



An improved Kalman filtering approach for the estimation of unsaturated flow parameters by assimilating photographic imaging data

Mohammad Mahdi Rajabi, Benjamin Belfort, François Lehmann, Sylvain Weill, Behzad Ataie-Ashtiani, Marwan Fahs

► To cite this version:

Mohammad Mahdi Rajabi, Benjamin Belfort, François Lehmann, Sylvain Weill, Behzad Ataie-Ashtiani, et al.. An improved Kalman filtering approach for the estimation of unsaturated flow parameters by assimilating photographic imaging data. Journal of Hydrology, 2020, 590, pp.125373. 10.1016/j.jhydrol.2020.125373 . hal-03452430

HAL Id: hal-03452430

<https://hal.science/hal-03452430>

Submitted on 26 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proofs

Research papers

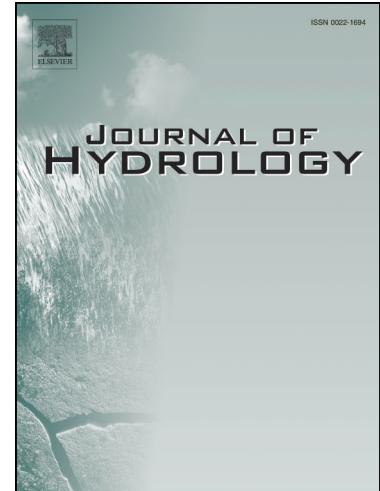
An Improved Kalman Filtering Approach for the Estimation of Unsaturated Flow Parameters by Assimilating Photographic Imaging Data

Mohammad Mahdi Rajabi, Benjamin Belfort, François Lehmann, Sylvain Weill, Behzad Ataie-Ashtiani, Marwan Fahs

PII: S0022-1694(20)30833-7
DOI: <https://doi.org/10.1016/j.jhydrol.2020.125373>
Reference: HYDROL 125373

To appear in: *Journal of Hydrology*

Received Date: 6 May 2020
Revised Date: 28 June 2020
Accepted Date: 30 July 2020



Please cite this article as: Mahdi Rajabi, M., Belfort, B., Lehmann, F., Weill, S., Ataie-Ashtiani, B., Fahs, M., An Improved Kalman Filtering Approach for the Estimation of Unsaturated Flow Parameters by Assimilating Photographic Imaging Data, *Journal of Hydrology* (2020), doi: <https://doi.org/10.1016/j.jhydrol.2020.125373>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Research Article

An Improved Kalman Filtering Approach for the Estimation of Unsaturated Flow Parameters by Assimilating Photographic Imaging Data

Mohammad Mahdi Rajabi^{1, *}, Benjamin Belfort², François Lehmann², Sylvain Weill²,
Behzad Ataie-Ashtiani^{3, 4}, Marwan Fahs²

¹ Civil and Environmental Engineering Faculty, Tarbiat Modares University, PO Box 14115-397, Tehran, Iran.

² Laboratoire d'Hydrologie et Géochimie de Strasbourg, Université de Strasbourg/EOST/ENGEEES, CNRS, 1 rue Blessig 67084 Strasbourg, France.

³ Department of Civil Engineering, Sharif University of Technology, PO Box 11155-9313, Tehran, Iran.

⁴ National Centre for Groundwater Research & Training, College of Science & Engineering, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia

First Submitted: April 2020

Revision Submitted: June 2020

1. Introduction

Modeling of soil moisture dynamics is of key importance in characterizing the processes occurring in the vadose zone and the interaction of the vadose zone with the underlying groundwater and the overlaying land surface, vegetation and atmosphere. For instance, soil moisture simulations play an important role in the prediction of groundwater recharge, and the understanding of how contaminant from the land surface or shallow subsurface migrate to the groundwater. In hydrology and climatology, these simulations are used in predicting the partitioning of water and energy at the land surface, and in ecology and agriculture, such models are employed in forecasting the availability of water to the root zone. As a result, numerical models for soil moisture simulation have received increasing attention in the last several decades.

The majority of currently available unsaturated flow models rely on the Richards' equation in conjunction with the van Genuchten-Mualem model (van Genuchten, 1980) to describe soil moisture dynamics (Vereecken et al., 2008). It is well-known and well-documented that the ability of such models in providing realistic predictions is highly dependent on the accurate estimation of parameters related to the soil water retention curve and the hydraulic conductivity function (Li and Ren, 2011; Moret-Fernández et al., 2017; Younes et al., 2018). These parameters often exhibit high spatial variability, and their direct measurement is infeasible. Hence they are often either inferred from soil particle size distributions using pedotransfer functions (e.g. Schaap et al., 2001; Van Looy et al., 2017; Zhang and Schaap, 2017), or estimated through inverse modeling by employing soil moisture data. The latter has been the subject of numerous studies with some promising results. Because of the nonlinear relationship that exists between the inputs and outputs of soil moisture simulation models, and in the heterogeneous case, the high-dimensionality of their parameter space, certain inverse methods have been of particular interest in this context. These

include frequentist methods such as least square estimation (e.g. [Abbaspour et al., 2001](#)) and maximum likelihood estimation (e.g. [Hollenbeck and Jensen, 1998](#)); batch Bayesian methods such as Markov Chain Monte Carlo ([Scharnagl et al., 2011](#)); and sequential Bayesian methods, including most notably the extended Kalman filter (e.g. [Lu et al., 2011](#)), the ensemble Kalman filter (EnKF) (e.g. [Zhu et al., 2017](#); [Wang et al. 2018](#)), and the particle filter ([Montzka et al., 2011](#)).

The surge in the use of inverse modeling methods for the estimation of unsaturated soil hydraulic parameters is partly due to increasing availability of soil moisture data. During the past 30 years, there has been significant advances in the development of different soil moisture measurement techniques ranging from the point-scale to the continental-scale ([Crow and Zhan, 2007](#)). These include techniques based on dielectric characterization of the soil (e.g. time domain reflectometry and frequency domain reflectometry), hydro-geophysical methods (e.g. ground penetrating radars (GPR)), and air-borne/space-borne remote sensing (RS) methods (e.g. passive microwave, synthetic aperture radars and scatterometers) ([Fang and Lakshmi, 2014](#); [Su et al., 2014](#); [Tran et al., 2014](#)).

Despite the abundance of measurement techniques, accurate and non-invasive characterization of the spatiotemporal variations of soil moisture is still considered challenging ([Klotzsche et al., 2018](#)). In this respect, photographic imaging techniques offer some interesting potentials, enabling mapping of the water content at very fine resolutions in both space and time ([Tidwell and Glass, 1994](#)). Examples of photographic imaging techniques for soil moisture mapping include the light transmission method ([Hoa, 1981](#)), the multispectral image analysis method ([Kechavarzi et al., 2000](#)) and the light reflection method ([Yoshimoto et al. 2011](#)). All these methods incorporate some sort of image processing to convert the reflected or transmitted light intensities into soil moisture content. Tracers are commonly employed to improve visualization. Recently, [Belfort et al \(2017b\)](#)

and [Belfort et al. \(2019\)](#) proposed a novel imaging technique which combines the analysis of photographic images and direct measurements, to characterize a two-dimensional (2D) field without the need for a tracer. The wealth of soil moisture data provided by such photographic imaging techniques makes them particularly appealing for the estimation of distributed fields of unsaturated soil hydraulic parameters, especially in lab-scale experiments.

This paper proposes a novel methodology for parameter estimation, in the context of variably saturated flow in porous media, which is based on an improved EnKF method and takes into account information obtained from a recent photographic technique in terms of water content. In the next two introductory sub-sections, these different aspects are put into context and our objectives are clearly stated.

1.1. Estimating unsaturated flow parameters using photographic imaging

A suitable method capable of estimating unsaturated flow parameters through the use of photographic imaging data must be able to incorporate finely discretized measurements provided by the images to gradually reduce parameter estimation uncertainty and distinguishably quantify individual components of uncertainty such as model structural uncertainties and imaging data uncertainties. The EnKF has both characteristics ([Reichle et al., 2008](#); [Rajabi et al., 2018](#)), and is hence a natural choice for the estimation of unsaturated soil parameters using imaging data. Furthermore, the EnKF does not require storage of past data, which is a bonus with respect to the potentially high volume of data produced by imaging. The EnKF has been used in the state estimation context for assimilation of soil moisture data in numerous studies since the early 2000s (e.g. [Reichle et al., 2002](#)). Estimation of unsaturated flow parameters using the EnKF, often done in conjunction with state estimation, is more recent and sparse, and we provide a review of some key studies in **Table 1**. The majority of studies reviewed in this table deal with ‘hypothetical’ one-

dimensional (1D) column or 2D rectangular problems, with a few other studies focusing on 1D ‘field experiments’ (e.g. [Jiang et al., 2019](#)). Previous studies have employed RS ([Li et al., 2012](#)), GPR ([Tran et al., 2014](#)) and contact-based, point-scale soil moisture data for the estimation of unsaturated flow parameters using the EnKF. However, to the author’s knowledge, no attempt has been previously made in using the EnKF for assimilation of photographic images in the unsaturated zone.

Note that photographic imaging data have some unique characteristics compared to data obtained from contact-based, GPR or RS techniques, which affect the parameter estimation process. Contact-based techniques provide point-scale data whereas photographic imaging can be used to characterize the entire soil moisture distribution. GPR is considered more appropriate for intermittent rather than continuous measurement of the soil moisture profile ([Vereecken et al., 2008](#)), and RS is suitable for low resolution measurements of surface soil moisture across large areas ([Lu et al., 2011](#)). GPR, which has been successfully applied under laboratory conditions, allows indirect visualization of water flow in two and three dimensions, but data analysis as well as the interpretation remain a challenging topic ([Allroggen et al., 2015](#)). Comparatively, photography provides detailed 2D structural and phenomenological information with high spatial and temporal resolutions, well suited for characterizing experiments at the lab-scale when sections of porous media are directly accessible. This feature is important because the estimation of heterogeneous fields of unsaturated soil parameters from sparse data obtained from other methods, results in highly underdetermined problems which are prone to ill-posedness and non-uniqueness. The affluence of data provided by photographic imaging has the potential to solve this problem. However, imaging data is highly susceptible to noise resulting from brightness fluctuations and reflections from the surroundings, and is also influenced by the color space used in the analysis

(Kashuk et al., 2014). Therefore, parameter estimation based on photographic imaging data is vulnerable to instability.

Based on the above notion, this study tends to extend previous researches by: (1) proposing a novel framework for the estimation of unsaturated soil hydraulic parameters using photographic imaging data, (2) presenting a customized version of the EnKF capable of handling potentially noisy image data, and (3) validating the proposed methodology using lab-scale data. Data obtained from the novel imaging technique proposed by Belfort et al (2017b) is used as the basis for the validations. After accomplishing these objectives, we will try to answer a key question: how do the parameter estimation results obtained from using image data, compared with those acquired from using direct measurements of soil moisture?

1.2. Developing relationship between light intensities and water content

In photographic imaging, what is measured is the light intensities not the water content. Therefore, direct measurements of the water content are also required to develop or calibrate an equation that transforms light intensities to water content. This is a key component of almost every previous study on the use of photographic imaging techniques for soil moisture mapping, and previous studies often rely on statistical regression to develop such an equation. However, especially in lab-scale experiments, the properties of the soil may be known in advance. Hence, a simulated version of the moisture content distribution can also be generated, which is expected to closely resemble that of the real distribution. This gives rise to a new question: can we use prior information about the unsaturated flow parameters of the soil to calibrate an equation that relates light intensities to water content, without the need for direct measurements of the water content? In this paper, we will try to address this question, by developing a customized version of the EnKF that employs numerical model outputs to estimate such a relationship.

