

Water Resources Research®

RESEARCH ARTICLE

10.1029/2021WR030747

Key Points:

- We propose two parameter-estimation methods to infer evapotranspiration rates and root water uptake profiles from soil-moisture sensor data
- Our innovation is to treat the evapotranspiration sink term in the Richards equation, rather than soil moisture, as the observable to update
- The new methods are two orders of magnitude faster than the current implementations of ensemble Kalman filter

Correspondence to:

D. M. Tartakovsky,
tartakovsky@stanford.edu

Citation:

Li, W., Wainwright, H. M., Yan, Q., Zhou, H., Dafflon, B., Wu, Y., et al. (2021). Estimation of evapotranspiration rates and root water uptake profiles from soil moisture sensor array data. *Water Resources Research*, 57, e2021WR030747. <https://doi.org/10.1029/2021WR030747>

Received 4 JUL 2021
Accepted 3 NOV 2021

Author Contributions:

Conceptualization: Haruko M. Wainwright, Qina Yan, Baptiste Dafflon, Yuxin Wu

Data curation: Weiyu Li

Formal analysis: Weiyu Li, Haruko M. Wainwright, Qina Yan

Funding acquisition: Haruko M. Wainwright, Roelof Versteeg, Daniel M. Tartakovsky

Investigation: Weiyu Li, Haruko M. Wainwright, Qina Yan

Methodology: Weiyu Li, Haruko M. Wainwright, Qina Yan, Haiyan Zhou, Baptiste Dafflon, Yuxin Wu, Daniel M. Tartakovsky

Project Administration: Haruko M. Wainwright, Roelof Versteeg

Software: Weiyu Li, Haruko M. Wainwright, Qina Yan, Haiyan Zhou

Supervision: Daniel M. Tartakovsky

Validation: Weiyu Li, Haruko M. Wainwright

Visualization: Weiyu Li

© 2021. American Geophysical Union.
All Rights Reserved.

Estimation of Evapotranspiration Rates and Root Water Uptake Profiles From Soil Moisture Sensor Array Data

Weiyu Li¹, Haruko M. Wainwright² , Qina Yan², Haiyan Zhou³, Baptiste Dafflon² , Yuxin Wu² , Roelof Versteeg³ , and Daniel M. Tartakovsky¹ 

¹Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA, ²Lawrence Berkeley National Laboratory, Berkeley, CA, USA, ³Subsurface Insights, Hanover, NH, USA

Abstract Evapotranspiration is arguably the least quantified component of the hydrologic cycle. We propose two complementary strategies for estimation of evapotranspiration rates and root water uptake profiles from soil-moisture sensor-array data. One is our implementation of ensemble Kalman filter (EnKF); it treats the evapotranspiration sink term in the Richards equation, rather than soil moisture, as the observable to update. The other is a maximum likelihood estimator (MLE) applied to the same observable; it is supplemented with the Fisher information matrix to quantify uncertainty in its predictions. We use numerical experiments to demonstrate the accuracy and computational efficiency of these techniques. We found our EnKF implementation to be two orders of magnitude faster than either the standard EnKF or MLE, and our MLE procedure to require an order of magnitude fewer iterations to converge than its counterpart applied to soil moisture. These findings render our methodologies a viable and practical tool for estimation of the root water uptake profiles and evaporation rates, with the MLE technique to be used when the prior knowledge about evapotranspiration at the site is elusive.

1. Introduction

Evapotranspiration (ET), an amalgam of evaporation from the soil to the atmosphere and plant transpiration in the root zone, is a critical component of the hydrologic cycle. The ability to quantify ET is crucial to science-based predictions of an ecosystem's dynamics and health (Fath, 2018; Fisher et al., 2011); and the impact of droughts, precipitation patterns and snow-melt on groundwater recharge (Dezsi et al., 2018; Doble & Crosbie, 2017), and on plant growth (Severino & Tartakovsky, 2015). It is also of paramount importance for smart agriculture (Kaur Saggi & Jain, 2020; Torres et al., 2020; Yan et al., 2019) and sustainable groundwater management (Jakeman et al., 2016; Loáiciga, 2017).

Satellite-based remote sensing is routinely used to estimate ET at the regional scale (Carter & Liang, 2019; Chen & Liu, 2020). It yields extensive data sets that inform regional-scale models and decisions informed by such models. Yet, this approach relies on a number of physical assumptions to convert the observables, such land surface temperature and wind speed, into ET measurements; consequently, their measurement errors can be hard to ascertain and are often biased (Long et al., 2014; Talsma et al., 2018; Weerasinghe et al., 2020). In addition, remote sensing-based methods often do not consider soil textures and other soil properties, which play key roles in ET rates (Miller & Aarstad, 1973). Moreover, the spatiotemporal resolution of these ET data undermines their usefulness for freshwater resource monitoring and calibration of hydrologic models (Herman et al., 2018), and for other above-mentioned applications that require point-wise information.

In parallel, several efforts have been made to estimate ET from soil moisture time-series data (Guderle & Hildebrandt, 2015; Hupet et al., 2002). In its simplest form, this strategy relies on the water balance approach (Breña Naranjo et al., 2011; Green & Clothier, 1995). Evaporation and root water uptake profiles are inferred from soil moisture time-series by solving an inverse problem, in which these phenomena are represented as either boundary conditions for or a source term in the (partial differential) equation of water flow in the vadose zone. All these approaches assume that the soil water flux to groundwater is negligible, which allows one to compute the source term in the Richards equation as the difference between measured water contents at two observation times. Recent advances in low-cost soil moisture sensors (and sensor networks) and real-time telemetry connections (Jackisch et al., 2020; Protim Goswami et al., 2019; Xaver

Writing – original draft: Weiyu Li
Writing – review & editing: Haruko M. Wainwright, Qina Yan, Haiyan Zhou, Baptiste Dafflon, Yuxin Wu, Roelof Versteeg, Daniel M. Tartakovsky

et al., 2020) provide an opportunity to improve ET estimation by accounting for local hydrologic conditions such as microtopography and soil properties.

To the best of our knowledge, these hardware advances remain decoupled from effective inverse modeling tools. A brute-force numerical inversion of the Richards equation, based on Monte Carlo exploration of the parameter space, yields accurate estimates of both total ET and root water uptake; yet it is computationally expensive and requires considerable knowledge (an accurate “informative prior”) of the model parameters (Hupet et al., 2002). Deterministic alternatives for solving this inverse problem (Guderle & Hildebrandt, 2015; Zuo et al., 2004) can be less expensive, but their convergence is not guaranteed and their robustness to measurement and model errors is questionable.

As an inverse modeling tool, Bayesian data assimilation techniques for parameter estimation proved its worth in various settings, including flow and transport in the vadose zone (Bauser et al., 2021, and references therein). Computationally efficient approximations of the Bayesian update are provided by various flavors of Kalman filter, including ensemble Kalman filter or EnKF (Evensen, 1994; Yang et al., 2020). They assume the observable (e.g., the system state) to be Gaussian, which generally requires the underlying model to be linear. While they often work for nonlinear problems, for example, for estimation of soil water content and total ET rates (Pan & Wood, 2006; Reichle et al., 2008), their performance is not guaranteed. For example, the use of EnKF to infer the spatial distribution of root water uptake from soil moisture data had limited success (Hupet et al., 2002).

We develop two alternative algorithms both to estimate total ET and root water uptake profiles from soil moisture measurements and to quantify uncertainty inherent in such estimators. Section 2 contains our formulation of this estimation problem. In Sections 3, we present our implementations of EnKF (Section 3.1) and maximum likelihood estimation (MLE) (Section 3.2), whose key innovation is to treat the sink term, rather than saturation, as a random variable to be estimated. Another innovation is the deployment of the Fisher information matrix to quantify uncertainty in the MLE predictions of ET and root water uptake. Numerical experiments reported in Section 4 serve to investigate the accuracy and computational efficiency of these two parameter estimation strategies and to contrast them with those of the EnKF implementation by Reichle et al. (2008). Main conclusions drawn from this study are summarized in Section 5.

2. Problem Formulation

Following Guderle and Hildebrandt (2015), we consider an array of N_{sen} moisture sensors placed vertically at locations z_n ($n = 1, \dots, N_{\text{sen}}$) throughout a soil column of length L . Each sensor takes N_{obs} measurements of the (true, yet unknowable) volumetric soil water content $\tilde{\theta}(z, t)$ at times t_k ($k = 1, \dots, N_{\text{obs}}$) separated by a time interval Δt_{obs} . These measurements, denoted by θ_n^k , differ from their “true” values $\tilde{\theta}(z_n, t_k)$ by a random measurement error ϵ_n^k , such that

$$\theta_n^k = \tilde{\theta}(z_n, t_k) + \epsilon_n^k, \quad n = 1, \dots, N_{\text{sen}}, \quad k = 1, \dots, N_{\text{obs}}. \quad (1)$$

In the absence of measurement bias, the ensemble mean of these errors is $\langle \epsilon_n^k \rangle = 0$ for all n and k ; and the errors at different space-time locations are mutually uncorrelated, $\langle \epsilon_n^k \epsilon_v^l \rangle = \sigma^2 \delta_{nv} \delta_{kl}$ with σ^2 and δ_{ij} denoting the noise strength (variance) and the Kronecker delta, respectively. Our goal is to estimate evaporation rate $E(\theta, t)$ and root water uptake (transpiration) $T(\theta, z, t)$ from the data set $\mathbf{d} = \{\theta_n^k : n = 1, \dots, N_{\text{sen}}; k = 1, \dots, N_{\text{obs}}\}$.

