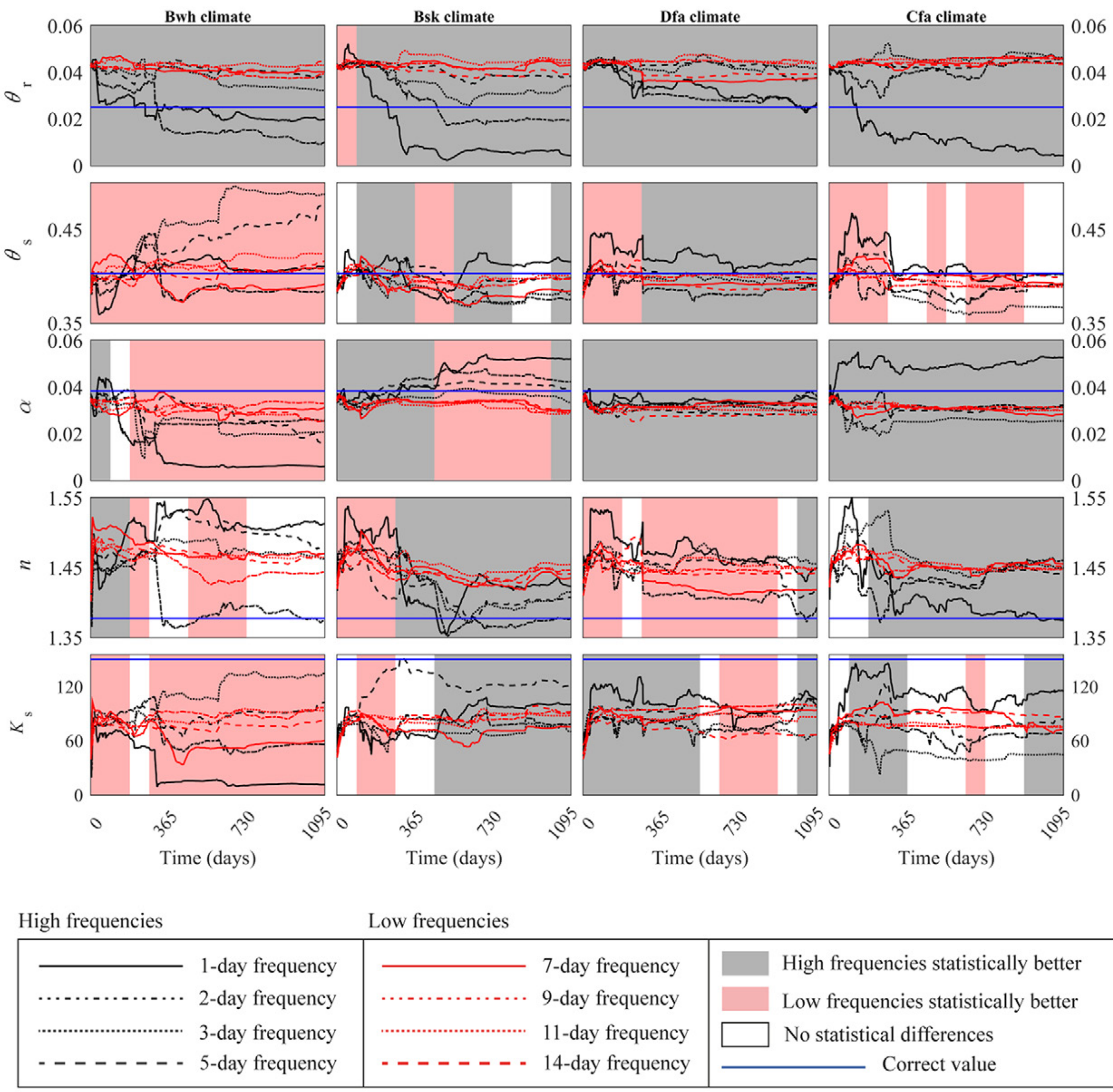

Fig. 1. Temporal evolution of the soil hydraulic parameters of sandy loam soil obtained with data assimilation modeling using different updating frequencies. Statistical analysis of the results is depicted as background shadows that highlight the existence of significant differences between high (shorter than a week) and low frequencies (equal to or longer than a week). Four different climates are evaluated. Bwh, hot semiarid; Bsk, cold semiarid; Dfa, humid continental; Cfa, humid subtropical.