Table 1. Review of studies on the use of the EnKF for estimation of unsaturated flow parameters

Reference	Case study	Data	Estimated hydraulic parameters	KF algorithm	Numerical model
Li and Ren (2011)	Hypothetical 1D columns + field drainage experiment	Soil water pressure head	k, α	EnKF (aug.)	Hydrus-1D (Šimůnek et al., 2012)
Erdal et al. (2014)	Hypothetical 1D column, 1D field experiment	Point-scale θ	k_{sat}, α_{Gr}, e	EnKF (aug.)	FV approach
Song et al. (2014)	Hypothetical 2D 200×200 rectangle	Point-scale θ and h	Spatially variable k	Confirming, Restart, modified Restart EnKF	SWMS-2D code (Šimůnek et al., 1995)
Tran et al. (2014)	Hypothetical 1D column	Time-lapse GPR data of the θ profile (created artificially from model outputs)	k, α, n	MLEF (aug.)*	Hydrus-1D
Shi et al. (2015)	Hypothetical 180 cm column	Point-scale θ, h, GL	k, α, n	EnKF (aug.)	Hydrus-1D
Bauser et al. (2016)	1D field experiment	Point-scale θ	k, α, n	standard, closed-eye (aug.)	MuPhi (Ippisch et al., 2006)
Zhu et al. (2017)	Hypothetical $9000 \times 9000 \times 3$ m cuboid domain	Multi-scale near-surface (0–5 cm) θ observations	Spatially variable k	EnKF (aug.)	Model developed by Zhu et al. (2012)
Chaudhuri et al. (2018)	Hypothetical, 200×200×7 m cuboid domain	Point-scale θ observations at numerous locations	Spatially variable k, α, n	iterative EnKF	FEHM CVFE code (Zyvoloski, 2007)
Wang et al. (2018)	1D field experiment	Various θ observation scenarios for data-worth analysis	k, α, n	EnKF (aug.)	Code based on the Ross method (Ross, 2006)
Jiang et al. (2019)	1D field experiment	Point-scale θ and C	k_{sat}, α, SL	EnKF	Hydrus-1D
Man et al (2019)	Hypothetical 2D 100×200 cm rectangular domain	Point-scale C	Spatially variable k_{sat}	adaptive restart PCKF	---

θ : Soil moisture, h : soil water head, GL : groundwater level, C : solute concentration, k : Permeability, k_{sat} : saturated hydraulic conductivity, SL : solute longitudinal dispersivity
 α and n : shape parameters of the Mualem-van Genuchten model, e : porosity, CVFE: Control Volume Finite Element, FV: finite volume
PCKF: probabilistic collocation Kalman filter, aug.: augmentation method, MLEF: Maximum Likelihood Ensemble Filter
* Deterministic alternative to the KF

2. Theoretical framework

The main building blocks of the proposed approach for unsaturated flow parameter estimation are (1) a Richards' equation type unsaturated flow model, (2) a carefully chosen instance of the EnKF, and (3) generalized polynomial chaos expansions (PCEs) which are used for accelerating the EnKF computations. The theoretical framework of these three building blocks are described in this section.

2.1. Numerical simulation approach

Assuming that the media is rigid, air remains at atmospheric pressure, and the flow velocity is consistent with the Darcy-Buckingham's law; flow in a variably saturated porous media can be described using the mixed form of the Richards' equation (Richards, 1931):

$$\frac{\partial \theta}{\partial t} + S_s S_w \frac{\partial H}{\partial t} + \nabla \cdot (-K(h) \cdot \nabla H) = f_s \quad (1)$$

Where t [T] is time, θ [-] is volumetric soil water content, and H [L] and h [L] are the hydraulic and pressure heads respectively, so that $H = h + z$ (in which z [L] is depth taken positive upward). Furthermore, $K(h)$ [LT⁻¹] is unsaturated hydraulic conductivity as a function of pressure head, f_s [T⁻¹] is the sink/source term, S_s [L⁻¹] is specific storage, and S_w [-] is relative saturation (where $S_w = \theta / \theta_s$ in which θ_s is the saturated water content). The relationships between θ and h (called the water retention curve), and the function $K(h)$ can be characterized by the Mualem-van Genuchten model (Mualem, 1976; Van Genuchten, 1980):

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha h|^n)^m} \quad (2)$$

$$K(h) = k_s S_e^{1/2} \left[1 - (S_e^{1/m})^m \right]^2 \quad (3)$$

Where θ_r is the residual water content, k_s [LT^{-1}] is the saturated conductivity, α [L^{-1}] and n [-] are empirical parameters related to the mean and uniformity of the pore size, respectively, m is related to n by the formula $m = 1 - 1/n$, and S_e [-] is the effective saturation defines as $S_e = (\theta - \theta_r)/(\theta_s - \theta_r)$.

In the context of variably saturated flows, the development of performant numerical tools for solving the Richards' equation remains a challenging issue that also affects the development and application of efficient methodologies for parameter estimation and our knowledge of model-data interactions (Farthing and Ogden, 2017; Rajabi et al., 2018). In this study, the Richards' equation is numerically solved using a combination of Lumped Mixed Hybrid Finite Element (LMHFE) and the Method of Lines (MOL) (see Fahs et al., 2009). This is done by employing the 2D-UWF code (Belfort et al., 2017a) developed at the Laboratoire d'Hydrologie et de Géochimie de Strasbourg (LHyGeS). The spatial discretization of Richards' equation is based on MHFE method which approximates simultaneously both pressure head and velocity and can handle general irregular grids with highly heterogeneous permeability. Our 2D-UWF code embeds a lumped formulation to tackle difficulties of classical numerical schemes in satisfying the maximum principle, and LMHFE is then effective in avoiding unphysical oscillations in the solution. This approach is coupled with a sophisticated ordinary differential equation (ODE) solver for time integration (DLSODIS) that combines a high order adaptive time integration with an appropriate time-stepping management. Although this work includes efficient numerical methods and our main goal is not the resolution of the Richards' equation, other 2D/3D codes based on finite element or finite volume methods and integrating unstructured meshes (see McBride et al, 2006; Paulus et al., 2013; Deng et al, 2017), can also be employed within the framework of our

methodology. In this study our focus will be on the estimation of k_s , α and n as the uncertain input parameters.

2.2. Ensemble Kalman filtering approach

EnKF is a Monte Carlo-based variant of the Kalman filter (KF), which extends the KF to the case of nonlinear model systems. It was first introduced by Evensen (1994) and later modified by Burgers et al. (1998). Two formulations have been used in the literature for estimation of unsaturated flow parameters using the EnKF: (1) the joint or state augmentation approach (e.g. Wang et al. 2018), in which the states and parameters are simultaneously estimated as a single augmented vector; and (2) the dual estimation approach (e.g. Lu et al., 2011), which employs two interactive parallel filters: a filter for the parameters and another for the states, with the parameters undergoing an artificial evolution (or random walk) while waiting to be updated indirectly by the state variables (Rajabi et al., 2018). As indicated by Table 1, previous studies on the use of the EnKF for estimation of unsaturated flow parameters have mostly relied on the state augmentation approach. However, some (e.g. Moradkhani et al., 2005) have argued that this method is vulnerable to instability and intractability in highly nonlinear problems, and so the dual estimation approach is deemed superior in such circumstances. As Richards's equation-type models are known to be highly nonlinear (Erdal et al., 2014), we adapt the dual estimation approach in the current study. As a type of Bayesian filtering, the EnKF considers the uncertain model parameters and states as random variables, and hence represents them with probability distributions. These probability distributions are numerically approximated by an ensemble of realizations, which are constructed by sampling from the known prior distribution of each random variable at the onset of calculations. Each of these realizations is then separately propagated through time by a two-stage prediction/update process to estimate the marginal posterior distributions (or filtering

distributions) based on the measurements available in each time step t_i (i.e. discrete time intervals in which new measurements become available).

The potentially noisy nature of imaging data means that in the update stage the parameters can be perturbed to unrealistic values to obtain a closer match with the defective instances of data. Unsurprisingly, artificial evolution in the prediction stage may also lead to parameter estimations that are beyond physically reasonable limits. So it is necessary to account for the physical bounds on the parameters while computing their estimates. Therefore a ‘constrained’ version of the dual EnKF, based on the works of [Moradkhani et al. \(2005\)](#), [Bavdekar et al \(2013\)](#), and [Chen et al \(2018\)](#), is developed in the current study. The sequential steps of the adapted EnKF approach are describe in the following and presented in **Fig. 1**:

1. *Algorithm initiation*: the initial ensemble of parameter realizations is built by sampling from their respective prior distributions. These prior distributions characterize what is initially known about the possible range of variability for the model parameters, and the algorithm can often work well even if the prior distribution does not contain the correct values. The resulting ensemble of parameter realizations for the initial time step (denoted by ϕ_{t_0}) is made of n_e vectors of the uncertain parameters:

$$\phi_{t_0} = [k_s, \alpha, n] \quad (4)$$

Each of these realizations is then separately propagated through time by an iterative two-stage prediction/update process to estimate the marginal posterior distributions with respect to measurements available in each time step t_i .

2. *Prediction*: The input to the prediction stage is ϕ_{t_0} for the time step t_1 , and the outcome of the previous update stage for the subsequent time steps t_i ($i > 1$). The latter is denoted by $\phi_{t_i}^u$ where the superscripts u refers to the updated values. In the prediction step, first, an

artificial evolution (or random walk) of the parameter realizations is implemented by adding a normally distributed noise term (ϑ_t) (i.e. random walk term) with zero mean and a predefined covariance to each vector:

$$\phi_{t_{i+1}}^p = \phi_{t_i}^u \text{ (or } \phi_{t_0}) + \vartheta_t \quad (5)$$

In Eq. (5), the superscripts p refers to the forecasted values. Then, for each vector of the parameter realizations, a numerical simulation is performed to forecast the system state in the next observation time and produce the state ensemble $S_{t_{i+1}}^p$.

Given the strong dependence of soil moisture content on the model input parameters, the simulations are done from the initial time (t_0) to obtain the updated state variables in each assimilation step, and hence the name restart EnKF (re-EnKF) is given to this algorithm (Xu and Gómez-Hernández, 2016). As indicated by Song et al. (2014), using the restart version of the EnKF helps solve the inconsistency problem in which the updated parameters and state variables no longer follow the Richards' equation. The resulting process of state updating can be formulated as:

$$S_{t_{i+1}}^p = f(S_{t_0}, \phi_{t_{i+1}}^p) + \omega_t \quad (6)$$

Where $f()$ is a function representing the nonlinear model of the unsaturated flow, S_{t_0} denotes the state initial condition, and ω_t is the stochastic model error vector often assumed to be Gaussian with zero mean and a predefined covariance $Cov_{t_{i+1}}^f$ representing structural uncertainty in model predictions. Here, $S_{t_i}^p$ is a vector containing soil moisture content values in all nodes of the numerical grid, as there is an observation (i.e. set of pixel in the obtained images) for every numerical model element.