The one-dimensional Richards equation,

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[K(\theta) \left(\frac{\partial \psi}{\partial z} + 1 \right) \right] + S(\theta, z, t), \quad 0 < z < L, \quad t > 0, \quad (2)$$

relates volumetric soil water content $\theta(z, t)$ to these two quantities of interest. Here, t is the time, z is the vertical coordinate from the soil surface and positive downwards, $K(\psi)$ is the unsaturated hydraulic conductivity of the soil that varies with the pressure head $\psi(z, t)$, and the source/sink term,

$$S(\theta, z, t) = T(\theta, z, t), \quad (3)$$

is referred to as transpiration (root water uptake function). This equation of water flow in partially saturated soils is closed by specifying the constitutive relations $\psi = \psi(\theta)$, $K = K(\theta)$ and $T = T(\theta)$. To be specific, we select the van Genuchten model for the first two expressions. We use a canonical form (Perrochet, 1987),

$$T(\theta, z, t) = T_{\max}(t)\gamma_T(\theta)F_{\text{root}}(z), \quad (4)$$

to relate the root water uptake function $T(\theta, z, t)$ (1/hr) to the potential transpiration rate $T_{\max}(t)$ (m/hr), the root density (1/m) (Schenk & Jackson, 2002)

$$F_{\text{root}}(z) = -\frac{c}{z_{50}} \left(\frac{z}{z_{50}}\right)^{c-1} \left[1 + \left(\frac{z}{z_{50}}\right)\right]^{-2} \quad \text{with} \quad c = \frac{1.27875}{\lg(z_{50}) - \lg(z_{95})}, \quad (5)$$

and the root uptake water-stress response function (Guswa et al., 2002; Porporato et al., 2003)

$$\gamma_T(\theta) = \begin{cases} 0 & 0 \leq \theta < \theta_w \\ \frac{\theta - \theta_w}{\theta^* - \theta_w} & \theta_w < \theta \leq \theta^* \\ 1 & \theta^* < \theta \leq \theta_p. \end{cases} \quad (6)$$

In the expressions above, z_{50} and z_{95} represent the soil depth above which 50% and 95% of water uptake occurs; θ_w is the saturation at the wilting point at which the uptake is zero and a plant wilts; θ^* is the saturation at the point of stomatal closure where the uptake is equal to the demand; and θ_p is the field capacity that is generally larger than θ^* . One can verify that $\int_0^L F_{\text{root}}(z)dz = 1$. Finally, while evaporation can occur over a finite soil depth, we follow the standard practice by restricting it to the soil surface. That relegates $E(\theta, t)$ to the boundary condition at $z = 0$, as detailed below.

Mass conservation at the soil surface gives rise to the boundary condition

$$K(\theta) \left(\frac{\partial \psi}{\partial z} + 1 \right) = P(t) - E(\theta, t), \quad z = 0, \quad t > 0, \quad (7)$$

where $P(t)$ is the infiltration (m/hr); and the evaporation rate is defined, in analogy to Equations 4–6, as

$$E(\theta(0, t), t) = E_{\max}(t)\gamma_E(\theta(0, t)). \quad (8)$$

Here, E_{\max} is the maximum evaporation rate, and the evaporation reduction factor γ_E is given by (Hale & Orcutt, 1987)

$$\gamma_E(\theta(0, t)) = \begin{cases} 0 & 0 \leq \theta(0, t) < \theta_h \\ \frac{\theta(0, t) - \theta_h}{\theta_w - \theta_h} & \theta_h < \theta(0, t) \leq \theta_w \\ 1 & \theta_w < \theta(0, t) \leq \theta_p, \end{cases} \quad (9)$$

where θ_h is the hygroscopic saturation at which evaporation diminishes. At the bottom of the soil column, $z = L$, we impose a free-drainage condition,

$$\frac{\partial \psi}{\partial z} + 1 = 0, \quad z = L, \quad t > 0. \quad (10)$$

Finally, we assume the initial water content in the soil column, θ_0 , to be uniform, giving rise to the initial condition

$$\theta(0 \leq z \leq L, 0) = \theta_0. \quad (11)$$

If all the model parameters in Equations 2–11 were known with certainty, one could compute temporal snapshots of the moisture profile throughout the soil column, $\theta(z, t)$, and, hence, predict the evaporation rate $E(\cdot, t)$ and root water uptake profile $T(\cdot, z, t)$ with a required precision. In reality, a few, if any, of these parameters are measured and the model predictions can be thought of as an “educated guess” that must be refined as relevant observations become available. In our study, the static parameters, including hydraulic conductivity and van Genuchten parameters, are assumed to be known or estimated in advance through infiltration tests. Of all the parameters in the flow model, Equations 2–11, we focus on the two, $T_{\max}(t)$ and $E_{\max}(t)$, that are temporally variable. A specific goal of our study is to estimate $T_{\max}(t)$ and $E_{\max}(t)$ from the measurements of water content, $\theta_n(t) \equiv \theta(z_n, t)$ with $n = 1, \dots, N_s$.

3. Efficient Strategies for Parameter Estimation

We introduce two strategies for estimating the evaporation rate $E(\cdot, t)$ and root water uptake profile $T(\cdot, z, t)$. Section 3.1 contains a description of our implementation of EnKF, whose key innovation is to treat the source function $S(z, t)$ and driving functions at the soil surface (rather than soil water content) as state variables in the Bayesian update. The second strategy, presented in Section 3.2, is the MLE of these state variables, which is enriched by the Fisher information matrix to quantify predictive uncertainty.

To facilitate comparison between the two parameter estimation techniques, both utilize the same data processing strategy to construct new observables, E^k and T_n^k , from the directly observed θ_n^k in Equation 1. We start by subdividing the soil column $[0, L]$ into N_{sen} elements of length Δz_{obs} such that each element contains one sensor. (In general, each element can have a different length if the sensors are nonuniformly spaced; we assume all the elements to have the same length Δz_{obs} to simplify the notation). Then, the total ET rate at time t_k , S_{tot}^k , is related to the observed infiltration rate $P^k \equiv P(t_k)$ and the soil moisture observations $\mathbf{d} = \{\theta_n^k\}$ by water balance (Breña Naranjo et al., 2011; Wilson et al., 2001),

$$S_{\text{tot}}^k \Delta z_{\text{obs}} = P^k - \sum_{n=1}^{N_{\text{sen}}} \frac{\theta_n^k - \theta_n^{k-1}}{\Delta t_{\text{obs}}} \Delta z_{\text{obs}}, \quad k = 1, \dots, N_{\text{obs}}. \quad (12)$$

Let $\theta^*(z, t)$ denote a solution to Equations 2–11 without the ET losses, that is, with $E \equiv 0$ and $T \equiv 0$; it is computed on the time interval $[t_{k-1}, t_k]$, for the initial condition given by Equation 11 if $k = 1$, or by the posterior mean of the water content at time t_{k-1} if $k > 1$ (see below). In either case, $\theta^*(z, t)$ is deterministic. Guderle and Hildebrandt (2015) also used a solution to Equations 2–11 without the ET losses within their inverse modeling framework. This is assuming that soil moisture changes in a unit time can be separated into the vertical flow and sink terms. This is a valid assumption when $K(\theta)$ does not change significantly within δt , and Equation 3 can be solved as a linear system. In reality, soil moisture changes rapidly during infiltration, but slowly during the drying period (see, e.g., Figure 2 in Guderle & Hildebrandt, 2015). Our formulation is, therefore, valid in the drying period.

Both $\theta^*(z, t)$ and the full solution $\theta(z, t)$ are obtained by solving Equations 2–11 numerically on a mesh consisting of N_{el} elements of size Δz , with time step Δt . (Sensor array design and the accuracy of a numerical solution dictate that $\Delta x \ll \Delta x_{\text{obs}}$ and $\Delta t \ll \Delta t_{\text{obs}}$). Motivated by Equation 12, we approximate an uncertain prediction of the rate of change in water content of the n th element of the numerical mesh, $[z_{n-1}, z_n]$ during the k th observational time step, $[t_{k-1}, t_k]$, as

$$\frac{\theta(z, t) - \theta(z, t - \Delta t_{\text{obs}})}{\Delta t_{\text{obs}}} \Delta z = \frac{\theta^*(z, t) - \theta(z, t - \Delta t_{\text{obs}})}{\Delta t_{\text{obs}}} \Delta z - S(z, t) \Delta z, \quad (13)$$

for $z \in [z_{n-1}, z_n]$ and $t \in [t_{k-1}, t_k]$, with $n = 1, \dots, N_{\text{el}}$ and $k = 1, \dots, N_{\text{obs}}$. In the first ($n = 1$) element, the source function represents the ET, $S(z, t) \Delta z \equiv E(\theta(0), t) + T(\theta(0), t) \Delta z$, while in the remaining ($n > 1$) elements, $S(z, t) \equiv T(\theta(z), t)$ for $z \in [z_{n-1}, z_n]$. As we are using a uniform numerical grid with size Δz , from now on, we will use $S(z, t)$ (m/hr) to represent $S(z, t) \Delta z$. Rearranging the terms yields

$$S(z, t) = \frac{\theta^*(z, t) - \theta(z, t)}{\Delta t_{\text{obs}}} \Delta z, \quad z \in [z_{n-1}, z_n], \quad t \in [t_{k-1}, t_k], \quad (14)$$

for $n = 1, \dots, N_{el}$ and $k = 1, \dots, N_{obs}$. Implicit in this calculation is an assumption that the ET does not change during the time interval Δt_{obs} between any two adjacent observation times t_{k-1} and t_k .

3.1. Ensemble Kalman Filter

Between any two observation times t_{k-1} and t_k , the uncertain input parameters T_{max}^k and E_{max}^k are treated as independent Gaussian random variables. Their probability density functions (PDFs) $f_{T_{max}}(\xi_{T_{max}})$ and $f_{E_{max}}(\xi_{E_{max}})$ have respective means $\mu_{T_{max}}$ and $\mu_{E_{max}}$ and standard deviations (SDs) $\sigma_{T_{max}}$ and $\sigma_{E_{max}}$,

$$f_p(\xi_p) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(\xi_p - \mu_p)^2}{2\sigma_p^2}\right], \quad p = T_{max}, E_{max}. \quad (15)$$

The predicted values of the sink term $S(z, t)$ in Equation 14 are arranged into a vector $\mathbf{S}(t_k) = (S_1, \dots, S_{N_{el}})^T$ of length N_{el} . Although many stochastic analyses of unsaturated flow found a solution of the Richards Equation 2 to be non-Gaussian (Lu et al., 2002; Tartakovsky, Guadagnini, & Riva, 2003; Tartakovsky, Lu, et al., 2003, among others), we follow the standard practice in EnKF by treating \mathbf{S} , at any time t_k , as multivariate Gaussian with PDF

$$f_{\mathbf{S}}(\mathbf{s}; t_k) = \frac{1}{(2\pi)^{N_{el}/2} |\boldsymbol{\Sigma}_{\mathbf{S}}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{s} - \boldsymbol{\mu}_{\mathbf{S}})^T \boldsymbol{\Sigma}_{\mathbf{S}}^{-1} (\mathbf{s} - \boldsymbol{\mu}_{\mathbf{S}})\right]. \quad (16)$$

Monte Carlo simulations, which comprise N_{MCS} solves of Equation 14 for N_{MCS} realizations of T_{max} and E_{max} drawn from Equation 15, are used to estimate the $N_{el} \times 1$ vector of sample mean $\boldsymbol{\mu}_{\mathbf{S}}(t_k)$ and the $N_{el} \times N_{el}$ sample covariance matrix $\boldsymbol{\Sigma}_{\mathbf{S}}(t_k)$ for $\mathbf{S}(t_k)$.