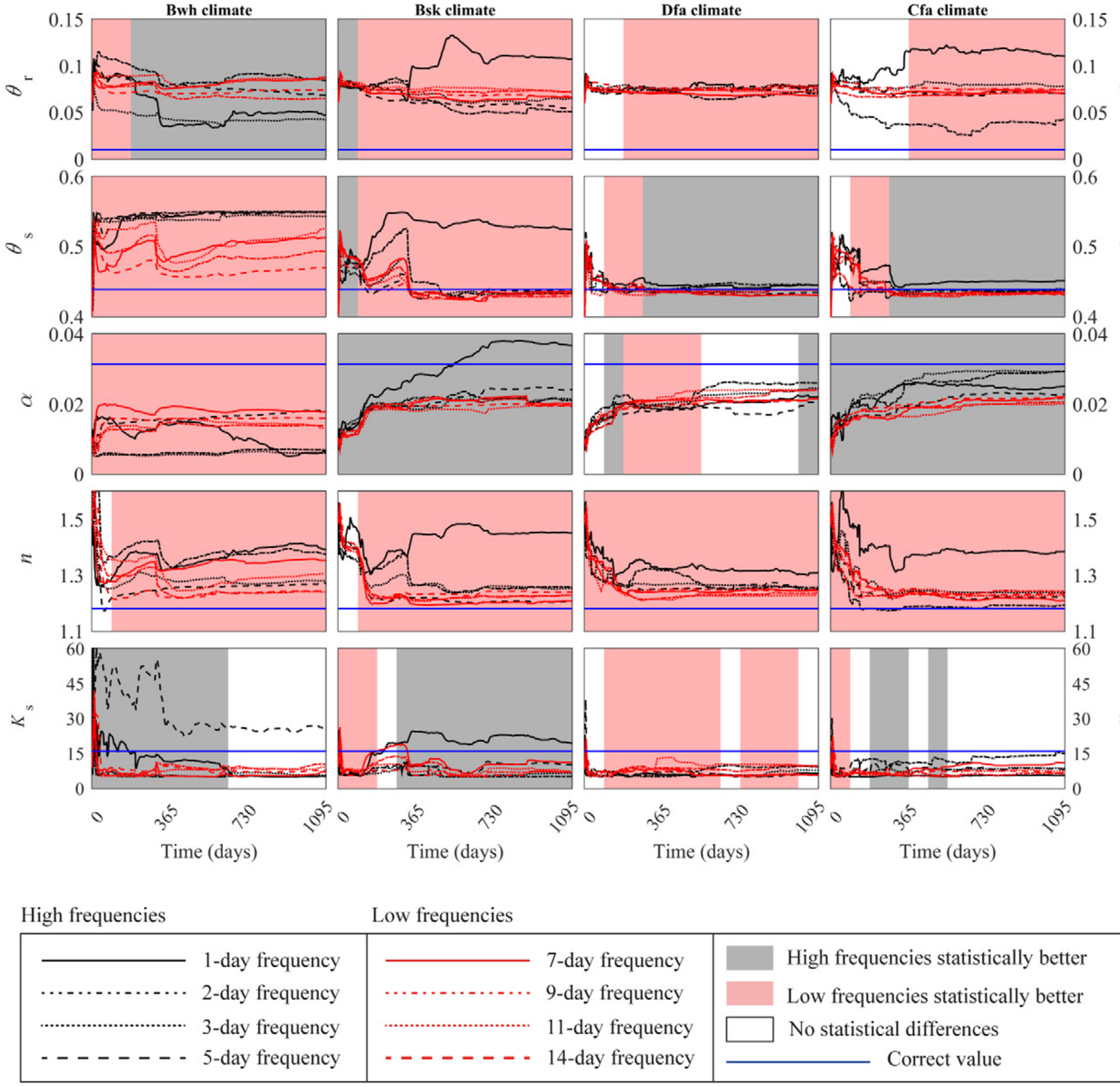

Fig. 2. Temporal evolution of the soil hydraulic parameters of loamy soil obtained with data assimilation modeling using different updating frequencies. Statistical analysis of the results is depicted as background shadows that highlight the existence of significant differences between high (shorter than a week) and low frequencies (equal to or longer than a week). Four different climates (Bwh, Bsk, Dfa and Cfa) are evaluated. Bwh, hot semiarid; Bsk, cold semiarid; Dfa, humid continental; Cfa, humid subtropical.

With respect to θ_s, the DA frequencies showed different behavior depending on climate: low frequencies seem to provide better results in Bwh and Cfa climates, while the opposite trend (advantage of low frequencies) is shown in Dfa climate and no clear conclusion can be obtained from Bsk climate. On a general trend, θ_s, α, and n were well reproduced for all frequencies in Bsk, Dfa, and Cfa climates. Parameter α was better captured with the high-frequencies updates for humid climates, while the parameter n exhibited a different behavior: it was better estimated by the high-frequencies group in the Cfa climate and by the low-frequencies group in the Dfa climate. Nevertheless, since all frequencies provided good results, significant differences moved eventually from one group to another. Finally, K_s was always underestimated for all climates and it was the parameter with the highest differences among the different assimilation frequencies. Consequently, significant differences did not lead to conclude that the high or low frequencies were better, except for the Bwh climate, where the low-frequencies group was consistently better than the high-frequencies group.