3. *Update*: In the update stage, a new set of measurements of the system state (denoted by $O_{t_{i+1}}$) is used to update the probability distributions obtained in the prediction stage by

applying Bayes rule. In our case, the system state (i.e. soil water content) is measured indirectly and hence measurement data are not of the same nature as the model output states. So a measurement model (or observation equation, denoted by $h(\cdot)$) is required to map the states to the specific type of data available from the images. In general, the measurement model can be formulated as (Liu et al., 2012):

$$M_{t_{i+1}} = h(S_{t_{i+1}}^p) + \eta_t \quad (7)$$

Where $M_{t_{i+1}}$ is the prediction ensemble, and the stochastic term η_t represents observation error. η_t is assumed to have a Gaussian distribution with a mean value representing systematic bias and a covariance that signifies the observation data uncertainty. For η_t , we assume that the mean is zero (no bias) and the covariance to have a predefined value denoted by $Cov_{t_{i+1}}^M$. The update of the parameter and state realizations includes the estimation of two Kalman gains for the parameters and states, represented by $K_{t_{i+1}}^\phi$ and $K_{t_{i+1}}^S$ respectively:

$$K_{t_{i+1}}^\phi = Cov_{t_{i+1}}^{\phi M} [Cov_{t_{i+1}}^{MM} + Cov_{t_{i+1}}^M]^{-1} \quad (8)$$

$$K_{t_{i+1}}^S = Cov_{t_{i+1}}^{SM} [Cov_{t_{i+1}}^{MM} + Cov_{t_{i+1}}^M]^{-1} \quad (9)$$

Where $Cov_{t_{i+1}}^{\phi M}$ is the cross-covariance of the parameter and prediction ensembles, $Cov_{t_{i+1}}^{MM}$ is the forecast error covariance matrix of the prediction, $Cov_{t_{i+1}}^{SM}$ is the cross covariance of state and prediction ensembles. Updating the state and parameter ensembles are subsequently done according to the standard KF equation:

$$\phi_{t_{i+1}}^u = \phi_{t_{i+1}}^p + K_{t_{i+1}}^\phi (O_{t_{i+1}} - M_{t_{i+1}}) \quad (10)$$

$$S_{t_{i+1}}^u = S_{t_{i+1}}^p + K_{t_{i+1}}^S (O_{t_{i+1}} - M_{t_{i+1}}) \quad (11)$$

In drawing random samples for algorithm initiation and the prediction and update stages, if any of the realizations of the parameters violates its constraints, it is projected to the boundary of the constrained space (similar to [Bavdekar et al., 2013](#)). Apart from the reasons previously mentioned, another key purpose of applying a constraint version of the EnKF is to prevent unsuccessful model runs when employing the numerical model in EnKF computations. The sequential prediction/update continues until the last measurement time step.

We call the above described algorithm, the constrained restart dual EnKF (CRD-EnKF), and in the subsequent sections of the paper, we intend to employ it for estimation of unsaturated flow parameters using photographic imaging data.

2.3. Polynomial chaos expansion (PCE)

A key challenge in the implementation of EnKF methods is the computational burden imposed by the requirement to perform a model simulation for each member of the ensemble in each time step. Large ensembles (often $\geq 1,000$) are required to obtain consistent results when the EnKF is repeated ([Erdal et al., 2014](#)). However, since numerical models of unsaturated flow are computationally demanding, using such large ensembles becomes computationally intractable. To tackle this computational challenge, we employ PCEs. The notion is to train PCEs for each model output quantity of interest (QoIs) using a set of numerical model input-output pairs, and then use the resulting PCEs as meta-models in EnKF computations. Details about PCE can be found in e.g. [Le Maître and Knio \(2010\)](#), [Rajabi et al., \(2015\)](#) and [Rajabi \(2019\)](#). Here we only give a brief overview.

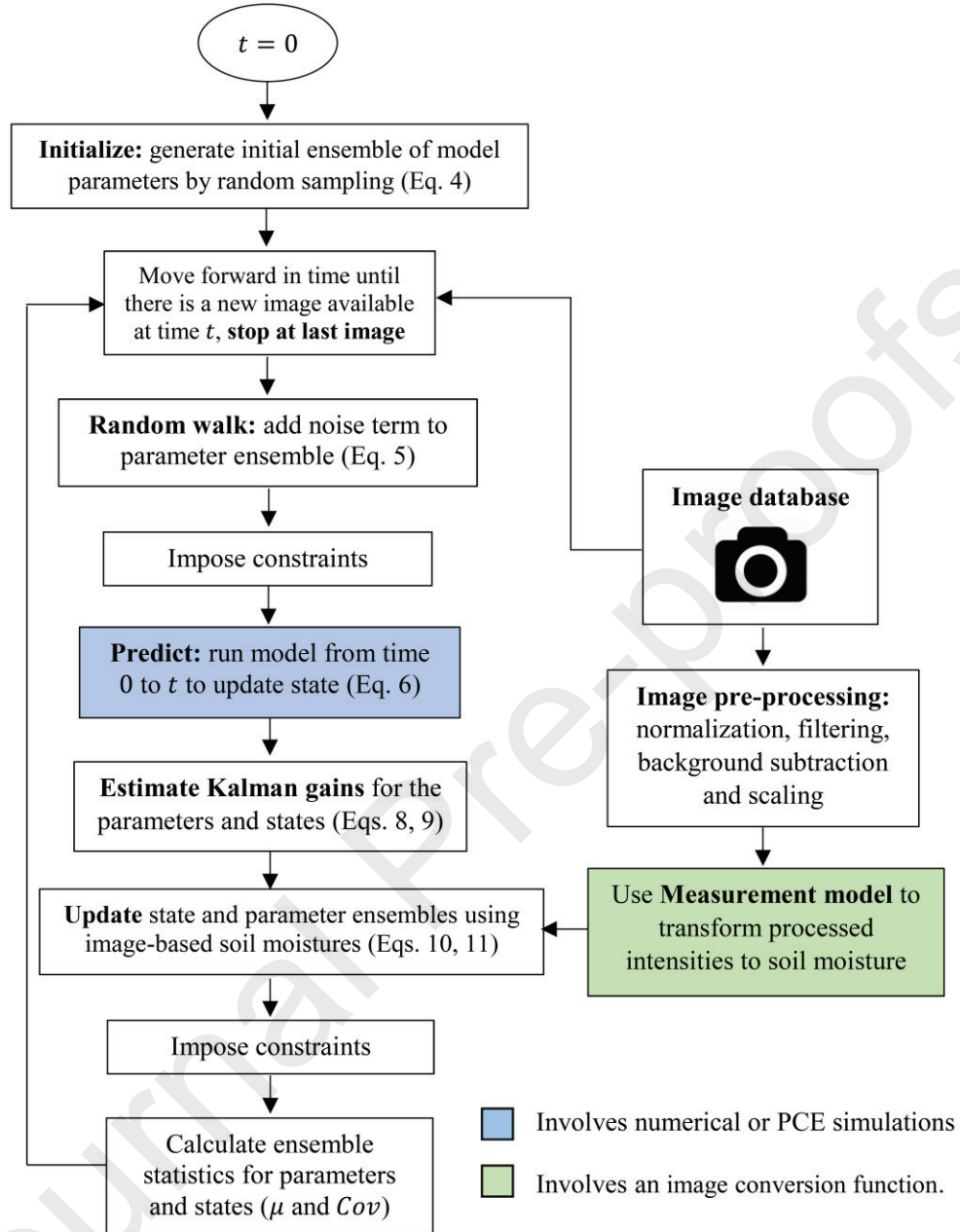


Fig. 1. Algorithm flowchart for estimation of soil moisture and unsaturated flow parameters with the CRD-EnKF

We consider x to be the vector of the input variables (here, consisting of k_s , α and n), and y the set of model output QoIs (e.g. soil moisture values). x is assumed to be affected by uncertainty, and is hence represented by a random vector X with a particular probability density function

(PDF). Due to the uncertainty in model inputs, the output QoIs are also random variables denoted by $Y = \Omega(X)$. Assuming that the input random variables in X are independent, Y can be expanded onto an orthogonal polynomial basis as follows (Ghanem, 1998; Li and Zhang, 2007):

$$Y = \Omega(X) = \sum_{i=0}^{\infty} \alpha_i \psi_i(X) \quad (12)$$

Where the series in Eq. (12) converges in the sense of the L_2 -norm. Every element of this series is the product of a deterministic coefficient α_i and a stochastic multivariate polynomial $\psi_i(X)$. The optimal choice of the polynomial type for $\psi_i(X)$ depends on the probability distribution for X . This is often done using the Askey family of hypergeometric polynomials (Askey and Wilson, 1985). The series in Eq. (12) is commonly referred to as PCE. For computational purposes, Eq. (12) is truncated after a number of terms (denoted by O_p) (Blatman and Sudret, 2011):

$$\Omega(X) = \sum_{i=0}^{O_p} \alpha_i \psi_i(X) \equiv \alpha^T \psi(X) \quad (13)$$

Where α and ψ gather the PCE coefficients and basis polynomials respectively.

The deterministic PCE coefficients α_i must be estimated for every QoI and this can be done using non-intrusive methods (such as the regression method) which treat the physics of the system as a black-box and use a training dataset for the estimation of α_i (Rajabi et al., 2015; Rajabi, 2019). The training dataset consists of a set of N realizations of the input random vector generated by e.g. Monte Carlo sampling, which we denote by $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$, and the corresponding model outputs which are represented by $\{\Omega(x^{(1)}), \Omega(x^{(2)}), \dots, \Omega(x^{(N)})\}$. In the regression method, the set of PCE coefficients (α) are estimated by (Sudret, 2008; Blatman and Sudret, 2011):

$$\alpha = \left(\frac{1}{N} \sum_{i=1}^N \psi(x^{(i)}) \psi^T(x^{(i)}) \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \psi(x^{(i)}) \Omega(x^{(i)}) \right) \quad (14)$$

2.4. Extending EnKF for conversion of light intensities to water contents

The proposed CRD-EnKF can also be applied to calibrate an equation that transforms light intensities to water content, without resorting to statistical regression based on data from direct measurements of the water content. As previous indicated, this can be done on the condition that the properties of the soil are known in advance. Using CRD-EnKF for conversion of light intensities to soil water contents requires modification to the algorithm described in sub-section 2.2, and in the following we describe the sequential steps of the algorithm modified for this purpose:

1. *Algorithm initiation*: For the sake of simplicity, we assume that a particular functional form (e.g. linear function) is chosen in prior for the equation that relates light intensities to water contents (which we call the ‘conversion function’), and the CRD-EnKF algorithm is meant to calibrate this equation by estimating its coefficients (c_i). Hence the vector of uncertain parameters in Eq. (4) becomes $\phi_{t_0} = [c_0, c_1, \dots, c_n]$. As described in sub-section 2.2, the initial ensemble of parameter realizations is constructed by sampling from the respective prior distributions of the coefficients, and then each realization is separately propagated through time by an iterative two-stage prediction/update process.
2. *Prediction*: In the random walk process described by Eq. (5), a normally distributed noise term is added to the most recent values of the coefficients $[c_0, c_1, \dots, c_n]$, and then for each vector of the parameter realizations, the soil moisture content is forecasted based on the photographic imaging data of the next observation time. To estimate the

coefficients $[c_0, c_1, \dots, c_n]$, the $f(\cdot)$ function in Eq. (6) is no longer the Richards' equation-type numerical simulator or its PCE surrogate, and is instead the conversion function that relates light intensities to soil water contents (with the latest values of its coefficient). Pre-processed light intensities for all points in the problem domain are stored in a database, and the ones pertaining to the desired time step are fed into the conversion function in the prediction stage of the EnKF.