At any measurement time t_k ($k = 1, \dots, N_{obs}$), the N_{sen} -dimensional vector of observed soil moisture values, \mathbf{d}^k , is multivariate Gaussian with mean $\bar{\mathbf{d}}^k$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{d}}$. The off-diagonal terms of the latter are all 0, and the diagonal terms are σ^2 . Given the data model in Equation 1, and assuming a discretized solution $\boldsymbol{\theta}^{(l)}(t) = (\theta_1^{(l)}, \dots, \theta_{N_{el}}^{(l)})^T$ of Equations 2–11 in its l th Monte Carlo realization ($l = 1, \dots, N_{MCS}$) to be a perfect representation of reality, $\bar{\mathbf{d}}^k$ and $\boldsymbol{\theta}^{(l)}(t_k)$ are related by

$$\bar{\mathbf{d}}^k = \mathbf{H}_{obs}^k \boldsymbol{\theta}^{(l)}. \quad (17)$$

The $N_{sen} \times N_{el}$ observational matrix \mathbf{H}_{obs}^k ensures that the numerical solution $\boldsymbol{\theta}^{(l)}$, computed with Δz and Δt , is available at the same locations and times as the observations \mathbf{d} collected with Δz_{obs} and Δt_{obs} , if necessary via interpolation; the superscript k in \mathbf{H}_{obs}^k indicates that this mapping is done for time t_k .

Consider a linear transformation of the soil moisture data at time t_k ,

$$\mathbf{Y}^k = \frac{\mathbf{H}_{obs}^k \boldsymbol{\theta}^* - \mathbf{d}^k}{\Delta t_{obs}} \Delta z, \quad k = 1, \dots, N_{obs}, \quad (18)$$

where the $N_{el} \times 1$ vector $\boldsymbol{\theta}^*(t)$ is the discretized in space analog of $\theta^*(z, t)$ for $t \in [t_{k-1}, t_k]$. Since \mathbf{d}^k is multivariate Gaussian, so is \mathbf{Y}^k . It follows from Equation 18 and the discretized version of Equation 14 that its mean, $\boldsymbol{\mu}_{\mathbf{Y}}(t_k)$, and covariance matrix, $\boldsymbol{\Sigma}_{\mathbf{Y}}(t_k)$, are given by

$$\boldsymbol{\mu}_{\mathbf{Y}}(t_k) = \mathbf{H}_{obs}^k \mathbf{s} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{Y}}(t_k) = \frac{\Delta z^2}{\Delta t_{obs}^2} \boldsymbol{\Sigma}_{\mathbf{d}}(t_k). \quad (19)$$

When this conditional PDF of \mathbf{Y}^k ,

$$f_{\mathbf{Y}^k|\mathbf{S}=\mathbf{s}}(\mathbf{y}; \mathbf{s}; t_k) = \frac{1}{(2\pi)^{N_{el}/2} |\boldsymbol{\Sigma}_{\mathbf{Y}}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}})^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}})\right], \quad (20)$$

is treated as function of \mathbf{s} , it is referred to as the likelihood function in the Bayesian update. The latter is used to update the prior PDF of $\mathbf{S}(t_k)$, $f_{\mathbf{S}}(\mathbf{s}; t_k)$ in Equation 16, with the observation $\mathbf{Y}^k = \mathbf{y}^k$,

$$f_{\mathbf{S}|\mathbf{Y}^k=\mathbf{y}^k}(\mathbf{s}; t_k) \sim f_{\mathbf{Y}^k|\mathbf{S}=\mathbf{s}}(\mathbf{s}; \mathbf{y}^k; t_k) f_{\mathbf{S}}(\mathbf{s}; t_k), \quad (21)$$

Algorithm 1. Estimation of Evapotranspiration Via EnKF

For $k = 1$:

1. Read infiltration $P(t_1)$
2. Obtain $\theta^*(z, t_1)$ by solving Equations 2–11 with $E \equiv 0$ and $T \equiv 0$
3. Post-process soil moisture data to compute \mathbf{y}^1 in Equation 18
4. For the l th Monte Carlo realization (l from 1 to N_{MCS})
 - (a) Sample T_{max}^1 and E_{max}^1 from their respective PDFs in Equation 15
 - (b) Compute $T(z, t_1)$ and $E(t_1)$ from Equations 4–9, for these values of T_{max}^1 and E_{max}^1 , and with the functionals $\gamma_T(\theta(t))$ and $\gamma_E(\theta(0, t))$ evaluated at $\theta(z, t_0)$, abiding by the explicit scheme
 - (c) Compute the sink term $S(z, t_1)$ using Equation 3
5. Compute the prior mean $\boldsymbol{\mu}_S(t_1)$ and covariance $\boldsymbol{\Sigma}_S(t_1)$ of $\mathbf{S}(t_1)$ in Equation 16
6. Use Equations 23 and 24 to calculate the posterior mean $\hat{\boldsymbol{\mu}}_S(t_1)$ and covariance $\hat{\boldsymbol{\Sigma}}_S(t_1)$ of $\mathbf{S}(t_1)$
7. Solve Equations 2–11 with S given by $\hat{\boldsymbol{\mu}}_S(t_1)$ to forecast $\theta(z, t_1)$

For k from Equation 2 to N_{obs} :

1. Read infiltration $P(t)$ at time t_k
 2. Obtain $\theta^*(z, t_k)$ by solving Equations 2–11 with $E \equiv 0$ and $T \equiv 0$ while the initial condition in Equation 11 is given by $\theta(z, t_{k-1})$
 3. Post-process soil moisture data to compute \mathbf{y}^k in Equation 18
 4. For the l th Monte Carlo realization (l from 1 to N_{MCS})
 - (a) Sample T_{max}^k and E_{max}^k from their respective PDFs in Equation 15
 - (b) Compute $T(z, t_k)$ and $E(t_k)$ from Equations 4–9, for these values of T_{max}^k and E_{max}^k , and with the functionals $\gamma_T(\theta(t))$ and $\gamma_E(\theta(0, t))$ evaluated at $\theta(z, t_{k-1})$, abiding by the explicit scheme
 - (c) Compute the sink term $S(z, t_k)$ using Equation 3
 5. Compute the prior mean $\boldsymbol{\mu}_S(t_k)$ and covariance $\boldsymbol{\Sigma}_S(t_k)$ of $\mathbf{S}(t_k)$ in Equation 16
 6. Use Equations 23 and 24 to calculate the posterior mean $\hat{\boldsymbol{\mu}}_S(t_k)$ and covariance $\hat{\boldsymbol{\Sigma}}_S(t_k)$ of $\mathbf{S}(t_k)$
 7. Solve Equations 2–11 with S given by $\hat{\boldsymbol{\mu}}_S(t_k)$ to forecast $\theta(z, t_k)$
-

yielding the posterior PDF of $\mathbf{S}(t_k)$, $f_{\mathbf{S}|\mathbf{Y}}(\mathbf{s}; t_k)$. This Bayesian approach to data assimilation can be computationally expensive, and often prohibitively so (Boso & Tartakovsky, 2020). EnKF assumes the posterior PDF $f_{\mathbf{S}|\mathbf{Y}}(\mathbf{s}; t_k)$ to be multivariate Gaussian,

$$f_{\mathbf{S}|\mathbf{Y}}(\mathbf{s}; t_k) = \frac{1}{(2\pi)^{N_{\text{obs}}/2} |\hat{\boldsymbol{\Sigma}}_S|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{s} - \hat{\boldsymbol{\mu}}_S)^\top \hat{\boldsymbol{\Sigma}}_S^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_S) \right], \quad (22)$$

that is, replaces the task of computing the full PDF in Equation 21 with that of computing the posterior mean, $\hat{\boldsymbol{\mu}}_S(t_k)$, and posterior covariance, $\hat{\boldsymbol{\Sigma}}_S(t_k)$, of $\mathbf{S}(t_k)$ via the Kalman update,

$$\hat{\boldsymbol{\mu}}_S = \boldsymbol{\mu}_S + \mathbf{K}(\mathbf{y}^k - \mathbf{H}_{\text{obs}}^k \boldsymbol{\mu}_S), \quad \hat{\boldsymbol{\Sigma}}_S = (\mathbf{I} - \mathbf{K} \mathbf{H}_{\text{obs}}^k) \boldsymbol{\Sigma}_S \quad (23)$$

where \mathbf{I} is the $N_{\text{el}} \times N_{\text{el}}$ identity matrix, and the $N_{\text{el}} \times N_{\text{sen}}$ matrix

$$\mathbf{K} = \boldsymbol{\Sigma}_S \mathbf{H}_{\text{obs}}^{k\top} (\mathbf{H}_{\text{obs}}^k \boldsymbol{\Sigma}_S \mathbf{H}_{\text{obs}}^{k\top} + \boldsymbol{\Sigma}_Y)^{-1} \quad (24)$$

is called the Kalman gain. Algorithm 1 provides our implementation of EnKF for estimation of ET from soil moisture data.

Our implementation of EnKF allows one to estimate an ET profile from a vertical array of soil moisture measurements. It is more accurate than the EnKF applied to the total water balance in Equation 12 (Pan & Wood, 2006; Reichle et al., 2008), which ignores the vertical soil water flux distribution. It is also significantly more efficient than their EnKF, which corrects the soil moisture in the Kalman update, because it avoids solving the Richards equation for all ensemble members.

Like any implementation of EnKF, ours requires prior knowledge about the model inputs T_{max} and E_{max} , which is fundamentally uncertain. In its standard incarnations, it is of limited validity, which can be hard to

Algorithm 2. Inference of Evapotranspiration Via MLE

1. Set $T_{\max} = 0$ and $E_{\max} = 0$ at the beginning of each iteration process
2. Compute $T(z, t)$ and $E(t)$ via Equations 4–9
3. Compute $\theta(z, t)$ by solving Equations 2–11
4. For a prescribed learning rate α , update T_{\max} and E_{\max} according to

$$\text{and } T_{\max} = T_{\max} - \frac{\alpha \Delta z}{\Delta t_{\text{obs}} N_{\text{sen}}} \sum_{i=1}^{N_{\text{sen}}} [\theta_i^k - \theta(z_i, t_k)] \gamma_T(\theta(z_i, t_k)) \bar{F}_{\text{root}}(z_i)$$

$$E_{\max} = E_{\max} - \frac{\alpha \Delta z}{\Delta t_{\text{obs}} N_{\text{sen}}} [\theta_1^k - \theta(0, t_k)] \gamma_E(\theta(0, t_k))$$

5. Repeat steps 2–4 until convergence
-

ascertain a priori (Bocquet et al., 2015). As an alternative strategy that is free of this requirement, we deploy an MLE supplemented with the Fisher information matrix to quantify predictive uncertainty.