For the sandy soil (Fig. 1), general conclusions about the best frequency strategy cannot be drawn from the temporal evolution of soil hydraulic parameters when looking at all soil parameters simultaneously. We could only suggest that high frequencies are better, which may be explained by the fact that response of soil moisture conditions to changes in meteorological conditions (both in evapotranspiration and precipitation) is quicker in the case of sandy soils because of the higher permeability and the use of low frequencies may not be able to assimilate those quick changes into the model.

Larger differences between the two groups of assimilation frequencies were found with the loamy soil (Fig. 2) and the capability of high frequencies to better detect the soil water parameters is not so predominant as it was with the sandy soil. The soil water regime in loamy soils, with slower and smaller variations (compared with sandy soils), could explain the different performance of DA. These differences in DA, when applied to different soils, were also observed in a previous study (Valdes-Abellan et al., 2018b), which used a constant update frequency of $(7\ d)^{-1}$. The value of θ_r was overestimated with all frequencies and climates and the low-frequencies group was significantly better ($p = 0.05$) for the Bsk, Dfa, and Cfa climates. Such clear overestimation of θ_r was not detected with the sandy loam soil; this performance of DA could be explained by the fact that θ_r does not play a key role in soil water modeling in this work, basically because of the large difference between θ_r and the average soil water content observed in the field. Overall, in loam soil, water content used in the updates was larger and therefore less sensitive to θ_r vs. sandy loam soil. Differences between the correct or estimated θ_r and the observed and simulated soil water content were higher in the case of sandy loam soil than in the case of loam soil (e.g., the mean soil water content measured in soil was 0.29 for loam soil when the correct θ_r was 0.01). Therefore, the impact of this parameter in simulations is smaller with loamy soil and then the ability to the engine to find out the correct value is lower with this soil.

Loamy soil (Fig. 2) exhibits major differences between results of the different DA frequencies. In general, low update frequencies (i.e.: assimilation of data every 7 d or longer) lead to better results for θ_s and α in Bwh climate and to worse results in Cfa climate. In Bsk climate, statistical differences detected between the two groups of DA frequencies point out controversial results because θ_s was better captured with low frequencies and α with high frequencies. Parameter n was overestimated in all climates and the opposite occurs with α, which was underestimated in the majority of cases. Overall, the best results were accomplished with the low frequencies. Finally, K_s was also underestimated except in a very few cases of high frequencies, although variations were found regarding the best fit of results: for Bwh and Bsk climates, high DA frequencies showed better results, while high DA frequencies were preferable for Dfa climate; in Cfa climate differences between the two groups moved from one to another with no clear trend.

The temporal evolution of the RMSE for all soil–climate combinations and all the considered DA frequencies are shown in Fig. 3. This figure shows the combined ability of all soil parameters estimated by the DA to predict the soil water content. The RMSE was very low in all climates with the sandy loam soil and below 0.05 in all combinations with the loam soil. In contrast to the soil hydraulic parameters evolution, where no very clear conclusion was found whether the low- or high-frequency group is better, the analysis of the RMSE temporal evolution showed a clear better performance, on average, for the low-frequencies group, being that difference statistically significant for almost all periods, climates, and soils. Focusing on the sandy loam soil, it is worth noting that the high-frequency group never reported better results; at best, the differences between the two groups were not significant. Nevertheless, it is fair to say that the absolute values of RMSE were acceptable ($<0.01\ m^3\ m^{-3}$) in all cases (both for the high and low DA frequencies) and similar to typical soil moisture measurement errors (Campbell Scientific, 2012; Decagon Devices, 2010)

Focusing on RMSE results for the loam soil, the high-frequencies group reported better performance only for the initial stages in Bsk climate, when during the first 3 mo of data assimilation provided better statistics for the high-frequencies group. After that initial period in that particular climate, the RMSE values indicated that updating with low-frequencies yields statistically better simulations. It is worth mentioning that the aforementioned initial period with high frequencies group statistically better than low frequencies or with no statistically significant differences was not enough to get good absolute values of RSME; the authors considered that 1 yr is the optimal minimum period to get steady and good statistics. Therefore, the better performance of high frequencies took place only when it was not really useful. Additionally, in some study cases, the high frequencies not only showed poorer results but also showed an erratic behavior: for example, the RMSE plot for the 1-d updating case led to worse statistics (i.e., higher RMSE) in all climates instead of approaching better simulations. Such erratic behavior was much more common in the high-frequencies group as it can be seen in the Bwh climate with sandy loam soil and in almost all climates with loam soil.