3. *Update*: The data used for updating the unknown coefficients of the conversion function in each assimilation step (i.e. $M_{t_{i+1}}$), is the output of the numerical model at the same time step and in every point in the numerical grid, assuming that the input parameters of the numerical model (e.g. k_s, α and n) are known. The numerical simulation (single one required) is done before initiation of the CRD-EnKF algorithm, and the relevant model outputs are stored for use by the EnKF. In the update stage, the numerical model outputs are used to update the probability distributions obtained in the prediction stage by applying Bayes rule (see Eqs. 8, 9, 10, 11).

The whole process of using the CRD-EnKF for conversion of light intensities to water contents has a key challenging. In the algorithm described in the sub-section 2-2, it's conceptually easy to define physically plausible prior distributions and constraints for the unsaturated flow parameter based on previous knowledge of soil characteristics. But the coefficients of the conversion function have no physical limits (i.e. can vary on the interval $(-\infty, +\infty)$), and there is no readily understandable relation between these coefficients and soil characteristics or other physical traits of the unsaturated flow system to be exploited for the definition of priors and constraints. We can solve this problem by noting that for any combination of these coefficients, the conversion function should produce physically reasonable moisture content values for all the available pre-processed

light intensities. Hence the outcome of the conversion function (not the individual coefficients) can have meaningful constraints. Algorithm initiation may be conducted with a very broad (i.e. uninformative) prior distribution. Then each sample drawn from the prior is checked for this constraint and only the ones that satisfy this constraint makeup the initial ensemble. This constraint is also imposed in the prediction and update stages of the CRD-EnKF algorithm. The CRD-EnKF algorithm for conversion of light intensities to water contents is depicted in **Fig. S1** of the **Supplementary Material**.

3. Application

Data obtained from the imaging technique proposed by [Belfort et al \(2017b\)](#) is used as the basis for validation of the proposed methodology. The experimental setup and procedure is explained in sufficient detail in [Belfort et al. \(2017b\)](#), and so here we provide a brief description.

3.1. Experimental setup

The experiment setup included a Plexiglas tank of 40 cm × 14 cm × 6 cm inner dimensions. The walls of the tank were transparent, hence allowing visual observation of water content variations. The tank was filled with monodisperse quartz sand. A Nikon digital camera D80, was placed on a fixed metal tripod at a distance of 2 m from the tank to capture linear images from the tank in the 12-bit RAW format. In these linear images, the value at every pixel was directly related to the amount of light received from the tank during the exposure.

To improve visualization, a light device was placed at the bottom front of the tank, and a black curtain was arranged around the tank and camera. White or black paper strips placed along the side, bottom and top edge of the tank were intended to capture the maximum and minimum light intensities for the normalization of data. A Theta probe (Type ML2x, made by Delta-T Devices

Ltd.) was also placed vertically in the middle of the tank surface (covering an area of $4 \text{ cm} \times 6 \text{ cm}$) to provide contact-based direct measurement of the water content. This apparatus had an accuracy of $\pm 0.01 \text{ m}^3 \cdot \text{m}^{-3}$ for measurements between 0.05 and $0.6 \text{ m}^3 \cdot \text{m}^{-3}$. The accumulated mass of water entering and leaving the flow tank was additionally measured. A schematic representation of the experimental setup is shown in **Fig. 2**.

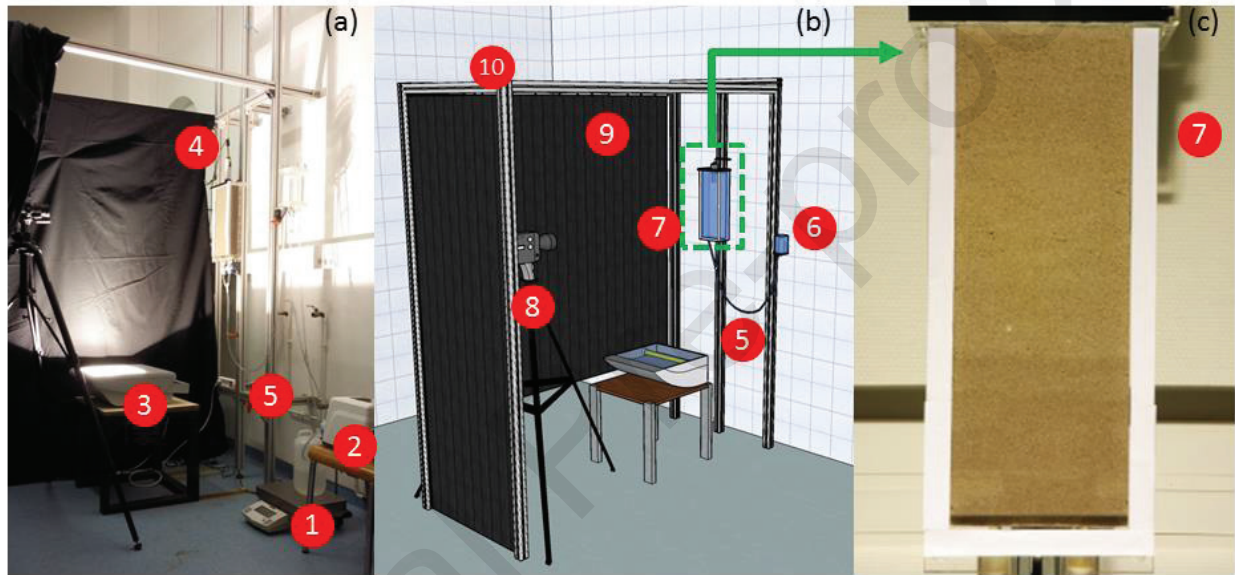


Fig. 2. The non-invasive imaging technique: (a) photograph of the experimental setup, (b) its schematic representation, and (c) image of the flow tank. (1: Sartorius digital balance; 2: peristaltic Cole-Parmer pump; 3: floodlight (400 W); 4: Theta probe ML2x connected to data logger; 5: pipes between tank - balance - overflow outlet; 6: overflow outlet (can be moved along the vertical axis); 7: Plexiglas flow tank (40 cm x 14 cm x 6 cm); 8: NIKON digital camera D80; 9: black curtain (all around the experiment); 10: Norcan metallic structure)

3.2. Experimental procedure

We tend to estimate unsaturated soil parameters using images obtained from a single drainage/imbibition cycle as described in [Belfort et al. \(2017b\)](#). As shown in **Fig. 3**, successive

drainage and imbibition steps are achieved by moving an overflow outlet connected to the bottom of the flow chamber. The entire cycle takes 122,400 seconds, and images are taken every 1,800 seconds during the experiment, except for the first 3 images which are taken with 600 second intervals. This results in a total of 70 images, from which the first 69 images are used for the data assimilation process. The Raw images are pre-processed before they are employed in the parameter estimation process. The pre-processing includes the following successive steps: (1) RAW images are first converted to 16-bit tiff images, (2) the resulting tiff images are transformed into the RGB color space, (3) normalization is done on the light intensities using the minimum and maximum intensities obtained from the black and white paper strips, and (4) the value of the background intensity is subtracted from the resulting normalized intensities, pixel by pixel, to obtain the corrected intensities which we denote by I_{pr} .

3.3. Numerical modeling

A numerical model of the sandbox experiment was also developed using the 2D-UWF code. The sandbox was modeled as a 2D system, where the top and lateral boundaries were assumed to be impermeable, and a varying pressure head corresponding to the outlet overflow displacements, was assumed at the bottom boundary. The domain was discretized to 560 irregular triangular elements, each having an area of about 1 cm² (see **Fig. 3**). As the model initial condition, the domain was assumed to be fully saturated. Further details about the numerical model, and its validation through comparison with the real observed data can be found in [Belfort et al. \(2017b\)](#).

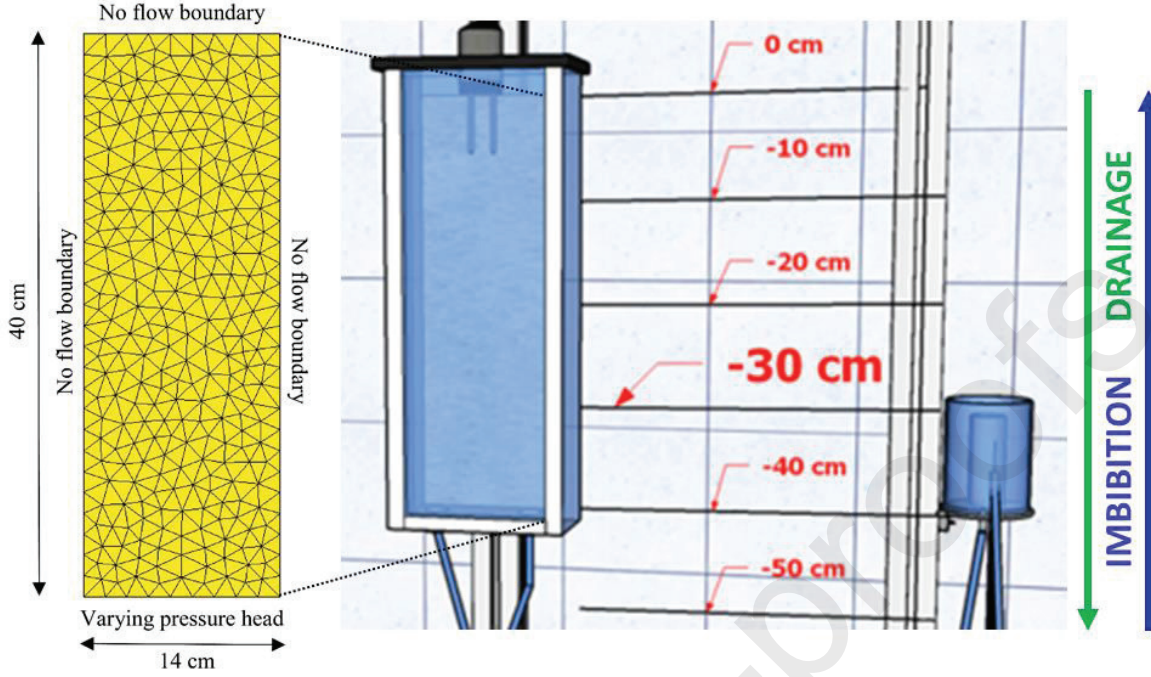


Fig. 3. Conceptual model of the 2D flow tank and its numerical grid.

We tend to employ the proposed CRD-EnKF approach to estimate k_s , α and n , and hence these three parameters are represented by independent random variables. The sand used in the present setup has been characterized by other column experiments subject to parameter estimation by classical 1D inverse modeling. Combined with the results of the study dealing with the photometric imaging technique (i.e. [Belfort et al., 2017b](#)), values of these parameters can be fixed in advance (see **Table 2**), and these target values serve as a benchmark for evaluation of the CRD-EnKF results. Other input parameters of the model were considered to be deterministic, so $\theta_r = 0.088$ and $\theta_r = 0.376$ (both in cm^3/cm^3) in all model simulations.