3.2. Maximum Likelihood Estimator

As a deterministic procedure, MLE ignores (random) measurement errors, that is, sets $\epsilon_n^k \equiv 0$ in the data model Equation 1. At any time t_k , the mismatch between the water content data $\mathbf{d} = \{\theta_n^k : n = 1, \dots, N_{\text{sen}}; k = 1, \dots, N_{\text{obs}}\}$ and the water content prediction $\theta(z, t)$ obtained by solving Equations 2–11 can be quantified in terms of the loss function

$$\tilde{\mathcal{L}}^k = \frac{1}{2N_{\text{sen}}} \sum_{i=1}^{N_{\text{sen}}} [\theta_i^k - \theta(z_i, t_k; T_{\max}^k, E_{\max}^k)]^2. \quad (25)$$

The notation $\theta(\cdot; T_{\max}^k, E_{\max}^k)$ serves to remind the reader that, in our formulation, the model prediction depends on the choice of $T_{\max}(t)$ and $E_{\max}(t)$, all the other input parameters being assumed to have been measured. As before, we define the solution of Equations 2–11 with $T_{\max} \equiv 0$ and $E_{\max} \equiv 0$ by $\theta^*(z, t)$. Then, we rewrite Equation 25 as

$$\tilde{\mathcal{L}}^k = \frac{1}{2N_{\text{sen}}} \sum_{i=1}^{N_{\text{sen}}} [(\theta^*(z_i, t_k) - \theta(z_i, t_k; T_{\max}^k, E_{\max}^k)) - (\theta^*(z_i, t_k) - \theta_i^k)]^2. \quad (26)$$

Accounting for Equation 14,

$$\mathcal{L}^k = \frac{1}{2N_{\text{sen}}} \sum_{i=1}^{N_{\text{sen}}} \left[S(z_i, t_k; T_{\max}^k, E_{\max}^k) - \frac{\Delta z}{\Delta t_{\text{obs}}} (\theta^*(z_i, t_k) - \theta_i^k) \right]^2, \quad (27)$$

where $\tilde{\mathcal{L}}^k = (\Delta z / \Delta t_{\text{obs}})^2 \mathcal{L}^k$. Optimal values of T_{\max}^k and E_{\max}^k are those that minimize \mathcal{L}^k .

A solution of this optimization problem is obtained by setting to 0 the derivatives of \mathcal{L}^k with respect to T_{\max}^k and E_{\max}^k . For the ET model in Equations 4–8, this yields

$$\frac{\partial \mathcal{L}^k}{\partial T_{\max}^k} = 0 = \sum_{i=1}^{N_{\text{sen}}} [\theta_i^k - \theta(z_i, t_k)] \gamma_T(\theta(z_i, t_k)) \bar{F}_{\text{root}}(z_i) \quad (28a)$$

$$\frac{\partial \mathcal{L}^k}{\partial E_{\max}^k} = 0 = [\theta_1^k - \theta(0, t_k)] \gamma_E(\theta(0, t_k)), \quad (28b)$$

where $\bar{F}_{\text{root}} = F_{\text{root}} / \sum_{i=1}^{N_{\text{sen}}} F_{\text{root}}(z_i)$ is the normalized root distribution function at the sensor locations. It follows from Equation 28b that MLE of the parameter E_{\max} requires a soil moisture sensor to be placed at the soil surface (the boundary cell of the model). Algorithm 2 provides our implementation of MLE for inference of ET from soil moisture data.

The convergence criterion is defined in terms of the normalized difference of the cost function \mathcal{L}^k in Equation 27 between two successive iterations, ν and $\nu - 1$,

$$\left| \frac{\mathcal{L}_{(v)}^k - \mathcal{L}_{(v-1)}^k}{\mathcal{L}_{(v-1)}^k} \right| \leq \epsilon, \quad v = 1, 2, \dots, \quad (29)$$

where ϵ is a prescribed tolerance. The convergence criterion is met when the gradient descent makes very small changes in the loss function that is close to its optimal value. A suitable choice of the learning rate α can accelerate convergence without overshooting the minimum.

We use the Fisher information matrix \mathbf{I} to quantify uncertainty in the MLE predictions of T_{\max} and E_{\max} . In analogy with Equation 20, one defines a multivariate-Gaussian likelihood function $f_{\mathbf{d}|\theta}$ for the discretized water content $\theta(t_k) = (\theta_1, \dots, \theta_{N_{\text{sen}}})^\top$, such that its logarithm is

$$\ln f_{\mathbf{d}|\theta} = C - \frac{1}{2} (\mathbf{d}^k - \theta(t_k))^\top \Sigma_{\mathbf{d}}^{-1} (\mathbf{d}^k - \theta(t_k)), \quad (30)$$

where C is the normalizing constant. As in Equation 27, we reformulate this log-likelihood function as

$$\ln f_{\mathbf{d}|\theta} = C - \frac{1}{2} \left(\mathbf{S}(t_k) - \Delta z \frac{\theta^*(t_k) - \mathbf{d}^k}{\Delta t_{\text{obs}}} \right)^\top \Sigma_{\mathbf{Y}}^{-1} \left(\mathbf{S}(t_k) - \Delta z \frac{\theta^*(t_k) - \mathbf{d}^k}{\Delta t_{\text{obs}}} \right), \quad (31)$$

where $\mathbf{S}(t_k) = (S_1, \dots, S_{N_{\text{sen}}})^\top$ is the discretized version of $S(z, t)$ in Equation 14 and is a function of T_{\max} and E_{\max} . For $S(z, t; T_{\max}, E_{\max})$ in Equations 2-11, the 2×2 Hessian matrix

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \ln f_{\mathbf{d}|\theta}}{\partial T_{\max}^2} & \frac{\partial^2 \ln f_{\mathbf{d}|\theta}}{\partial T_{\max} \partial E_{\max}} \\ \frac{\partial^2 \ln f_{\mathbf{d}|\theta}}{\partial E_{\max} \partial T_{\max}} & \frac{\partial^2 \ln f_{\mathbf{d}|\theta}}{\partial E_{\max}^2} \end{bmatrix} \quad (32a)$$

is computed analytically,

$$\mathbf{H} = \begin{bmatrix} -(\boldsymbol{\gamma}_{\text{str}}^T \circ \bar{\mathbf{F}}_r)^\top \Sigma_{\mathbf{Y}}^{-1} (\boldsymbol{\gamma}_{\text{str}}^T \circ \bar{\mathbf{F}}_r) & -(\boldsymbol{\gamma}_{\text{str}}^T \circ \bar{\mathbf{F}}_r)^\top \Sigma_{\mathbf{Y}}^{-1} (\boldsymbol{\gamma}_{\text{str}}^E) \\ -(\boldsymbol{\gamma}_{\text{str}}^E)^\top \Sigma_{\mathbf{Y}}^{-1} (\boldsymbol{\gamma}_{\text{str}}^T \circ \bar{\mathbf{F}}_r) & -(\boldsymbol{\gamma}_{\text{str}}^E)^\top \Sigma_{\mathbf{Y}}^{-1} (\boldsymbol{\gamma}_{\text{str}}^E) \end{bmatrix}. \quad (32b)$$

Here, the symbol \circ denotes the element-wise multiplication, $\boldsymbol{\gamma}_{\text{str}}^T = (\gamma_{w_T}(\tilde{\theta}_1^t), \dots, \gamma_{w_T}(\tilde{\theta}_{n_m}^t))^\top$, $\boldsymbol{\gamma}_{w_E} = [\gamma_{w_E}(\tilde{\theta}_1^t) \ 0 \ \dots \ 0]$, $\bar{\mathbf{F}}_r = [\bar{F}_r(z_1) \ \bar{F}_r(z_2) \ \dots \ \bar{F}_r(z_{n_m})]$. The ensemble mean of \mathbf{H} is the information matrix (Lehmann & Casella, 1998),

$$\mathbf{I} = -\mathbb{E}[\mathbf{H}]. \quad (33)$$

Its inverse is the variance-covariance matrix of MLEs of T_{\max} and E_{\max} (Efron & Hinkley, 1978),

$$\begin{bmatrix} \text{var}(T_{\max}) & \text{cov}(T_{\max}, E_{\max}) \\ \text{cov}(E_{\max}, T_{\max}) & \text{var}(E_{\max}) \end{bmatrix} = \mathbf{I}^{-1}. \quad (34)$$

4. Results of Numerical Experiments

We conduct synthetic experiments to test the accuracy and efficiency of the two new strategies for estimation of dynamic evaporation rates and root water uptake from data recorded by an array of soil-moisture sensors. Section 4.1 contains a description of our experimental design. Evaluation metrics of the relative performance of the two methods are specified in Section 4.2. We investigate the methods' ability to estimate the total ET rate (Section 4.3), and the vertical distributions of root water uptake (Section 4.4) and soil moisture (Section 4.5). The methods' computational costs and convergence rates are reported in Section 4.6.

Table 1
Known Hydraulic Soil Properties and Initial Guesses for the Statistics of Unknown T_{\max} and E_{\max}

Parameter	Symbol	Value	Units
Porosity	ϕ	0.4	-
Permeability	k	$1 \cdot 10^{-13}$	m^2
Residual water content	θ_i	0.05	-
Shape factor in van Genuchten model	α_{vG}	$1 \cdot 10^{-3}$	-
Shape factor in van Genuchten model	m_{vG}	0.5	-
Saturation at the wilting point	θ_w	0.1	-
Saturation at the point of stomatal closure	θ^*	0.2	-
Hygroscopic saturation	θ_h	0.05	-
Depth above which 50% water uptake occurs	z_{50}	0.1	m
Depth above which 95% water uptake occurs	z_{95}	0.6	m
Prior mean of T_{\max}	$\mu_{T_{\max}}$	$2 \cdot 10^{-4}$	m/hr
Prior SD of T_{\max}	$\sigma_{T_{\max}}$	$1 \cdot 10^{-4}$	m/hr
Prior mean of E_{\max}	$\mu_{E_{\max}}$	$4.17 \cdot 10^{-5}$	m/hr
Prior SD of E_{\max}	$\sigma_{E_{\max}}$	$2 \cdot 10^{-5}$	m/hr

4.1. Design of Numerical Experiments

We consider infiltration into a homogeneous soil following a 2-h long rain event. Observations are taken over the time interval of 200 hr, with the rain occurring between 48th and 50th h and the observation being taken with time step $\Delta t_{\text{obs}} = 2$ hr. The soil column is homogeneous, of length 1.5 m; its hydraulic properties are reported in Table 1. The infiltration rate is 0.04 m/h.