Analysis of RMSE showed that using low DA frequencies (equal or longer than a week) was better than using high DA frequencies (shorter than a week). This occurred for both types of studied soils and all the different climates. The RMSE determines the ability of the model to infer soil parameters and thus it indicates the reliability of the different studied DA updating frequencies. Although sandy loam soil presented better results from the point of view of predicting some soil parameters when using high DA updating frequencies (Fig. 1), the RMSE values clearly demonstrated that the use of low DA frequencies for soil modeling is not only acceptable but also more efficient than using the high DA frequencies alternative to estimate soil water content. One possible reason could be that when updating observations with high frequencies, the inclusion of the potential noise jointly with the observations is also included in the assimilation, which may be detrimental to the efficiency of the EnKF, exacerbated by small-size ensembles and low replications in the observations. The detrimental effects of noise for the DA with the EnKF have been noted, and various techniques were suggested to alleviate it (Ha et al., 2017; Lei et al., 2012). If the temporal variations in soil water contents and the high-frequency updates effectively creates the same effect as the additional uncertainty in measurements, then the DA will be less inclined to bring parameters and model accuracy to true values. This possibility presents an interesting avenue to explore. Properties of soils as atmospheric signal filters may also play a role in the update frequency efficiency. The optimum update frequency in the work of de Lannoy et al.

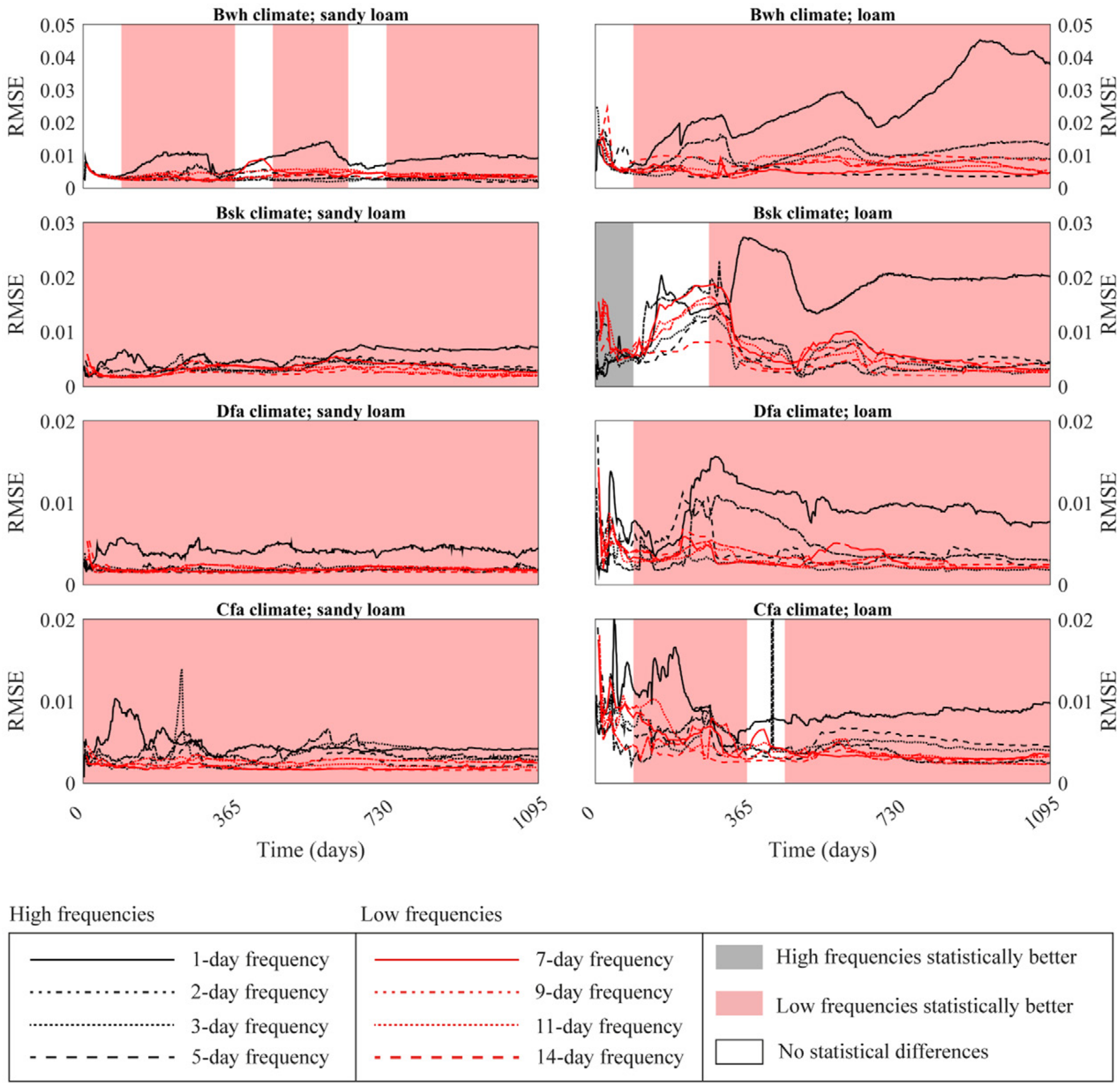

Fig. 3. The RMSE temporal evolution for both the sandy loam and the loamy soil under the influence of four different climates. Statistical analysis of the results is depicted as background shadows that highlight the existence of significant differences between high (shorter than a week) and low frequencies (equal to or longer than a week). Bwh, hot semiarid; Bsk, cold semiarid; Dfa, humid continental; Cfa, humid subtropical.