Table 2. Target values of the parameters to be estimated by the CRD-EnKF

Parameter	Target values	Unit
k_s	2.74×10^{-2}	cm.s^{-1}
α	0.054	cm^{-1}
n	7.37	--

3.4. PCE meta-model development

850 realizations of the input random vector $[k_s, \alpha, n]$ are generated by Monte Carlo sampling, and are then used as the basis for the same number of numerical model simulations. For each model run, soil moisture values at the 560 model elements, in addition to the estimate of mass entering or leaving the flow domain, are extracted from the model outputs files for each of the 69 time steps described in sub-section 3.2. 750 input-output pairs are used for PCE training, 50 for validation and the last 50 for testing. One PCE meta-model is required for each QoI, and hence $69 \times (560 + 1) = 38,709$ PCEs are developed by applying the regression method to the described training dataset. For each QoI, the order of polynomial (O_p) is chosen by comparing PCEs of different orders from $O_p = 0$ to 8, and the PCE order with the lowest normalized root mean square error (NRMSE) value with respect to the validation dataset is chosen.

4. Results and discussion

4.1. Evaluation of PCE results

In this sub-section we validate and assess the performance of the developed surrogate PCE meta-models. A scatterplot of numerical model vs. PCE meta-model soil moisture estimations is presented in **Fig. 4**. The plot is based on an arbitrarily chosen instance of the test dataset and shows the associated values for all 560 soil moisture estimations. Three additional straight lines are also depicted in the figure. The diagonal line indicates perfect calibration, and the two dashed lines bound an area in which the numerical model and PCE meta-model results differ by less than a factor of two. As shown, the points in this scatter plot fall close to the one-one line, and so no systematic bias is observed for PCE estimations. Scatterplots based on the rest of the test dataset are very similar to what is shown in **Fig. 4**.

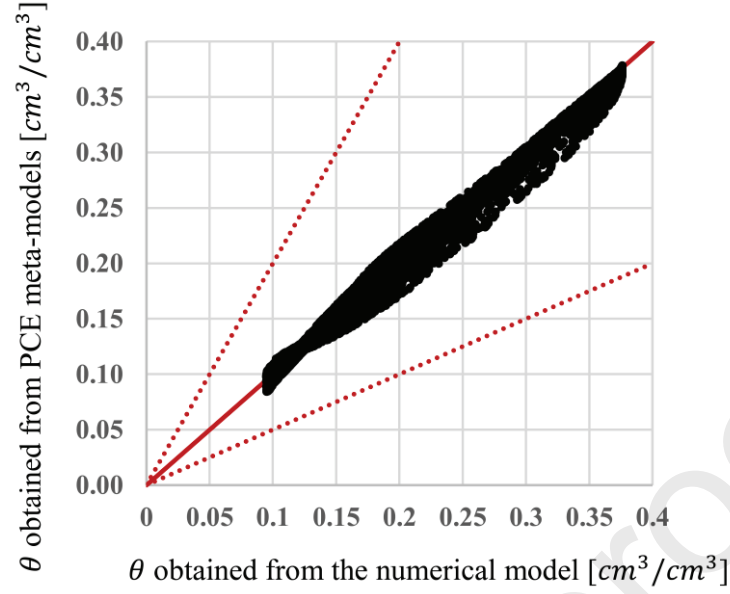


Fig. 4. Scatter plot of PCE vs. numerical model soil moisture (θ) estimations for an instance of the test dataset. Red dashed lines indicate a $\pm 100\%$ difference between the original model and its surrogate.

Additional statistical diagnostics for the constructed PCE meta-models are presented in **Table 3** along with their ideal and acceptable range of values. The table is based on the calculation of the following four diagnostics using the test dataset: (1) normalized mean square error (NMSE), (2) fractional bias (FB), (3) Pearson correlation coefficient (R), and (4) fraction of the PCE results located within a factor of two of the numerical model results (FAC2). As shown in table, the range of values for the 560 QoIs fall within the acceptable intervals and lie close to the ideal values. The outputs are free from bias, since both FB and NMSE are close to zero and FAC2 is one. A high degree of correlation exists between the numerical model and PCE meta-model outputs as R is close to one.

Table 3. Summary of statistical diagnostics for PCE performance assessment

Statistics	FAC2	NMSE	FB	R
Aim ^a	1	0	0	1
Acceptable Range ^a	> 0.5	(-0.3, +0.3)	< 1.5	> 0.8
Range of results	1	(6.78×10^{-5} , 7.05×10^{-4})	(-4.19×10^{-4} , 1.58×10^{-4})	(0.96, 0.99)

^aBased on [Chang and Hanna \(2005\)](#) and [Moonen and Allegrini \(2015\)](#).

4.2. State and parameter estimation using photographic imaging data

After testing the constructed PCE meta-models, we employ them in the framework of the CRD-EnKF algorithm. Several crucial filter parameter must be decide before the algorithm can be initiated, this includes characteristics of the initial ensemble, as well as ensemble size and parameter variation range in the constraints.

In general, the initial ensemble mean should be chosen so as to represent the initial guess for the model parameter values, and the spread of the ensemble must indicate the uncertainty of that initial guess ([Evensen, 2003](#)). But in our study the target values of model parameters are known. So the initial values of k_s , α and n are obtained by shifting from their target values, and are assumed to be equal to 0.005 cm.s^{-1} , 0.035 cm^{-1} and 4.5 respectively. We use an ensemble size of 10,000. Such a large ensemble size is only made possible by using the PCE meta-models. Other filter parameter values are chosen as follows: the covariance of the artificial evolution term (ϑ_t) is equal to $[0.01k_s, 0.001\alpha, 0.1n]$, model error (ω_t) covariance equals 10^{-3} , and observation error (η_t) covariance equals 10^{-4} . In the proposed PCE-based CRD-EnKF approach, the stochastic model error (ω_t) reflects the cumulative, inseparable effect of numerical model structural uncertainty and uncertainty due to surrogate modeling.

The parameter variation range can be constrained according to prior knowledge (Carsel and Parrish, 1988). It's recommended that a broad range of variation be used to prevent excessive external interference in the assimilation process (Li and Ren, 2011). Here, we define the constraints as follows: $1 \times 10^{-4} \leq k_s \leq 1 \text{ cm.s}^{-1}$, $0.02 \leq \alpha \leq 0.09 \text{ cm}^{-1}$, and $2 \leq n \leq 10$. Such constraints ensure that the CRD-EnKF results in parameter values that are physically plausible and in harmony with prior knowledge.

After choosing the hyper-parameter values, we assimilate the photographic imaging data to simultaneously update the state of the soil moisture content and the hydraulic parameters. Our photographic imaging data includes light intensity values at 1 by 1 cm pixels in a $14 \times 40 \text{ cm}$ domain, resulting in a total of 560 light intensity values being assimilated at each time step. This is the same as the number of model elements, but since the triangular shape of model elements do not coincide with the square shape of image pixels, a code was written to match every model element with the image pixel that has the maximum overlap with it.

Apart from photographic imaging data, we also apply the CRD-EnKF to data obtained from direct measurement of soil moisture content (using the Theta probe) at a single location (one data point per time step). **Fig. 5** shows the evolution of the estimated parameter values (i.e. ensemble mean) across assimilation steps for both datasets.

As demonstrated, the CRD-EnKF algorithm is able to significantly improve the wrong initial values of model parameters using both types of data. Estimates obtained from photographic imaging data are in better agreement with the target values of k_s and n compared to the ones acquired from the use of direct measurement of soil moisture content, while the reserve is true for α .

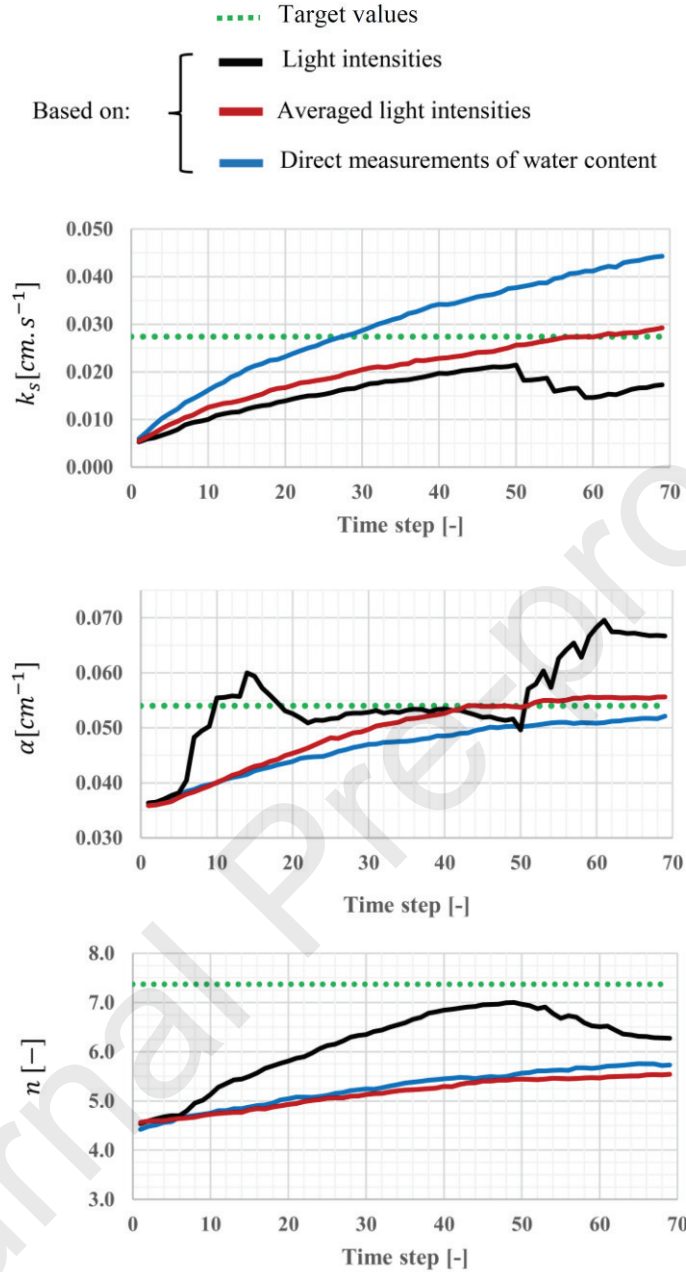


Fig. 5. Estimation of unsaturated flow parameters using different data. Plots show the time series of mean values of the ensembles.

One of the reasons why the parameter estimates obtained from photographic imaging data do not seem to lived up to the expectation (at least for α) is the presence of spatial variability in the experimental setup, which is captured by the images and not accounted for by the numerical model.

Despite all the precautions taken when filling the tank, sand compaction is not uniform and it creates different air penetration pathways and therefore a different distribution of water content. The intrinsic differences between the experimental reality captured by the images and the simplifying and unifying vision of the numerical model is illustrated in the exemplary **Fig. 6a**, which shows the soil moisture content distribution obtained from the image taken in the 25th time step (41,400 seconds after initiation of the experiment). **Fig. 6b** demonstrates the soil moisture content distribution acquired through numerical modeling (with the target parameter values) at the same time step. Comparison of the two figures clearly shows the existence of compaction heterogeneities with a significant orientation in specific regions of the domain in the photographic image (also see figure 7 in [Belfort et al, 2017b](#) for similar maps pertaining to other time steps). Image noise, which by nature could also be anisotropic due to disturbances such as lighting, have been reduced by making average of the intensity over squares of 1cm^2 . Nevertheless, both image noise and non-uniform compaction may affect parameter estimates.