We created the time series $T_{\max}(t)$ and $E_{\max}(t)$ as the ground truth (Figure 1). We assigned a discontinuous function of $T_{\max}(t)$ and $E_{\max}(t)$ to demonstrate the ability of the algorithm to capture their variability. The synthetic soil moisture data are then generated by solving the flow problem, Equations 2–11, with the numerical code PFLOTRAN (www.pfлотran.org/; executed on Dell Desktop Computer with Intel Core i7-8700 CPU @ 3.2 GHz and 15.4 GB memory). The soil column is discretized into $N_s = 30$ elements of length $\Delta z = 0.05$ m, and the simulation horizon of 200 hr into $N_t = 10,000$ time steps of duration $\Delta t = 0.02$ hr. The ground truth is reported in Figure 1 in terms of water content $\theta(z, t)$, spatiotemporal distribution of ET rate $S(\cdot, z, t)$, and total ET rate $S_{\text{tot}}(t)$. During the precipitation event, the soil moisture in the top soil increases dramatically and water infiltrates slowly into the soil column. The soil evaporation and root water uptake models imply that $S(z, t)$ is largest in the vicinity of the soil surface due to evaporation and relatively large root water uptake activity.

Unless specified otherwise, the water content predicted with Equations 2–11 is “measured” by an array of $N_{\text{sen}} = 8$ soil moisture sensors located at depths $z_1 = 2.5$ cm, $z_2 = 7.5$ cm, $z_3 = 12.5$ cm, $z_4 = 17.5$ cm, $z_5 = 32.5$ cm, $z_6 = 47.5$ cm, $z_7 = 62.5$ cm, and $z_8 = 97.5$ cm. The measurements at these locations, $\mathbf{d}^k = \{\theta_1^k, \dots, \theta_{N_{\text{sen}}}^k\}$, are generated at observation times t_k ($k = 1, \dots, N_{\text{obs}} = 100$) in accordance with the data model, Equation 1 with $\sigma_\epsilon = 0.001$.

4.2. Performance Evaluation Metrics

We use several complementary metrics to quantify the performance of the alternative strategies for inference of evapotranspiration from soil moisture measurements. We compare the discrepancy between the “real” total ET rate (S_{tot}) and its estimation ($S_{\text{tot}}^{\text{est}}$) in terms of the relative bias b , the correlation coefficient R , and the relative measure RV . These are defined as (Guderle & Hildebrandt, 2015; Gupta et al., 2009)

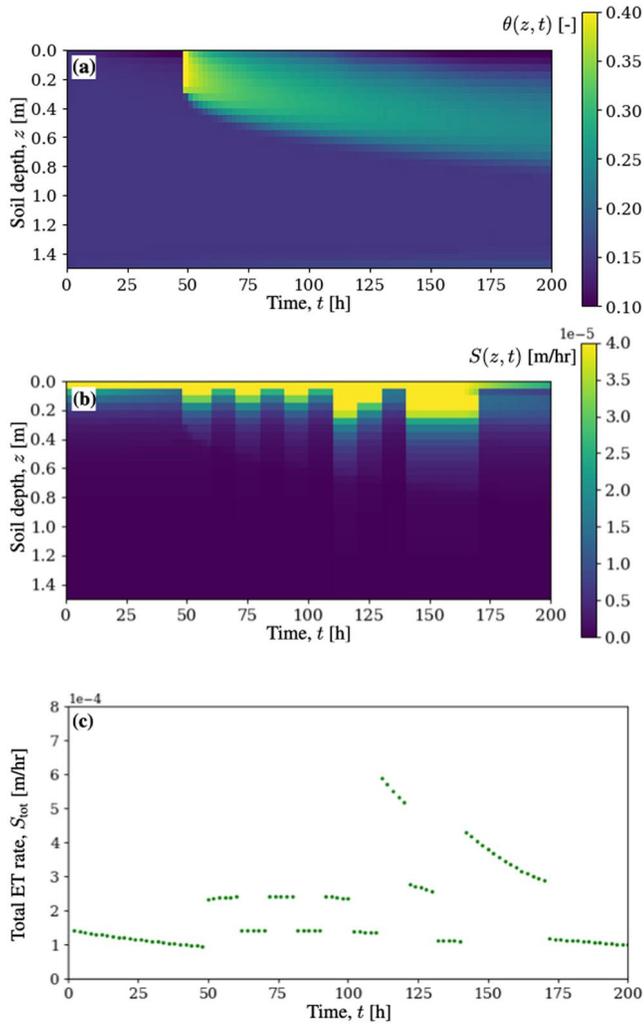


Figure 1. Ground truth values of (a) water content $\theta(z, t)$, (b) evapotranspiration rate $S(z, t)$, and (c) total evapotranspiration rate $S_{\text{tot}}(t)$ obtained as a solution of Equations 2–11 with known input parameters.

the posterior statistics of $\theta(z, t)$ and then evaluates the posterior statistics of $E_{\text{tot}}(t)$ from Equation 12. The MLE implementation has a fixed learning rate $\alpha = 0.2$, and $\epsilon = 10^{-4}$, most of the iteration processes at different time steps would converge with an average of 380 iterations. The two implementations of EnKF yield the $S_{\text{tot}}(t)$ estimates that have a comparable relative bias b that is small ($|b| = 1.6\%$ and 0.4% for EnKF(S) and EnKF(θ), respectively) and appreciably smaller than that of the MLE ($|b| = 4.6\%$). By the other two metrics, R and RV , EnKF(S) slightly outperforms EnKF(θ), and both of them slightly outperform the MLE. Given the same prior information about S , the uncertainty bound for the EnKF(S) estimator is narrower than that for EnKF(θ) and is similar to that for the MLE. That is because EnKF(θ)'s reliance on the water balance method, Equation 12, adds up the uncertainties in the soil moisture estimators at two consecutive time steps. The differences in these performance metrics are relatively minor, and all three methods provide an arguably adequate estimation of the total ET rate $S_{\text{tot}}(t)$. It is therefore remarkable that our EnKF implementation, EnKF(S), does so two orders of magnitude faster than either EnKF(θ) or MLE(S).

$$b = \frac{\langle S_{\text{tot}}^{\text{est}} \rangle - \langle S_{\text{tot}} \rangle}{\langle S_{\text{tot}} \rangle}, \quad R = \frac{\text{cov}(S_{\text{tot}}^{\text{est}}, S_{\text{tot}})}{\sigma_{S_{\text{tot}}^{\text{est}}} \sigma_{S_{\text{tot}}}}, \quad RV = \frac{\sigma_{S_{\text{tot}}^{\text{est}}}}{\sigma_{S_{\text{tot}}}}, \quad (35)$$

where $\langle \cdot \rangle$, $\text{cov}(\cdot, \cdot)$, and σ indicate the ensemble mean, covariance, and SD, respectively. A small absolute b value, and R and RV close to 1 represent a good estimation.

The quality of estimation ($\tilde{\mathcal{V}}$) of spatially varying quantities $\mathcal{V}(z, t)$, that is, ET S and soil moisture θ , is ascertained in terms of the root mean square error (RMSE), $\mathcal{E}(t)$, and the total RMSE, \mathcal{E}_{tot} ,

$$\mathcal{E}(t_k) = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (\tilde{\mathcal{V}}_i^k - \mathcal{V}_i^k)^2},$$

$$\mathcal{E}_{\text{tot}} = \sqrt{\frac{1}{N_t N_s} \sum_{t_j=1}^{N_t} \sum_{i=1}^{N_s} (\tilde{\mathcal{V}}_i^k - \mathcal{V}_i^k)^2}, \quad \mathcal{V} = S \text{ or } \theta. \quad (36)$$

Here, $\tilde{\mathcal{V}}_i^k = \tilde{\mathcal{V}}(z_i, t_k)$ and $\mathcal{V}_i^k = \mathcal{V}(z_i, t_k)$ are the estimated and true values of the quantity \mathcal{V} , respectively.

We compare the performance of our Bayesian inference method, which is based on treating the source term $S(\cdot, z, t)$ as an observable, with the original approach of Guderle and Hildebrandt (2015) that treats the state variable $\theta(z, t)$ as such and then computes the statistics of the quantities of interest at the post-processing stage. To distinguish between these two implementation of EnKF, we refer to ours as EnKF(S) and to the original as EnKF(θ); the MLE of S is denoted by MLE(S). The prior statistics of the S parameters for EnKF(θ) and EnKF(S) are collated in Table 1.

4.3. Total Evapotranspiration Rate

Figure 2 shows estimates of the total ET rate, $S_{\text{tot}}^{\text{est}}(t)$, obtained via the MLE, EnKF(S), and EnKF(θ). All three estimates are accompanied by the confidence intervals defined as $\pm \sigma_{S_{\text{tot}}^{\text{est}}}(t)$. The two implementations of EnKF start with the same prior, contain 200 ensemble members, and estimate $S_{\text{tot}}(t)$ in terms of its posterior mean. EnKF(θ) first computes

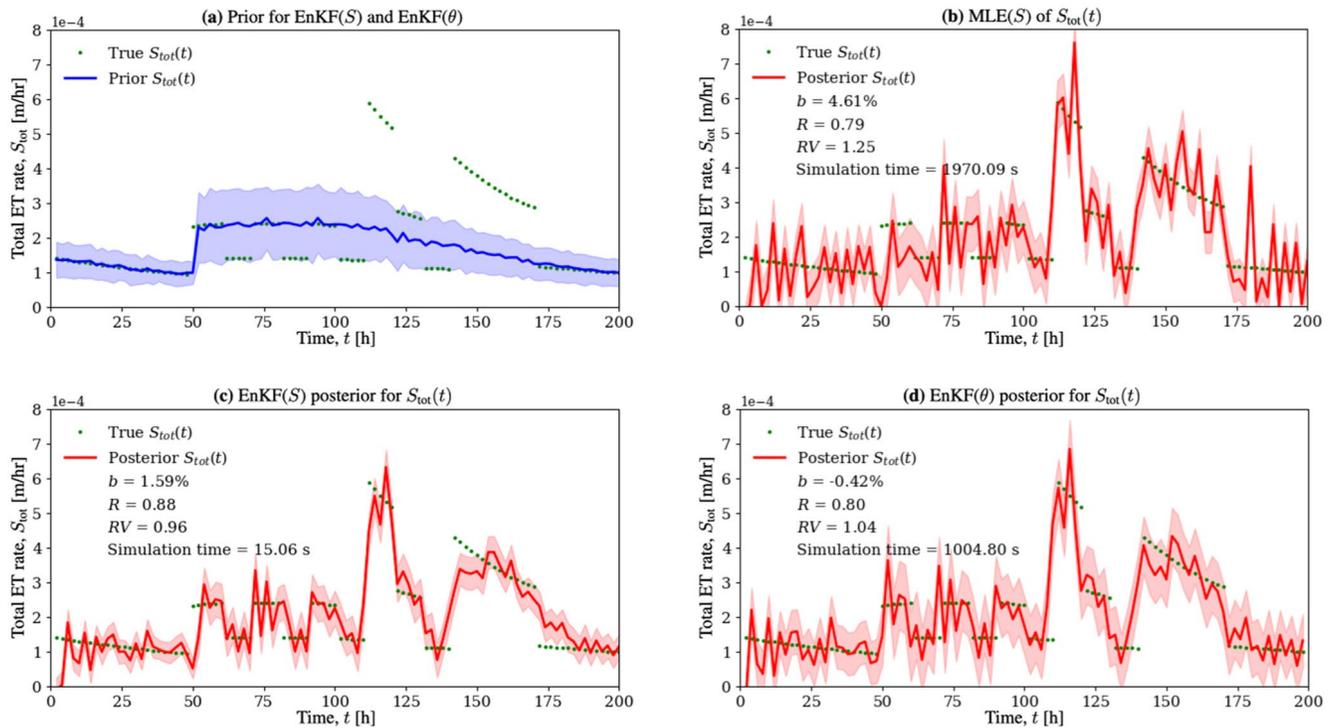


Figure 2. (a) Initial guess and estimation of the total evapotranspiration rate, $S_{tot}(t)$, via (b) MLE, (c) EnKF(S), and (d) EnKF(θ), from an array of 8 moisture sensors. Whereas the initial guess serves as a prior for both implementations of EnKF, it is not used by MLE. The EnKF estimates of $S_{tot}(t)$ are given by the posterior means (solid lines). All three estimates are accompanied by the confidence intervals (shaded regions) defined as \pm SD.