(2007), who used DA to improve modeling states without changing parameters, related their optimum frequency to the autocorrelation length for soil moisture and, ultimately, to the atmospheric forcing.

We note that the model performance metric may be an additional factor affecting the selection of the assimilation frequency. Different metrics, such as bias, Nash–Sutcliffe efficiency index, R^2, multi-objective optimization Pareto-type metrics, etc., could produce different results in terms of the DA update frequency. Therefore, the intended model application can be the update control. To illustrate the case, temporal evolution of r^2, instead of RMSE, was provided in Supplemental Fig. S3; it can be observed that r^2 did not show the same results of periods with or without statistically significant differences between the two groups that can be inferred from the analysis of RMSE. Nevertheless, from this figure, low assimilation frequencies are also better than high frequencies.

There is an apparent contradiction between evaluations of the update frequency by RMSE and by results of soil hydraulic parameters estimation. In other words, focusing on RMSE, the low-frequency update strategy is clearly better than high-frequency updates. However, this conclusion is less definite when focusing on the differences between correct and predicted soil hydraulic parameters. From the authors' point of view, combinations of parameters that better reproduce the observed data (even if those parameters do not fit accurately the correct values) are more preferable than combinations of parameters that better fit the correct soil parameter value but do not reproduce the observed data very well. Additionally, when comparing the temporal evolution of soil hydraulic parameters and RMSE at the same time, periods with varying soil hydraulic parameters but RMSE remaining almost constant can be observed; this fact highlights the nonuniqueness set of parameters able to provide good simulations. The presence of equifinality problems (Beven and Binley, 1992; Hamraz et al., 2015; Mannina et al., 2010) is expected to increase with more complex soil profiles where the number of model parameters also increases.

We would like to indicate that results of this work may be dependent on choices made when the study was designed. One key variable is the ensemble size. By setting the ensemble size equal to 100 we achieved a practically tenable research timeline for simulations (the computational time exceeded 200 h with an Intel Core 7-4790 3.6GHz processor and 16 GB RAM) with the highly nonlinear flow model. The election of smaller ensemble sizes might lead to overly small uncertainties in the parameters and that performance of high assimilation frequencies could improve for larger ensemble sizes. Efficient methods have been suggested to compensate the relatively small size of the ensemble by using damping factors and inflation (Bauser et al., 2018; Hendricks Franssen and Kinzelbach, 2008) or localization (Bauser et al., 2016; Houtekamer and Mitchell, 2001). We noted that application of both inflation and damping factors require some optimization of their settings for the specific study (e.g., Bauser et al., 2018) or, as in case of the localization strategy, some extra knowledge of the relations among the soil parameters that a conventional user may not have when deciding to apply DA. These facts are behind the authors' decision of not using those strategies in the present study; however, applying these methods presents an interesting avenue for further research. Such research should also address an unexplored issue of the effect of erroneously low parameter and measurement uncertainty on the results of assimilation at different frequencies. Higher frequencies may reduce the parameter uncertainty faster (especially with a limited ensemble size) and consequently cannot include information at later times effectively. Additionally, the possible effect of the measurement errors jointly with the number of replication in the data and the potential manifestation of temporal stability in different locations at the same depth on the EnKF performance should be studied and they limit the results from the present study.

The distance between the initial value of soil hydraulic parameters and the true values can be an additional factor that might affect the results of this study. In the present study, this distance is different for the two soils as a consequence of applying the methodology that adopted initial values provided by Rosetta and the covariance matrixes for soil parameters provided by Faulkner et al. (2003). Therefore, the difference between results for the two soils might be partially attributed to that reason. Alternatively, the same distance between initial and correct values with higher uncertainty in parameters may be adopted regardless other previous information.

Overall, results of our work are limited to the present study and are far from exhaustive. They do, however, elucidate the fact that the measurement frequency can have a nontrivial influence on results of application of EnKF along with control parameters and modifications of this DA method (Keller et al., 2018; Bauser et al., 2018).

Finally, the reader has to know that the study in its present form does not consider model errors since the observations were obtained synthetically. Those synthetic observations were obtained after the inclusion of white noise based on variance and covariance observed in real data, which, in part, may counteract the application of a nonerror model. The replication of the study with real data and the consideration of model errors could lead to different results. The impact of the covariance matrix for observations adopted, either from previous studies (as in this work) or user-defined, is also a topic that has to be tackled in future investigations because monitoring setups with only one observation for each depth, which are useless for quantifying the covariance matrix in the observations, are quite common in the scientific literature.