As one way to reduce the effect of image noise and non-uniform compaction, we employ the ‘lateral average’ of light intensities obtained from photographic imaging as the basis for parameter estimations. So instead of a 2D field, we employ a vertical profile, and as a result, we have 40 values per assimilation time step instead of the previous 560. Parameter estimates obtained from applying these averaged values are shown with red curves in **Fig. 5**. The curves show that the target values of k_s and α are reproduced closely with the averaged values of light intensity, while the estimate of n has moved farther from the target value. For all three parameters (k_s , α and n) the ensemble means have less fluctuation in the evolution process if averaged values are used instead of the entire 2D field.

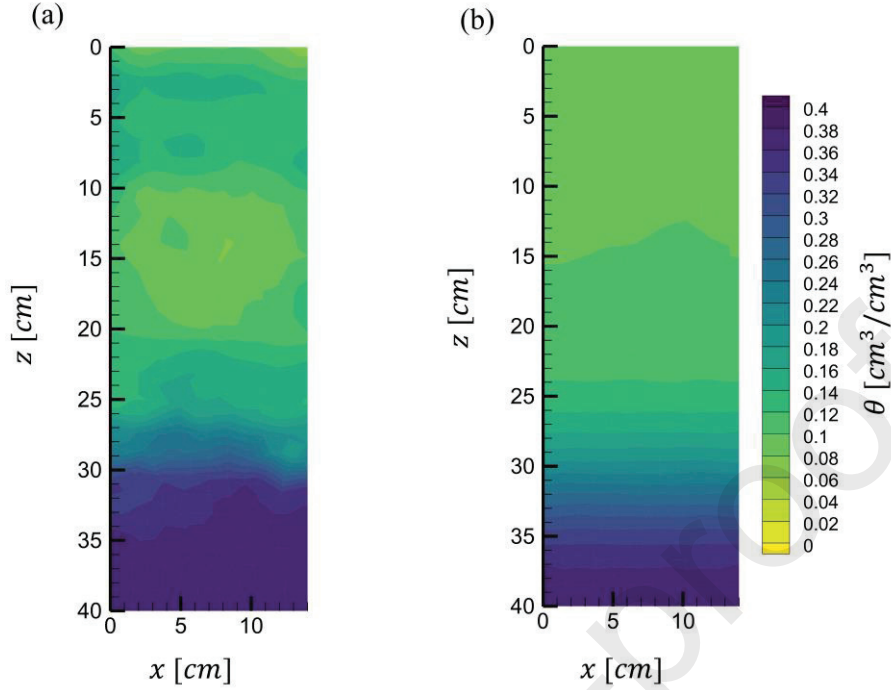


Fig. 6. Comparison of soil moisture distributions obtained from: (a) photographic image, and (b) numerical model, for the 25th time step (41400 seconds after initiation of the experiment)

Table 4 demonstrates the standard deviations (σ) of the final ensembles as a measure of estimation uncertainties (the associated cumulative density functions (CDF) are presented in **Fig. S2** of the **Supplementary Material**). The figures show that estimates obtained from using the entire photographic imaging dataset have the lowest σ and hence the narrowest spreads of the ensembles, for all three parameters (k_s , α and n). Parameter estimates based on averaged light intensities have the second lowest σ values, while estimates acquired through the use of direct measurements of water content have the highest σ values. The difference in σ values for the three datasets are more significant for k_s and α compared to n . The difference in the ensemble spread and the associated σ values for the three datasets is related to the ‘volume of data’ provided by each. When using the entire photographic imaging dataset, 560 data points are provided in each time step, whereas

averaged light intensities include 40 data points, and direct measurement a single data point per unit time step.

Table 4. Comparison of the standard deviations of the unsaturated flow parameters at the last time step obtained using different types of data.

Data type used in parameter estimations		$\sigma(k_s)[cm.s^{-1}]$	$\sigma(\alpha)[cm^{-1}]$	$\sigma(n)[-]$
Based on:	light intensities	0.0165	0.0164	2.328
	averaged light intensities	0.0249	0.0214	2.412
	direct measurements of water content	0.0374	0.0222	2.547

To analyze the performance of the proposed algorithm in state estimations, we focus on the case in which averaged light intensities are used as input data to the assimilation process. In **Fig. 7a**, the solid red line marks the mean of the ensemble for soil moisture content at the location of the measurement probe. In this figure, the dashed green line shows the true moisture content (measured by the Theta probe) as a function of time. As demonstrated, the soil moisture values estimated by assimilating the imaging data are close to the values obtained from direct measurements of water content, and both follow the same pattern imposed by the drainage/imbibition cycle. Furthermore, **Fig. 7b** shows the evolution of the cumulative mass entering and leaving the flow tank. Again, the associated ensemble mean values obtained from applying the photographic imaging data to the CRD-EnKF algorithm (the solid red line) match rather well with the true values acquired through direct mass measurements.

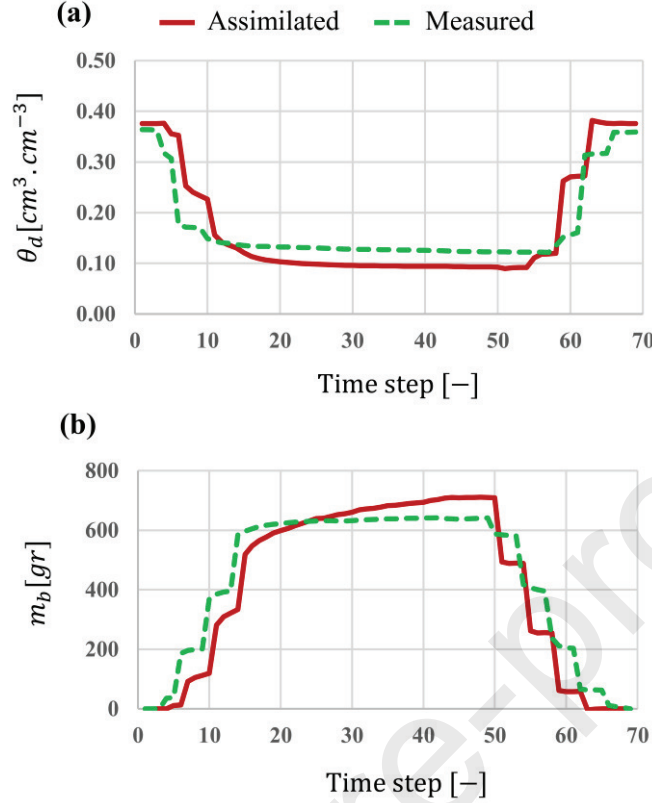


Fig. 7. Evolution of (a) soil moisture content at the direct measurement location (θ_d), and (b) cumulative mass entering and leaving the flow tank.

4.3. Comparison of PCE-based vs. numerical model-based CRD-EnKF

The logic behind the use of PCEs in the proposed algorithm can be explained using **Fig. 8**. This exemplary figure shows estimation of the shape parameter α obtained from two PCE-based and two numerical model-based repetitions of the CRD-EnKF. The ensemble size is 10,000 for PCE-based, and 12 for numerical model-based algorithms. Note that an ensemble size of 12 results in a total of 828 simulation for the 69 time steps, which is close to the number of numerical model runs previously used for PCE meta-model development (i.e. 850). The four repetitions of the CRD-EnKF are based on different initial values of α , but all other filtering parameters are the same. A close look at **Fig. 8** reveals two basic differences between PCE-based and numerical model-based CRD-EnKF results. First the PCE-based results are less affected by the initial ensemble, as

the two repetitions converge to nearly identical values. But the same cannot be implied for the numerical model-based result, as the two repetition converge to values that are somewhat different. Second, the PCE-based evolution curve is smoother than its numerical model-based counterpart, the latter having significant random fluctuations before converging to a specific value. **Fig. 8** is a representative figure, and the same traits are observed for the other input parameters (k_s and n) and different initial ensembles. Both of the above described distinctions are the result of the fact that, for nearly the same computational time, PCE allows for ensemble sizes that are several orders of magnitude larger than the case in which numerical models are directly incorporated into EnKF algorithms.

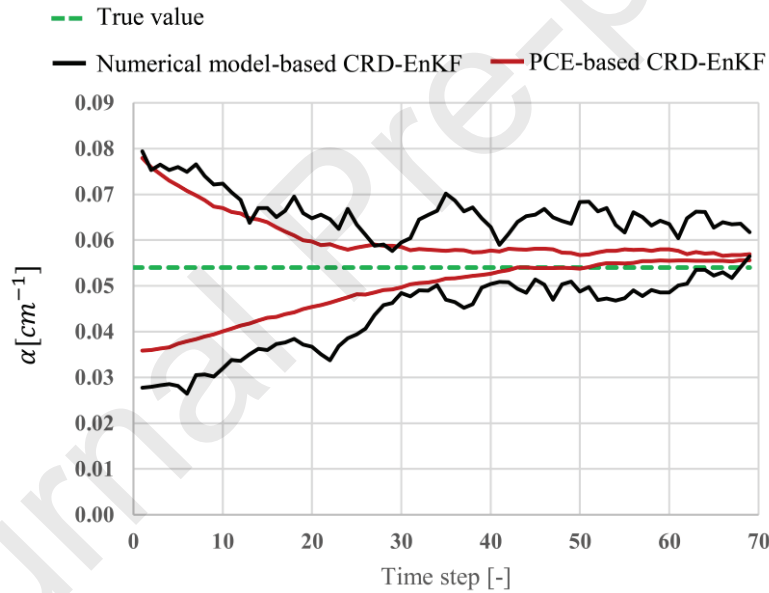


Fig. 8. PCE vs numerical model-based EnKF results: comparison of five iterations with different prior distributions.

4.4. Using the EnKF for developing relationship between light intensities and water content

We now apply the algorithm described in sub-section 2.4 to estimate the conversion function that relates light intensities to soil water content without the use of direct measurements of the soil water content. We assume a linear conversion function of the following form:

$$\theta_{img}(x, y, t) = c_0 - c_1 I_{pr} \quad (15)$$

In this experiment, a fully saturated soil has $\theta = 0.3635$, and from the images we know that the saturated soil has an I_{pr} equal to zero. Hence c_0 is also equal to 0.3635, and the CRD-EnKF algorithm is left with the task of estimating c_1 . [Belfort et al \(2017b\)](#) employed direct measurement of water content to estimate c_1 , and their estimated value can serve as a benchmark for evaluation of the CRD-EnKF results. A numerical simulation is done before initiation of the algorithm with the target values of k_s , α and n (see **Table 2**) as inputs. The soil moisture content values in all model elements are extracted for the 69 time steps, and are stored for use by the CRD-EnKF. The ensemble size is chosen to be 1,000 and the ensemble members are constrained so that the conversion function always produces a value of $0 \leq \theta_{img} \leq 0.3635$.

Fig. 9 illustrates the evolution of the ensemble mean (as the best estimate) for c_1 with the progress of assimilation steps. At the 32nd step, the c_1 value estimated by the CRD-EnKF algorithm is within a range of $\pm 8\%$ deviation from the value estimated by [Belfort et al \(2017b\)](#) (i.e. $c_1 = 2.3425$), and remains within this range until the end of the assimilation process. Hence the figure clearly demonstrates convergence to the true value of c_1 . This implies that the proposed algorithm is able to estimate the coefficient of the conversion function without the need for direct measurements of water content, on the condition that an accurate numerical model of the system is available.