4.4. Spatial Distribution of Root Water Uptake

By construction, the EnKF(θ) method cannot infer the spatial variability of root water uptake, that is, the source term $S(z, t)$, from soil moisture data. The remaining two methods, EnKF(S) and MLE(S), can be used for that task (Figure 3). Visual comparison of the spatiotemporal maps of $S(z, t)$ predicted by EnKF(S) and MLE(S) with the ground truth in Figure 1 shows the two methods' adequacy. Figure 3 also provides a quantitative assessment of their performance. At all times t , the RMSE $\mathcal{E}(t)$ of EnKF(S) is smaller than the RMSE of MLE, with the total RMSE \mathcal{E}_{tot} of EnKF(S) being half that of MLE(S).

4.5. Soil Water Content

While EnKF(θ) corrects the model's prediction of soil water content $\theta(z, t)$ in its Bayesian update, EnKF(S) and MLE(S) do so indirectly by updating the sink term $S(z, t)$ instead. One, therefore, would expect EnKF(θ) to yield the most accurate estimates of $\theta(z, t)$, provided the soil moisture data used for the update are sufficiently accurate. The predicted temporal evolution of $\theta(z, t)$ at soil depths where moisture sensors are not available ($z = 12.5, 27.5, \text{ and } 87.5$ cm) demonstrates that to be the case (Figure 4). At all times, the RMSE of the EnKF(θ)-based prediction of soil water content, $\mathcal{E}(t)$, is slightly smaller than that of its EnKF(θ)-based counterpart. RMSEs of both EnKF implementations are smaller than that of MLE. The total RMSEs, \mathcal{E}_{tot} , for all of three parameter-estimation methods are of the same order of magnitude, indicating their ability to predict soil water content $\theta(z, t)$ at depths where measurements are not available.

4.6. Computational Cost and Convergence Rate

The relative performance of EnKF(S) and EnKF(θ), in terms of their computational efficiency and convergence rate, is summarized in Table 2. The prediction accuracy is quantified by the relative bias b , correlation coefficient R , and relative variability RV in the predictions of total ET rate $S_{tot}(t)$, and by the total RMSE \mathcal{E}_{tot}

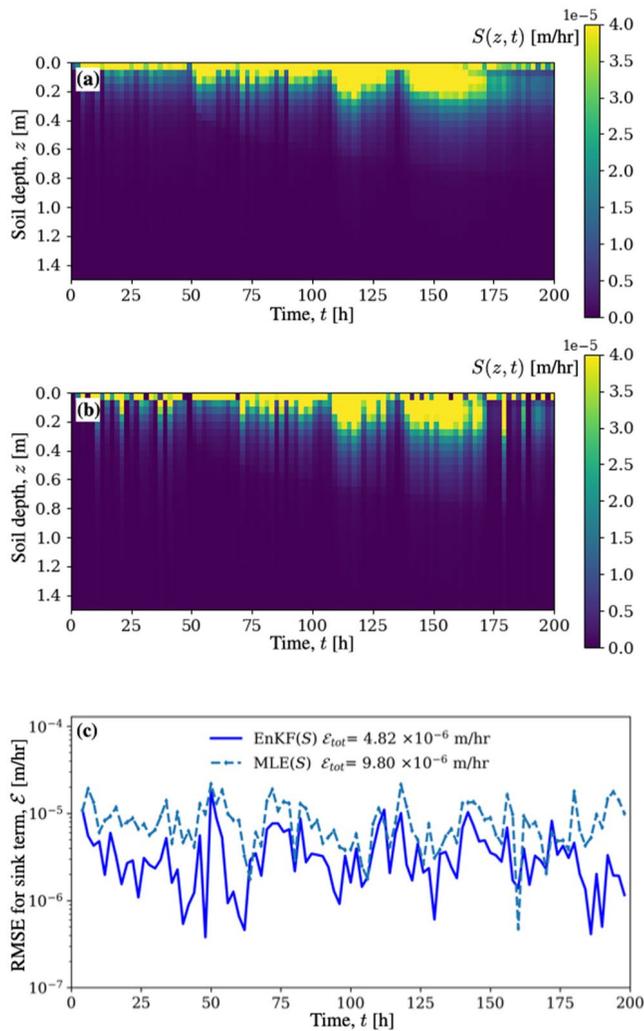


Figure 3. Spatiotemporal maps of root water uptake, $S(z, t)$, estimated with either (a) EnKF(S) or (b) MLE(S); and (c) their respective root mean square errors $\mathcal{E}(t)$ and \mathcal{E}_{tot} .

5. Conclusions

We proposed two parameter-estimation methods to infer ET rates and root water uptake profiles from soil-moisture sensor-array data. One, denoted by EnKF(S), is our implementation of ensemble Kalman filter (EnKF); it treats the ET sink term $S(\theta, z, t)$ in the Richards equation, rather than soil moisture $\theta(z, t)$, as the observable to update. The other is a maximum likelihood estimator (MLE) applied to the same observable and, hence, denoted by MLE(S); it is supplemented with the Fisher information matrix to quantify uncertainty in its predictions. We used numerical experiments to ascertain the ability of our methods to infer total ET rate $S_{tot}(t)$ and root water uptake profile $S(z, t)$ from soil moisture data collected by a sensor array. The performance of our methods was compared, in terms of accuracy and computational efficiency, to that of the original EnKF with water-content observable, EnKF(θ). Our analysis leads to the following major conclusions.

of the predictions of soil moisture $\theta(z, t)$. Both EnKF implementations use, alternatively, ensemble size $N_{MCS} = 200, 400, 600,$ and $1,000$. As expected, the simulation time of both model increases as $\mathcal{O}(N_{MCS})$, with EnKF(S) being two orders of magnitude faster than EnKF(θ). The error metrics for both methods are relatively insensitive to the ensemble size considered, suggesting that the ensemble size of $N_{MCS} = 200$ is sufficient to achieve convergence. By all the metrics considered, EnKF(θ) is slightly more accurate than EnKF(S); that is likely due to additional approximations underpinning EnKF(S), for example, leading to the derivation of Equation 13.

Being a quintessential deterministic procedure, MLE(S) does not rely on ensemble of realizations; its convergence refers to the iterative solution of the underlying minimization problem. Figure 5 exhibits the loss function value as function of the number of iterations for several observation times, and the number of iterations at convergence (i.e., when the loss function reaches its minimum within prescribed tolerance) as function of observation time. These results demonstrate that our MLE(S) procedure with learning rate $\alpha = 0.2$ requires on average 380 iterations to converge at all-time steps. Hence, the MLE(S) is a viable tool for estimation of the root water uptake profiles and evaporation rates when the prior knowledge of the parameters $T_{max}(t)$ and $E_{max}(t)$ is elusive.

4.7. Impact of the Number of Sensors

Finally, we investigate the estimation accuracy and predictive uncertainty of the three methods when the number of moisture sensors N_{obs} is reduced from 8 to 5. The setting of our numerical experiments remains the same, except for the number and locations of sensors. These sensors are now placed at depth $z_1 = 2.5$ cm, $z_2 = 7.5$ cm, $z_3 = 17.5$ cm, $z_4 = 52.5$ cm, and $z_5 = 97.5$ cm. All three parameter estimation methods enable one to infer the total ET rate, $S_{tot}(t)$, from the observations collected by this sensor array (Figure 6). These estimates of $S_{tot}(t)$ are less accurate than those obtained from the eight-sensor array (Figure 2), although not by as much as one could have expected. The largest impact of the reduction in the number of sensors is on the estimation uncertainty of EnKF(θ). Its confidence interval, defined as $\pm SD$, becomes much wider than those of EnKF(S) and MLE(S) when N_{obs} decreases from 8 to 5. This finding implies that our two new methods are advantageous in field applications wherein soil moisture data are scarce.

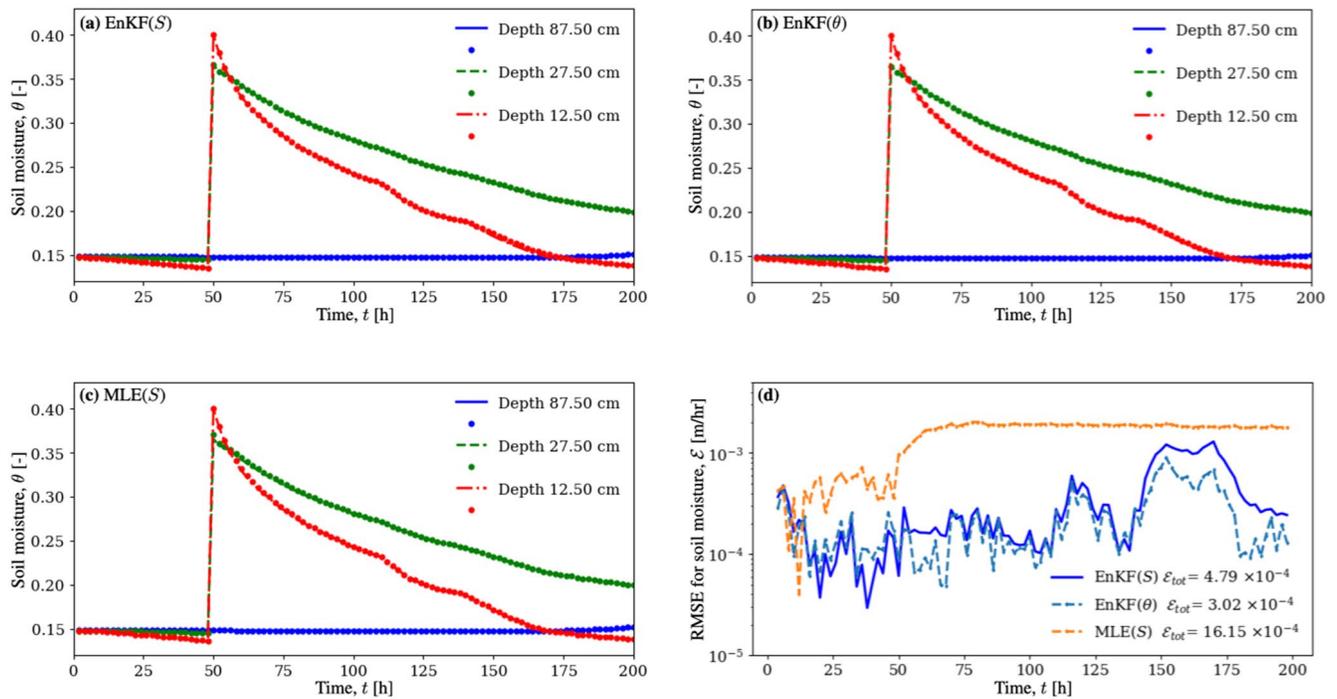


Figure 4. Time evolution of soil moisture at several depths predicted by (a) EnKF(S), (b) EnKF(θ), and (c) MLE(S); and (d) their respective RMSEs. Lines: ground truth; Dots: predictions.