Conclusions

The results of this study demonstrate that DA with high update frequencies does not provide results better than those obtained when using low frequencies. The statistical analysis suggests that for sandy loam the assimilation of observed data and model updating every 7 or more days yields better results than assimilation with a <1 wk frequency, which is true for all climatic conditions considered. In the case of loam soils, the same better performance of low-frequency updates is true, also for all climates, if the temporal length is 9 mo or longer. Impact of soil type predominates over climatic conditions when dealing with DA performance. Update frequency appears to be a parameter of the DA process that can be optimized. Conclusions from this study may have been impacted by some of the settings adopted during the research project design: the ensemble size (n = 100 units); the different distance between the initial and the correct soil hydraulic properties adopted for each soil type; the magnitude of the uncertainty in the soil parameters; and no adoption of damping factors, inflation, or localization strategies. Site-specific research with synthetic data may be useful to design soil water monitoring strategies that data will be used for DA.

Conflict of Interest

The authors declare that there is no conflict of interest.

Acknowledgments

This study forms part of the projects GRE15-19 and GRE17-12 financed by the University of Alicante. A post-doctoral research fellowship (CAS 15/00244) funded by the Spanish Ministry of Science and Innovation was awarded to J. Valdes-Abellan for this project.

References

Bauser, H.H., D. Berg, O. Klein, and K. Roth. 2018. Inflation method for ensemble Kalman filter in soil hydrology. Hydrol. Earth Syst. Sci. 22:4921–4934. doi:10.5194/hess-22-4921-2018

Bauser, H.H., S. Jaumann, D. Berg, and K. Roth. 2016. EnKF with closed-eye period: Towards a consistent aggregation of information in soil hydrology. Hydrol. Earth Syst. Sci. 20:4999–5014. doi:10.5194/hess-20-4999-2016

Beven, K., and A. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Processes 6:279–298. doi:10.1002/hyp.3360060305

Burgers, G., P.J. Van Leeuwen and G. Evensen. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126:1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Campbell Scientific. 2012. CS650 and CS655 water content reflectometers. Campbell Scientific, Logan, UT.

Camporese, M., C. Paniconi, M. Putti, and P. Salandin. 2009. Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. Water Resour. Res. 45:W10421. doi:10.1029/2008WR007031

Chen, W., C. Huang, H. Shen, and X. Li. 2015. Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation. Adv. Water Resour. 86:425–438. doi:10.1016/j.advwatres.2015.08.003

Chen, Y., and D. Zhang. 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Adv. Water Resour. 29:1107–1122. doi:10.1016/j.advwatres.2005.09.007

Chirico, G.B., H. Medina, and N. Romano. 2014. Kalman filters for assimilating near-surface observations into the Richards equation—Part 1: Retrieving state profiles with linear and nonlinear numerical schemes. Hydrol. Earth Syst. Sci. 18:2503–2520. doi:10.5194/hess-18-2503-2014

Chu, N., C.L. Huang, X. Li, and P.J. Du. 2015. Simultaneous estimation of surface soil moisture and soil properties with a dual ensemble Kalman smoother. Sci. China Earth Sci. 58:2327–2339. doi:10.1007/s11430-015-5175-6

Cosenza, P. 2016. Indirect determination of soil water content. In: P. Delage et al., editors, 3rd European Conference on Unsaturated Soils (E-UNSAT 2016), Paris. 12–14 Sept. 2016. E3S Web Conf. 9:04004. doi:10.1051/e3sconf/20160904004

Crosman, E.T., and J.D. Horel. 2010. Sea and lake breezes: A review of numerical studies. Boundary-Layer Meteorol. 137:1–29. doi:10.1007/s10546-010-9517-9

Das, N.N., and B.P. Mohanty. 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone J. 5:296–307. doi:10.2136/vzj2005.0033

Decagon Devices. 2010. 5TE operator manual. Version 6. Decagon Devices, Pullman, WA.

De Lannoy, G.J.M., P.R. Houser, V.R.N. Pauwels, and N.E.C. Verhoest. 2007. State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency. Water Resour. Res. 43:W06401. doi:10.1029/2006WR005100

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99:10143–10162. doi:10.1029/94JC00572

Evensen, G. 2009. Data assimilation: The ensemble Kalman filter. Springer, Berlin. doi:10.1007/978-3-642-03711-5

Faulkner, B.R., W.G. Lyon, F.A. Khan, and S. Chattopadhyay. 2003. Modeling leaching of viruses by the Monte Carlo method. Water Res. 37:4719–4729. doi:10.1016/S0043-1354(03)00419-6

Franssen, H.J.H., and W. Kinzelbach. 2009. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. J. Hydrol. 365:261–274. doi:10.1016/j.jhydrol.2008.11.033

Ha, S., C. Snyder, W.C. Skamarock, J. Anderson, and N. Collins. 2017. Ensemble Kalman filter data assimilation for the Model for Prediction Across Scales (MPAS). Mon. Weather Rev. 145:4673–4692. doi:10.1175/MWR-D-17-0145.1

Hamraz, B., A. Akbarpour, M.P. Bilondi, and S.S. Tabas. 2015. On the assessment of ground water parameter uncertainty over an arid aquifer. Arab. J. Geosci. 8:10759–10773. doi:10.1007/s12517-015-1935-z

Hendricks Franssen, H.J., and W. Kinzelbach. 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44:W09408. doi:10.1029/2007WR006505

Houtekamer, P.L., and H.L. Mitchell. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129:123–137. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Huang, C., X. Li, L. Lu, and J. Gu. 2008. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens. Environ. 112:888–900. doi:10.1016/j.rse.2007.06.026

Jacques, D. 2000. Analysis of water flow and solute transport at the field scale. Faculteit Landbouwkundige en Toegepaste Wetenschappen, K.U. Leuven, Belgium.