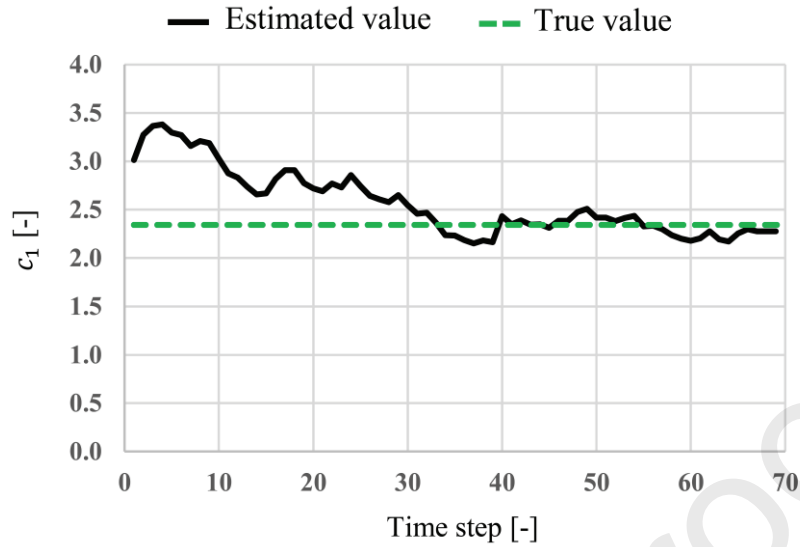


Fig. 9. Evolution of the ensemble mean for coefficient c_1 used in the conversion of light intensity to soil moisture content

5. Conclusion

In this study, a sequential data assimilation approach was developed to simultaneously update soil moisture content and soil hydraulic parameters using photographic imaging data as input. For this purpose, the study combines numerical modeling, PCEs, restart and constrained EnKF and the dual estimation approach to formulate a novel version of the EnKF.

We validated the proposed methodology by using lab-scale data obtained from images of a 2D Plexiglas flow tank in a drainage/imbibition cycle. The results indicated that the proposed method is able to significantly improve the wrong initial values of saturated conductivity (k_s) and empirical shape parameters α and n . Estimates obtained from assimilating photographic imaging data show better agreement with target values of k_s and n , compared to estimates obtained from employing sparse direct measurement of soil moisture content (θ). The ensemble spread, which represents uncertainty in estimation, was also significantly smaller for all the three estimated parameters (k_s , α and n) when employing imaging data instead of direct measurement of θ . To

reduce the effect of image noise and non-uniform compaction on the estimation, we also employed the lateral average of light intensities obtained from photographic imaging as the basis for parameter estimations. This resulted in better estimations of k_s and α . The soil hydraulic parameters were homogeneous in our problem domain, but the results obtained by this study show an encouraging way to estimate ‘distributed’ fields of unsaturated soil hydraulic parameters in heterogeneous soils, as photographic imaging data allows for cheap, non-invasive characterization of soil moisture content at very fine resolutions. This could be a topic for future research.

In this paper, we also propose an EnKF approach to estimate the coefficients of the conversion function that relates light intensities to soil moisture content. The approach uses numerical model outputs instead of direct measurements, and can be useful only when a good estimate of model input parameters is already available. Here we assumed a linear conversion function, but the method can be similarly applied to other form of conversion functions in future research.

The photographic approach requires access to the soil section, which is more easily achievable in laboratory experiments. The CRD-EnKF approach allows to exploit a large part of 2D structural mapped information to improve the parameterization of hydrodynamic models. The validation of this approach thus makes it possible to envisage its use both under other experimental conditions (e.g. employing more cohesive soils or testing heterogeneous porous media), and by relying on data from other sources, with significant work to be carried out on the effect of resolution and noise.

Acknowledgement

The study was partially funded by Strasbourg University and National School for Water and Environmental Engineering (ENGEES); we acknowledge their scientific councils for their confidence.

References

1. Abbaspour, K. C., Schulin, R., & Van Genuchten, M. T. (2001). Estimating unsaturated soil hydraulic parameters using ant colony optimization. *Advances in water resources*, 24(8), 827-841.
2. Allroggen, N., van Schaik, N. L. M., & Tronicke, J. (2015). 4D ground-penetrating radar during a plot scale dye tracer experiment. *Journal of Applied Geophysics*, 118, 139-144.
3. Askey, R., Wilson, J. (1985). Some Basic Hypergeometric Polynomials that Generalize Jacobi Polynomials, Memoirs of the American Mathematical Society. AMS, Providence, RI, pp. 319.
4. Bauser, H. H., Jaumann, S., Berg, D., & Roth, K. (2016). EnKF with closed-eye period—towards a consistent aggregation of information in soil hydrology. *Hydrology and Earth System Sciences*, 20(12), 4999-5014.
5. Bavdekar, V. A., Prakash, J., Shah, S. L., & Gopaluni, R. B. (2013, June). Constrained dual ensemble Kalman filter for state and parameter estimation. In 2013 American Control Conference (pp. 3093-3098). IEEE.
6. Belfort B., Fahs M., Lehmann F., (2017a). 2D-UWF User's guide. LHyGeS Report. University of Strasbourg, France.
7. Belfort, B., Weill, S., & Lehmann, F. (2017b). Image analysis method for the measurement of water saturation in a two-dimensional experimental flow tank. *Journal of hydrology*, 550, 343-354.
8. Belfort, B., Weill, S., Fahs, M., & Lehmann, F. (2019). Laboratory Experiments of Drainage, Imbibition and Infiltration under Artificial Rainfall Characterized by Image Analysis Method and Numerical Simulations. *Water*, 11(11), 2232.
9. Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6), 2345-2367.
10. Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6), 1719-1724.
11. Carsel RF, S Parrish R. (1988). Developing joint probability distributions of soil water retention characteristics. *Water Resources Resreach*, 24(5): 755–69.
12. Chang, J.C., Hanna, S.R., 2005. Technical Descriptions and User's Guide for the BOOT Statistical Model Evaluation Software Package. Available through. www.harmo.org/kit.
13. Chaudhuri, A., Franssen, H. J. H., & Sekhar, M. (2018). Iterative filter based estimation of fully 3D heterogeneous fields of permeability and Mualem-van Genuchten parameters. *Advances in Water Resources*, 122, 340-354.
14. Chen, Z., Gómez-Hernández, J. J., Xu, T., & Zanini, A. (2018). Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter. *Journal of hydrology*, 564, 1074-1084.
15. Crow, W. T., & Zhan, X. (2007). Continental-scale evaluation of remotely sensed soil moisture products. *IEEE Geoscience and Remote Sensing Letters*, 4(3), 451-455.

16. Deng, B., & Wang, J. (2017). Saturated-unsaturated groundwater modeling using 3D Richards equation with a coordinate transform of nonorthogonal grids. *Applied Mathematical Modelling*, 50, 39-52.
17. Erdal, D., Neuweiler, I., & Wollschläger, U. (2014). Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model. *Water Resources Research*, 50(1), 132-147.
18. Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), 10143-10162.
19. Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4), 343-367.
20. Fahs, M., Younes, A., & Lehmann, F. (2009). An easy and efficient combination of the mixed finite element method and the method of lines for the resolution of Richards' equation. *Environmental Modelling & Software*, 24(9), 1122-1126.
21. Fang, B., & Lakshmi, V. (2014). Soil moisture at watershed scale: Remote sensing techniques. *Journal of hydrology*, 516, 258-272.
22. Farthing, M.W. and Ogden, F.L. (2017), Numerical Solution of Richards' Equation: A Review of Advances and Challenges. *Soil Science Society of America Journal*, 81: 1257-1269.
23. Ghanem, R. (1998). Probabilistic characterization of transport in heterogeneous media. *Computer methods in applied mechanics and engineering*, 158(3-4), 199-220.
24. Hoa, N. T. (1981). A new method allowing the measurement of rapid variations of the water content in sandy porous media. *Water Resources Research*, 17(1), 41-48.
25. Hollenbeck, K. J., & Jensen, K. H. (1998). Maximum-likelihood estimation of unsaturated hydraulic parameters. *Journal of hydrology*, 210(1-4), 192-205.
26. Ippisch, O., Vogel, H. J., & Bastian, P. (2006). Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation. *Advances in water resources*, 29(12), 1780-1789.
27. Jiang, Z., Huang, Q., Li, G., & Li, G. (2019). Parameters Estimation and Prediction of Water Movement and Solute Transport in Layered, Variably Saturated Soils Using the Ensemble Kalman Filter. *Water*, 11(7), 1520.
28. Kashuk, S., Mercurio, S. R., & Iskander, M. (2014). Visualization of dyed NAPL concentration in transparent porous media using color space components. *Journal of contaminant hydrology*, 162, 1-16.
29. Kechavarzi, C., Soga, K., & Wiart, P. (2000). Multispectral image analysis method to determine dynamic fluid saturation distribution in two-dimensional three-fluid phase flow laboratory experiments. *Journal of Contaminant Hydrology*, 46(3-4), 265-293.
30. Klotzsche, A., Jonard, F., Looms, M. C., van der Kruk, J., & Huisman, J. A. (2018). Measuring soil water content with ground penetrating radar: a decade of progress. *Vadose Zone Journal*, 17(1).

31. Le Maître, O., & Knio, O. M. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media.
32. Li, C., & Ren, L. (2011). Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter. *Vadose zone journal*, 10(4), 1205-1227.
33. Li, B., Toll, D., Zhan, X., & Cosgrove, B. (2012). Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint. *Hydrology and Earth System Sciences*, 16(1), 105-119.
34. Li, H., & Zhang, D. (2007). Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. *Water Resources Research*, 43(9).
35. Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., ... & Van Velzen, N. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities.
36. Lu, H., Yu, Z., Zhu, Y., Drake, S., Hao, Z., & Sudicky, E. A. (2011). Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. *Advances in Water Resources*, 34(3), 395-406.
37. Man, J., Zheng, Q., Wu, L., & Zeng, L. (2019). Improving parameter estimation with an efficient sequential probabilistic collocation-based optimal design method. *Journal of hydrology*, 569, 1-11.
38. McBride, D., Cross, M., Croft, N., Bennett, C., & Gebhardt, J. (2006). Computational modelling of variably saturated flow in porous media with complex three-dimensional geometries. *International journal for numerical methods in fluids*, 50(9), 1085-1117.
39. Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H. J. H., Canty, M., & Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of hydrology*, 399(3-4), 410-421.
40. Moonen, P., & Allegrini, J. (2015). Employing statistical model emulation as a surrogate for CFD. *Environmental Modelling & Software*, 72, 77-91.
41. Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in water resources*, 28(2), 135-147.
42. Moret-Fernández, D., Peña-Sancho, C., Latorre, B., Pueyo, Y., & López, M. V. (2017). Estimating the van Genuchten retention curve parameters of undisturbed soil from a single upward infiltration measurement. *Soil research*, 55(7), 682-691.
43. Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water resources research*, 12(3), 513-522.
44. Paulus, R., Dewals, B. J., Erpicum, S., Piroton, M., & Archambeau, P. (2013). Innovative modelling of 3D unsaturated flow in porous media by coupling independent models for vertical and lateral flows. *Journal of Computational and Applied Mathematics*, 246, 38-51.
45. Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2015). Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. *Journal of Hydrology*, 520, 101-122.

46. Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2018). Model-data interaction in groundwater studies: Review of methods, applications and future directions. *Journal of hydrology*, 567, 457-477.
47. Rajabi, M. M. (2019). Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and Gaussian process emulation. *Stochastic environmental research and risk assessment*, 33(2), 607-631.
48. Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1), 103-114.
49. 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).
50. Richards L.A. (1931). Capillary conduction of liquids through porous medium. *Physics*, 1:318-33.
51. Ross, P. J. (2006). Fast solution of Richards' equation for flexible soil hydraulic property descriptions. *Land and Water Technical Report, CSIRO*, 39(06).
52. Scharnagl, B., Vrugt, J. A., Vereecken, H., & Herbst, M. (2011). Bayesian inverse modelling of in situ soil water dynamics: using prior information about the soil hydraulic properties. *Hydrology & Earth System Sciences Discussions*, 8(1).
53. Schaap, M. G., Feike, J. L., van Genuchten, M. T. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251: 163–176.
54. Shi, L., Song, X., Tong, J., Zhu, Y., & Zhang, Q. (2015). Impacts of different types of measurements on estimating unsaturated flow parameters. *Journal of Hydrology*, 524, 549-561.
55. Simunek, J., Huang, K., & Van Genuchten, M. T. (1995). The SWMS_3D code for simulating water flow and solute transport in three-dimensional variably-saturated media. *US Salinity Laboratory Agricultural Research Service*, (139).
56. Šimunek, J., Van Genuchten, M. T., & Šejna, M. (2012). The HYDRUS software package for simulating the two-and three-dimensional movement of water, heat, and multiple solutes in variably-saturated porous media. *Technical manual, version, 2*, 258.
57. Song, X., Shi, L., Ye, M., Yang, J., & Navon, I. M. (2014). Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling. *Vadose Zone Journal*, 13(2).
58. Su, S. L., Singh, D. N., & Baghini, M. S. (2014). A critical review of soil moisture measurement. *Measurement*, 54, 92-105.
59. Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7), 964-979.
60. Tidwell, V. C., & Glass, R. J. (1994). X ray and visible light transmission for laboratory measurement of two-dimensional saturation fields in thin-slab systems. *Water Resources Research*, 30(11), 2873-2882.

61. Tran, A. P., Vanclooster, M., Zupanski, M., & Lambot, S. (2014). Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter. *Water resources research*, 50(4), 3131-3146.
62. Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil science society of America journal*, 44(5), 892-898.
63. Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., ... & Schaap, M. G. (2017). Pedotransfer functions in Earth system science: Challenges and perspectives. *Reviews of Geophysics*, 55(4), 1199-1256.
64. Vereecken, H., Huisman, J. A., Bogaen, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water resources research*, 44(4).
65. Wang, Y., Shi, L., Zha, Y., Li, X., Zhang, Q., & Ye, M. (2018). Sequential data-worth analysis coupled with Ensemble Kalman Filter for soil water flow: A real-world case study. *Journal of Hydrology*.
66. Xu, T., & Gómez-Hernández, J. J. (2016). Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. *Water Resources Research*, 52(8), 6587-6595.
67. Yoshimoto, N., Orense, R. P., Tanabe, F., Kikkawa, N., Hyodo, M., & Nakata, Y. (2011). Measurement of degree of saturation on model ground by digital image processing. *Soils and foundations*, 51(1), 167-177.
68. Younes, A., Zaouali, J., Fahs, M., Slama, F., Grunberger, O., & Mara, T. A. (2018). Bayesian soil parameter estimation: Results of percolation-drainage vs infiltration laboratory experiments. *Journal of hydrology*, 565, 770-778.
69. Zhang, Y. and Schaap, M.G., 2017. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, 547, pp.39-53.
70. Zhu, Y., Shi, L., Lin, L., Yang, J., & Ye, M. (2012). A fully coupled numerical modeling for regional unsaturated-saturated water flow. *Journal of hydrology*, 475, 188-203.
71. Zhu, P., Shi, L., Zhu, Y., Zhang, Q., Huang, K., & Williams, M. (2017). Data assimilation of soil water flow via ensemble Kalman filter: Infusing soil moisture data at different scales. *Journal of Hydrology*, 555, 912-925.
72. Zyvoloski, G. (2007). FEHM: A control volume finite element code for simulating subsurface multi-phase multi-fluid heat and mass transfer. *Los Alamos Unclassified Report LA-UR-07-3359*.

Supplementary Material

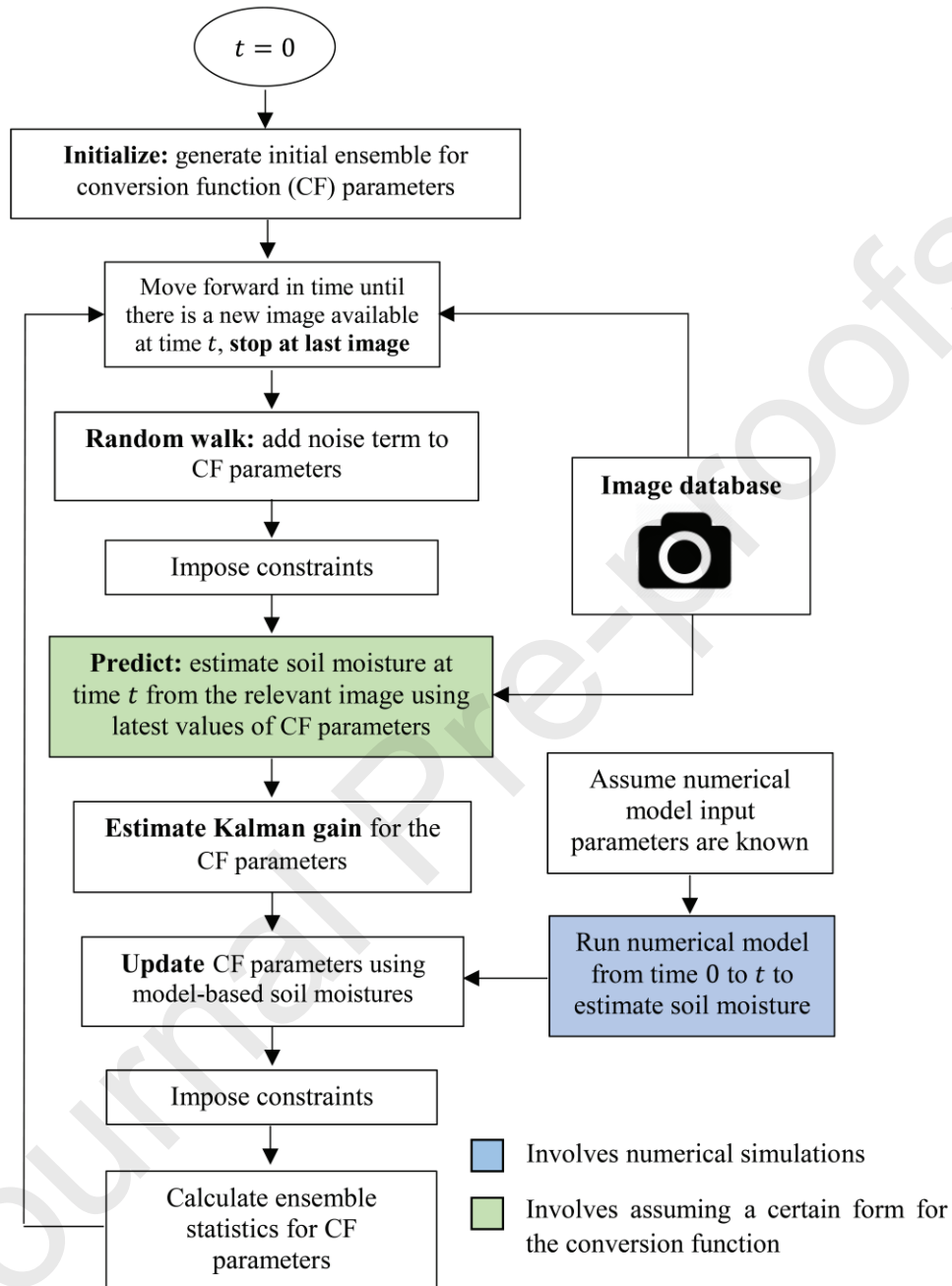


Fig. S1. Algorithm flowchart for estimation of conversion function parameters with the CRD-EnKF algorithm

Based on: {

- Light intensities
- Averaged light intensities
- Direct measurements of water content

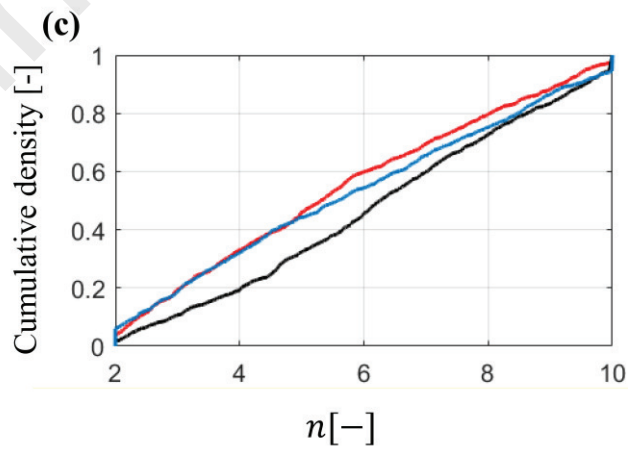
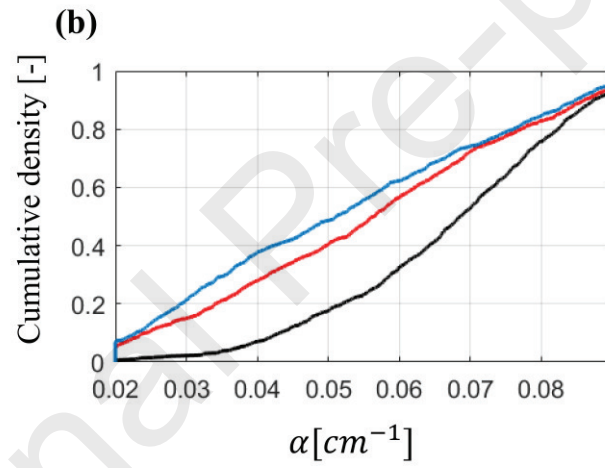
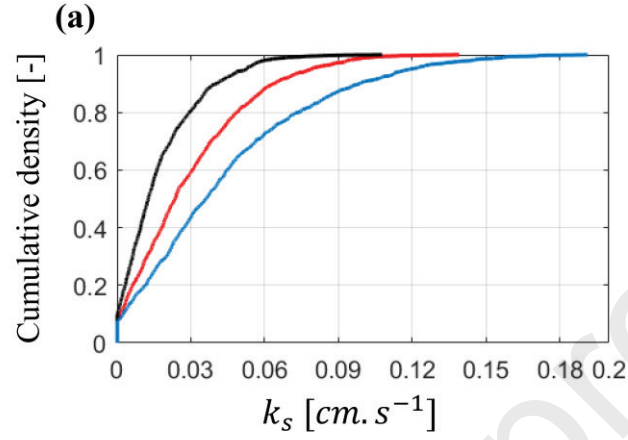


Fig. S2. Comparison of the cumulative distribution functions of the unsaturated flow parameters at the last time step obtained using different types of data.