1. All three methods provide accurate estimates of $S_{\text{tot}}(t)$ and $\theta(z, t)$. They can also capture the temporal dynamics and sudden changes in $S_{\text{tot}}(t)$ that can be attributed to diurnal cycles or plant mortality.
2. Our EnKF implementation, EnKF(S), does so two orders of magnitude faster than either EnKF(θ) or MLE(S).
3. The original EnKF implementation, EnKF(θ), cannot be used to infer $S(z, t)$ from soil moisture data, while EnKF(S) and MLE(S) can. EnKF(S) is about twice more accurate and two orders of magnitude faster than MLE(S) in performing this task.
4. The error metrics for both EnKF implementations are relatively insensitive to the number of forward solves of the Richards equation, suggesting that the ensemble size of $N_{\text{MCS}} = 200$ realizations is sufficient to achieve convergence.
5. Our MLE(S) procedure requires an order of magnitude fewer iterations to converge than its counterpart applied to $\theta(z, t)$. That renders MLE(S) a viable and practical tool for estimation of the root water uptake profiles and evaporation rates when the prior knowledge about ET at the site is elusive.

Table 2

Accuracy (Relative Bias b , Correlation Coefficient R , and Relative Variability RV in the Predictions of Total Evapotranspiration Rate S_{tot} ; and Total RMSE \mathcal{E}_{tot} of the Predictions of Soil Moisture θ) and Run Time of EnKF(S) and EnKF(θ) With Ensemble Size N_{MCS}

N_{MCS}	EnKF(S)					EnKF(θ)				
	b	R	RV	\mathcal{E}_{tot}	T_{run}	b	R	RV	\mathcal{E}_{tot}	T_{run}
200	1.59	0.88	0.96	4.79×10^{-4}	15s	-0.42	0.8	1.04	3.02×10^{-4}	1005s
400	1.43	0.89	0.96	4.82×10^{-4}	16.5s	-0.58	0.8	1.04	2.94×10^{-4}	2009s
600	1.50	0.89	0.96	4.76×10^{-4}	18s	-0.57	0.8	1.04	2.93×10^{-4}	3014s
1,000	1.53	0.89	0.96	4.78×10^{-4}	20.5s	-0.50	0.81	1.04	2.88×10^{-4}	5056s

Note. The relative bias b is reported in %, and the run time T_{run} in seconds.

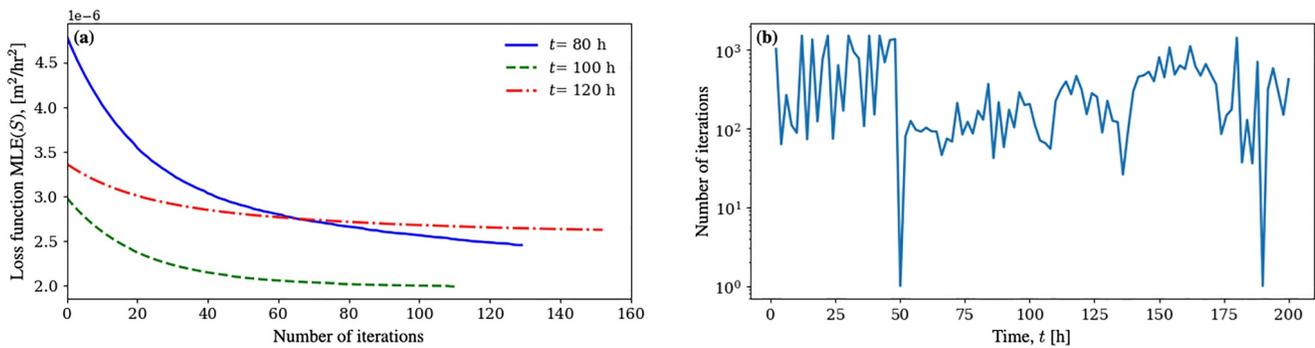


Figure 5. (a) Dependence of the loss function in the MLE(S) on the number of iterations at several times t , and (b) the number of iteration it takes the MLE(S) to convergence as function of time t .

- By focusing on the direct estimation of ET rather than its indirect estimation via soil moisture, our implementation of both EnKF and MLE is advantageous in field applications wherein soil moisture data are scarce. This finding suggests the use of these techniques in optimal design of sensor networks.

The conclusions reported above come with an important caveat. They are based on the assumption of unidirectional (vertical) flow in a homogeneous soil. While this assumption is often used in the inverse analyses of this kind, it represents a significant simplification of the vadose zone. One computationally efficient way of dealing with soil heterogeneity in the present context is to treat the soil as a collection of one-dimensional isolated flow tubes (Sinsbeck & Tartakovsky, 2015; Wang & Tartakovsky, 2011), in a manner consistent with the Dagan-Bresler parameterization (Dagan & Bresler, 1983). In addition, we assume the soil hydrologic parameters to be known with certainty. In real cases, one would estimate them

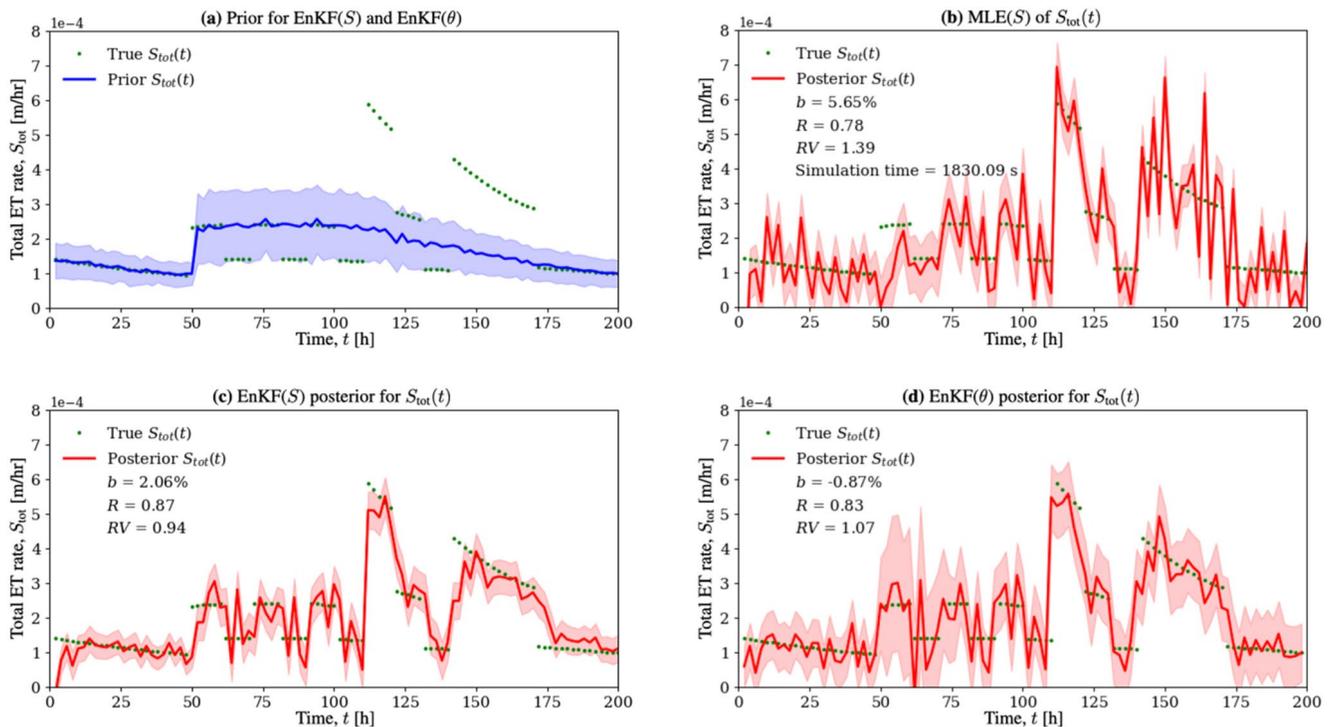


Figure 6. (a) Initial guess and estimation of the total evapotranspiration rate, $S_{tot}(t)$, via (b) MLE, (c) EnKF(S), and (d) EnKF(θ), from an array of five moisture sensors. Whereas the initial guess serves as a prior for both implementations of EnKF, it is not used by MLE. The EnKF estimates of $S_{tot}(t)$ are given by the posterior means (solid lines). All three estimates are accompanied by the confidence intervals (shaded regions) defined as \pm SD.

from infiltration tests in advance. Our data assimilation methods can be extended to include parametric uncertainty. We leave this and other extensions of our methods for follow-up studies.

Data Availability Statement

There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

Acknowledgments

Weiyu Li and Daniel M. Tartakovsky were supported in part by National Science Foundation Grant EAR-2100927, by UPS Endowment Fund at Stanford, and by a gift from Total. Baptiste Dafflon, Haruko M. Wainwright, Yuxin Wu, and Qina Yan, were funded by the U.S. Department of Energy, BER Contract DE-AC020SCH11231 to the LBNL Watershed Function Scientific Focus Area. Roelof Versteeg and Haiyan Zhou were funded by SBIR Award DE-SC0018447 to Subsurface Insights.