Kaufman, D.E., M.A.M. Friedrichs, J.C.P. Hemmings, and W.O. Smith. 2018. Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea. Biogeosciences 15:73–90. doi:10.5194/bg-15-73-2018

Keller, J., H.J. Hendricks Franssen, and G. Marquart. 2018. Comparing seven variants of the ensemble Kalman filter: How many synthetic experiments are needed? Water Resour. Res. 54:6299–6318. doi:10.1029/2018WR023374

Lei, L., D.R. Stauffer, S.E. Haupt and G.S. Young. 2012. A hybrid nudging-ensemble Kalman filter approach to data assimilation: I. Application in the Lorenz system. Tellus A 64:18484. doi:10.3402/tellusa.v64i0.18484

Li, C., and L. Ren. 2011. Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter. Vadose Zone J. 10:1205–1227. doi:10.2136/vzj2010.0159

Lu, H., K. Yang, and J. Shi. 2016. Constraining the water imbalance in a land data assimilation system through a recursive assimilation scheme. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) Beijing, China. 10–15 July 2016. Institute of Electrical and Electronics Engineers, Piscataway, NJ. p. 2993–2996. doi:10.1109/IGARSS.2016.7729773

Mannina, G., G. Di Bella, and G. Viviani. 2010. Uncertainty assessment of a membrane bioreactor model using the GLUE methodology. Biochem. Eng. J. 52:263–275. doi:10.1016/j.bej.2010.09.001

Martinez, G., L. Brocca, H.H. Gerke, and Y.A. Pachepsky. 2017. Soil variability and biogeochemical fluxes: Toward a better understanding of soil processes at the land surface. Vadose Zone J. 16(10). doi:10.2136/vzj2017.07.0145

Martínez, G., Y.A. Pachepsky, and H. Vereecken. 2014. Temporal stability of soil water content as affected by climate and soil hydraulic properties: A simulation study. Hydrol. Processes 28:1899–1915. doi:10.1002/hyp.9737

Mazzoleni, M., M. Verlaan, L. Alfonso, M. Monego, D. Norbiato, M. Ferri, and D.P. Solomantine. 2017. Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? Hydrol. Earth Syst. Sci. 21:839–861. doi:10.5194/hess-21-839-2017

Medina, H., N. Romano, and G.B. Chirico. 2014a. Kalman filters for assimilating near-surface observations into the Richards equation—Part 3: Retrieving states and parameters from laboratory evaporation experiments. Hydrol. Earth Syst. Sci. 18:2543–2557. doi:10.5194/hess-18-2543-2014

Medina, H., N. Romano, and G.B. Chirico. 2014b. Kalman filters for assimilating near-surface observations into the Richards equation: 2. A dual filter approach for simultaneous retrieval of states and parameters. Hydrol. Earth Syst. Sci. 18:2521–2541. doi:10.5194/hess-18-2521-2014

Monsivais-Huertero, A., W.D. Graham, J. Judge, and D. Agrawal. 2010. Effect of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF. Adv. Water Resour. 33:468–484. doi:10.1016/j.advwatres.2010.01.011

Montzka, C., H. Moradkhani, L. Weihermüller, H.J.H. Franssen, M. Canty, and H. Vereecken. 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399:410–421. doi:10.1016/j.jhydrol.2011.01.020

Moradkhani, H., S. Sorooshian, H.V. Gupta, and P.R. Houser. 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28:135–147. doi:10.1016/j.advwatres.2004.09.002

Mualem, Y. 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12:513–522. doi:10.1029/WR012i003p00513

Nicks, A.D., L.J. Lane, and G.A. Gander. 1995. Weather generator. In: D.C. Flanagan and M.A. Nearing, editors, USDA-Water Erosion Prediction Project: Hillslope profile and watershed model documentation. Natl. Soil Erosion Res. Lab., West Lafayette, IN. p. 2.1–2.22.

Or, D., and R.J. Hanks. 1992. Spatial and temporal soil water estimation considering soil variability and evapotranspiration uncertainty. Water Resour. Res. 28:803–814. doi:10.1029/91WR02585

Pachepsky, Y.A., A.K. Guber, and D. Jacques. 2005. Temporal persistence in vertical distributions of soil moisture contents. Soil Sci. Soc. Am. J. 69:347–352. doi:10.2136/sssaj2005.0347

Pachepsky, Y., F. Pan, and G. Martinez. 2014. Sensor network data assimilation in soil water flow modeling. In: W.G. Teixeira et al, editors, Application of soil physics in environmental analyses: Measuring, modelling and data integration. Springer, Cham, Switzerland. p. 239–260. doi:10.1007/978-3-319-06013-2_11

Pan, F., Y. Pachepsky, D. Jacques, A. Guber, and R.L. Hill. 2012. Data assimilation with soil water content sensors and pedotransfer functions in soil water flow modeling. Soil Sci. Soc. Am. J. 76:829–844. doi:10.2136/sssaj2011.0090

Quinn, G.P., and M.J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge Univ. Press, Cambridge, UK. doi:10.1017/CBO9780511806384

Reichle, R.H. 2008. Data assimilation methods in the Earth sciences. Adv. Water Resour. 31:1411–1418. doi:10.1016/j.advwatres.2008.01.001

Reichle, R.H., and R.D. Koster. 2003. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. J. Hydrometeorol. 4:1229–1242. doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2

Reichle, R.H., D.B. McLaughlin, and D. Entekhabi. 2002. Hydrologic data assimilation with the ensemble Kalman filter. Mon. Weather Rev. 130:103–114. doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2

Richards, L.A. 1931. Capillary conduction of liquids through porous mediums. J. Appl. Phys. 1:318–333.

Rosolem, R., T. Hoar, A. Arellano, J.L. Anderson, W.J. Shuttleworth, X. Zeng, and T.E. Franz. 2014. Translating aboveground cosmic-ray neutron intensity to high-frequency soil moisture profiles at sub-kilometer scale. Hydrol. Earth Syst. Sci. 18:4363–4379. doi:10.5194/hess-18-4363-2014

Schaap, M.G., and F.J. Leij. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163:765–779. doi:10.1097/00010694-199810000-00001

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251:163–176. doi:10.1016/S0022-1694(01)00466-8

Šimůnek, J., M. Šejna, H. Saito, M. Sakai, and M.Th. van Genuchten. 2009. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variability saturated media, Version 4.08. Dep. of Environ. Sci.. Univ. of California, Riverside.

Song, X., L. Shi, M. Ye, J. Yang, and I.M. Navon. 2014. Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling. Vadose Zone J. 13(2). doi:10.2136/vzj2013.05.0083

Twarakavi, N.K.C., J. Šimůnek, and M.G. Schaap. 2009. Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Sci. Soc. Am. J. 73:1443–1452. doi:10.2136/sssaj2008.0021

Valdes-Abellan, J., Y. Pachepsky, and G. Martinez. 2018a. MATLAB algorithm to implement soil water data assimilation with the Ensemble Kalman Filter using HYDRUS. MethodsX 5:184–203. doi:10.1016/j.mex.2018.02.008

Valdes-Abellan, J., Y. Pachepsky, and G. Martinez. 2018b. Obtaining soil hydraulic parameters from soil water content data assimilation under different climatic/soil conditions. Catena 163:311–320. doi:10.1016/j.catena.2017.12.022

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44:892–898. doi:10.2136/sssaj1980.03615995004400050002x

Vereecken, H., J.A. Huisman, H. Bogena, J. Vanderborght, J.A. Vrugt, and J.W. Hopmans. 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resour. Res. 46:W00D06. doi:10.1029/2008WR006829

Vrugt, J.A., C.G.H. Diks, H.V. Gupta, W. Bouten, and J.M. Verstraten. 2005a. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour. Res. 41:W01017. doi:10.1029/2004WR003059

Vrugt, J.A., B.A. Robinson, and V.V. Vesselinov. 2005b. Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. Geophys. Res. Lett. 32:L18408. doi:10.1029/2005GL023940

Wang, D., Y. Chen, and X. Cai. 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. Water Resour. Res. 45:W11416. doi:10.1029/2008WR007401

Wigley, T.M.L., K.R. Briffa and P.D. Jones. 1984. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. J. Clim. Appl. Meteorol. 23:201–213.

Williams, J.W., R.C. Izaurralde, and E.M. Steglich. 2008. Agricultural Policy/Environmental eXtender Model. Theoretical documentation v0604. AgriLIFE Research, Temple, TX.

Wösten, J.H.M., A. Lilly, A. Nemes, and C. Le Bas. 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90:169–185. doi:10.1016/S0016-7061(98)00132-3

Wu, C.C., and S.A. Margulis. 2013. Real-time soil moisture and salinity profile estimation using assimilation of embedded sensor datastreams. Vadose Zone J. 12(1). doi:10.2136/vzj2011.0176

Zhang, H., H.J.H. Franssen, X. Han, J.A. Vrugt, and H. Vereecken. 2017. State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. Hydrol. Earth Syst. Sci. 21:4927–4958. doi:10.5194/hess-21-4927-2017

Zhou, H., J.J. Gómez-Hernández, H.J. Hendricks Franssen, and L. Li. 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv. Water Resour. 34:844–864. doi:10.1016/j.advwatres.2011.04.014