References

- Bausner, H. H., Berg, D., & Roth, K. (2021). Technical note: Sequential ensemble data assimilation in convergent and divergent systems. *Hydrology and Earth System Sciences*, 25, 3319–3329. <https://doi.org/10.5194/hess-25-3319-2021>
- Bocquet, M., Raanes, P. N., & Hannart, A. (2015). Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 22, 645–662. <https://doi.org/10.5194/npg-22-645-2015>
- Boso, F., & Tartakovsky, D. M. (2020). Learning on dynamic statistical manifolds. *Proceedings of the Royal Society A*, 476(2239), 20200213. <https://doi.org/10.1098/rspa.2020.0213>
- Breña Naranjo, J. A., Weiler, M., & Stahl, K. (2011). Sensitivity of a data-driven soil water balance model to estimate summer evapotranspiration along a forest chronosequence. *Hydrology and Earth System Sciences*, 15(11), 3461–3473. <https://doi.org/10.5194/hess-15-3461-2011>
- Carter, C., & Liang, S. (2019). Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 78, 86–92. <https://doi.org/10.1016/j.jag.2019.01.020>
- Chen, J. M., & Liu, J. (2020). Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sensing of Environment*, 237, 111594. <https://doi.org/10.1016/j.rse.2019.111594>
- Dagan, G., & Bresler, E. (1983). Unsaturated flow in spatially variable fields: 1. Derivation of models of infiltration and redistribution. *Water Resources Research*, 19(2), 413–420. <https://doi.org/10.1029/wr019i002p00413>
- Dezsi, S., Mindrescu, M., Petrea, D., Kumar Rai, P., Hamann, A., & Nistor, M.-M. (2018). High-resolution projections of evapotranspiration and water availability for Europe under climate change. *International Journal of Climatology*, 38(10), 3832–3841. <https://doi.org/10.1002/joc.5537>
- Doble, R. C., & Crosbie, R. S. (2017). Review: Current and emerging methods for catchment-scale modelling of recharge and evapotranspiration from shallow groundwater. *Hydrogeology Journal*, 25, 3–23. <https://doi.org/10.1007/s10040-016-1470-3>
- Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 68(3), 467–487. <https://doi.org/10.1093/biomet/65.3.457>
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162. <https://doi.org/10.1029/94jc00572>
- Fath, B. (2018). *Encyclopedia of ecology* (Vol. 105, No. 3). Elsevier.
- Fisher, J. B., Whittaker, R. J., & Malhi, Y. (2011). ET come home: Potential evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, 20(1), 1–18. <https://doi.org/10.1111/j.1466-8238.2010.00578.x>
- Green, S. R., & Clothier, B. E. (1995). Root water uptake by kiwifruit vines following partial wetting of the root zone. *Plant and Soil*, 173(2), 317–328. <https://doi.org/10.1007/bf00011470>
- Guderle, M., & Hildebrandt, A. (2015). Using measured soil water contents to estimate evapotranspiration and root water uptake profiles—A comparative study. *Hydrology and Earth System Sciences*, 19, 409–425. <https://doi.org/10.5194/hess-19-409-2015>
- Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., & Kling, H. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Guswa, A. J., Celia, M. A., & Rodriguez-Iturbe, I. (2002). Models of soil moisture dynamics in ecohydrology: A comparative study. *Water Resources Research*, 38(9), 5. <https://doi.org/10.1029/2001wr000826>
- Hale, M. G., & Orcutt, D. M. (1987). *The physiology of plants under stress*. John Wiley & Sons.
- Herman, M. R., Nejadhashemi, A. P., Abouali, M., Hernandez-Suarez, J. S., Daneshvar, F., Zhang, Z., et al. (2018). Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *Journal of Hydrology*, 556, 39–49. <https://doi.org/10.1016/j.jhydrol.2017.11.009>
- Hupet, F., Lambot, S., Javaux, M., & Vanclooster, M. (2002). On the identification of macroscopic root water uptake parameters from soil water content observations. *Water Resources Research*, 38(12), 36. <https://doi.org/10.1029/2002wr001556>
- Jackisch, C., Germer, K., Graeff, T., Andrä, I., Schulz, K., Schiedung, M., et al. (2020). Soil moisture and matric potential—an open field comparison of sensor systems. *Earth System Science Data*, 12(1), 683–697. <https://doi.org/10.5194/essd-12-683-2020>
- Jakeman, A. J., Barreteau, O., Hunt, R. J., Rinaudo, J.-D., & Ross, A. (Eds.). (2016). *Integrated groundwater management*. Springer Nature. <https://doi.org/10.1007/978-3-319-23576-9>
- Kaur Saggi, M., & Jain, S. (2020). Application of fuzzy-genetic and regularization random forest (FG-RRF): Estimation of crop evapotranspiration (ETc) for maize and wheat crops. *Agricultural Water Management*, 229, 105907. <https://doi.org/10.1016/j.agwat.2019.105907>
- Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). Springer.
- Loáiciga, H. A. (2017). The safe yield and climatic variability: Implications for groundwater management. *Groundwater*, 171(3), 334–345. <https://doi.org/10.1111/gwat.12481>
- Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, 50(2), 1131–1151. <https://doi.org/10.1002/2013wr014581>
- Lu, Z., Neuman, S. P., Guadagnini, A., & Tartakovsky, D. M. (2002). Conditional moment analysis of steady state unsaturated flow in bounded, randomly heterogeneous soils. *Water Resources Research*, 38(4), 9. <https://doi.org/10.1029/2001WR000278>
- Miller, D., & Aarstad, J. (1973). Effective available water and its relation to evapotranspiration rate, depth of wetting, and soil texture. *Soil Science Society of America Journal*, 37(5), 763–766. <https://doi.org/10.2136/sssaj1973.03615995003700050036x>
- Pan, M., & Wood, E. F. (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *Journal of Hydrometeorology*, 7(3), 534–547. <https://doi.org/10.1175/jhm495.1>

- Perrochet, P. (1987). Water uptake by plant roots—A simulation model. I. Conceptual model. *Journal of Hydrology*, 95(1–2), 55–61. [https://doi.org/10.1016/0022-1694\(87\)90115-6](https://doi.org/10.1016/0022-1694(87)90115-6)
- Porporato, A., D'Odorico, P., Laio, F., & Rodriguez-Iturbe, I. (2003). Hydrologic controls on soil carbon and nitrogen cycles. i. modeling scheme. *Advances in Water Resources*, 26(1), 45–58. [https://doi.org/10.1016/s0309-1708\(02\)00094-5](https://doi.org/10.1016/s0309-1708(02)00094-5)
- Protim Goswami, M., Montazer, B., & Sarma, U. (2019). Design and characterization of a fringing field capacitive soil moisture sensor. *IEEE Transactions on Instrumentation and Measurement*, 68(3), 913–922. <https://doi.org/10.1109/TIM.2018.2855538>
- Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, 44(3). <https://doi.org/10.1029/2007wr006357>
- Schenk, H. J., & Jackson, R. B. (2002). The global biogeography of roots. *Ecological Monographs*, 72(3), 311–28. [https://doi.org/10.1890/0012-9615\(2002\)072\[0311:tgbor\]2.0.co;2](https://doi.org/10.1890/0012-9615(2002)072[0311:tgbor]2.0.co;2)
- Severino, G., & Tartakovsky, D. M. (2015). A boundary-layer solution for flow at the soil-root interface. *Journal of Mathematical Biology*, 70(7), 1645–1668. <https://doi.org/10.1007/s00285-014-0813-8>
- Sinsbeck, M., & Tartakovsky, D. M. (2015). Impact of data assimilation on cost-accuracy tradeoff in multifidelity models. *SIAM/ASA Journal of Uncertainty Quantification*, 3(1), 954–968. <https://doi.org/10.1137/141001743>
- Talsma, C. J., Good, S. P., Jimenez, C., Martens, B., Fisher, J. B., Miralles, D. G., et al. (2018). Partitioning of evapotranspiration in remote sensing-based models. *Agricultural and Forest Meteorology*, 237, 131–143. <https://doi.org/10.1016/j.agrformet.2018.05.010>
- Tartakovsky, D. M., Guadagnini, A., & Riva, M. (2003). Stochastic averaging of nonlinear flows in heterogeneous porous media. *Journal of Fluid Mechanics*, 492, 47–62. <https://doi.org/10.1017/S002211200300538X>
- Tartakovsky, D. M., Lu, Z., Guadagnini, A., & Tartakovsky, A. M. (2003). Unsaturated flow in heterogeneous soils with spatially distributed uncertain hydraulic parameters. *Journal of Hydrology*, 275(3–4), 182–193. [https://doi.org/10.1016/s0022-1694\(03\)00042-8](https://doi.org/10.1016/s0022-1694(03)00042-8)
- Torres, A. B. B., da Rocha, A. R., Coelho da Silva, T. L., de Souza, J. N., & Gondim, R. S. (2020). Multilevel data fusion for the internet of things in smart agriculture. *Computers and Electronics in Agriculture*, 171, 105309. <https://doi.org/10.1016/j.compag.2020.105309>
- Wang, P., & Tartakovsky, D. M. (2011). Reduced complexity models for probabilistic forecasting of infiltration rates. *Advances in Water Resources*, 34, 375–382. <https://doi.org/10.1016/j.advwatres.2010.12.007>
- Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., & van Griensven, A. (2020). Can we trust remote sensing evapotranspiration products over Africa? *Hydrology and Earth System Sciences*, 24(3), 1565–1586. <https://doi.org/10.5194/hess-24-1565-2020>
- Wilson, K. B., Hanson, P. J., Mulholland, P. J., Baldocchi, D. D., & Wullschleger, S. D. (2001). A comparison of methods for determining forest evapotranspiration and its components: Sap-flow, soil water budget, eddy covariance and catchment water balance. *Agricultural and Forest Meteorology*, 106(2), 153–168. [https://doi.org/10.1016/s0168-1923\(00\)00199-4](https://doi.org/10.1016/s0168-1923(00)00199-4)
- Xaver, A., Zappa, L., Rab, G., Pfeil, I., Vreugdenhil, M., Hemment, D., & Dorigo, W. A. (2020). Evaluating the suitability of the consumer low-cost Parrot Flower Power soil moisture sensor for scientific environmental applications. *Geoscientific Instrumentation, Methods and Data Systems*, 9(1), 117–139. <https://doi.org/10.5194/gi-9-117-2020>
- Yan, Q., Le, P. V., Woo, D. K., Hou, T., Filley, T., & Kumar, P. (2019). Three-dimensional modeling of the coevolution of landscape and soil organic carbon. *Water Resources Research*, 55(2), 1218–1241. <https://doi.org/10.1029/2018wr023634>
- Yang, L., Wang, P., & Tartakovsky, D. M. (2020). Resource-constrained model selection for uncertainty propagation and data assimilation. *SIAM/ASA Journal on Uncertainty Quantification*, 8(3), 1118–1138. <https://doi.org/10.1137/19M1263376>
- Zuo, Q., Meng, L., & Zhang, R. (2004). Simulating soil water flow with root-water-uptake applying an inverse method. *Soil Science*, 169(1), 13–24. <https://doi.org/10.1097/01.ss.0000112018.97541.85>