12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Soil Moisture Estimation for Large-scale Agro-hydrological Systems with Model Mismatch

Zhuangyu Liu [*,**] **Xiaoli Luan** [*] **Jinfeng Liu** [**] **Shunyi Zhao** [*]
Fei Liu [*] **Haiying Wan** [*]

[*] *Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Wuxi 214122, China (e-mail: liuzy95@stu.jiangnan.edu.cn, xlluan@jiangnan.edu.cn, shunyi@jiangnan.edu.cn, fliu@jiangnan.edu.cn, whywan@jiangnan.edu.cn)*
[**] *Department of Chemical & Materials Engineering, University of Alberta, Edmonton, AB, T6G 1H9, Canada. (e-mail: zhuangyu@ualberta.ca, jinfeng@ualberta.ca)*

Abstract: Developing a precise irrigation control strategy is essential for improving water use efficiency, and this requires accurate soil moisture information. However, certain challenges associated with state estimation must be addressed when dealing with large-scale fields. For instance, a vast farmland may be composed of different types of soil, making it challenging to obtain accurate parameters. Consequently, model mismatch becomes inevitable for agro-hydrological systems. In this study, we focus on addressing the issue of state estimation under such circumstance. A high dimensional nonlinear system is obtained by discretizing a 3D polar Richards equation that characterizes water movement dynamics. The proposed approach represents model mismatch as unknown inputs (UIs) relative to the state equations. To reduce computational complexity, a recursive expectation-maximization (EM) approach is modified from the existing batch EM algorithm to identify the UIs. The extended Kalman filter (EKF) is applied to calculate the posterior expectation of the states. Furthermore, an appropriate set of sensors is chosen to ensure complete observability of the system. The simulation results demonstrate the efficacy of the proposed estimation method.

Keywords: 3D agro-hrdological system, recursive EM algorithm, EKF, sensor placement

1. INTRODUCTION

Population growth and climate change have brought about water and food shortages, making it crucial to reduce the water usage in agriculture. According to relevant statistics, agricultural water consumption accounts for the utilization of about 70% of the world's accessible freshwater resources, with irrigation emerging as the principal and most substantial consumer in this regard. Regrettably, the global mean efficiency in water consumption for irrigation stands at a mere 50% Fischer et al. (2007), signifying a significant profligacy of this vital resource. In light of the escalating water scarcity issue, there exists an imperative requirement to ameliorate the efficiency of water utilization within irrigation practices. Closed-loop irrigation systems present a promising avenue for curtailing the strain on water reserves and augmenting the well-being of crops Mao et al. (2018). The establishment of such closed-loop systems necessitates the availability of precise irrigation models and instantaneous soil humidity data spanning the entirety of the agricultural field, a task that often presents formidable challenges Nahar et al. (2019). Addressing this difficulty involves estimating the soil moisture of the entire field using limited sensor data as a possible solution.

To obtain accurate soil moisture estimates, measurements from sensors alone are not sufficient. Various methods have been proposed for soil water content estimation, for example, the extended Kalman filter (EKF) Agyeman et al. (2021); Lü et al. (2011), the ensemble Kalman filter Moradkhani et al. (2005); Chen et al. (2015), the sequential Monte Carlo filter Montzka et al. (2011); Pan et al. (2008), and the moving horizon estimation (MHE) Bo et al. (2020); Bo and Liu (2020). Nonetheless, the precision of these estimations remains contingent upon the caliber of the agro-hydrological model, a metric inevitably influenced by model mismatch arising from the intricate characteristics of soil parameters. The soil water dynamics are elucidated through the utilization of the Richards equation, a partial differential equation employed for the comprehensive depiction of soil water content in relation to spatial and temporal parameters, with particular emphasis on its application in scenarios characterized by non-steady flow conditions. Within the scope of this investigation, we deploy a center pivot irrigation system, renowned for its efficacy in irrigating expansive agricultural plots. To encap-

⋆ This work was supported by the National Natural Science Foundation of China (grant number 61991402, 61973136, 61833007) and the China Scholarship Council.

sulate the intricacies of water flow dynamics, we employ the three-dimensional polar formulation of the Richards equation. Nonetheless, ascertaining the parameters governing a hydrological system remains a formidable task, largely due to the multifaceted nature of environmental conditions. Consequently, the primary aim of this study is to formulate a state estimation methodology capable of mitigating the challenges posed by model mismatch inherent to agro-hydrological systems, thereby furnishing reliable soil moisture estimations.

In our previous work, we addressed the challenge of state estimation for a 1D agro-hydrological system with model mismatch Liu et al. (2023). We proposed the recursive EM algorithm to estimate the states and unknown inputs, and utilized the sensitivity matrix to determine the optimal sensor placement. In this study, we extend the system from 1D to 3D, creating a large system with thousands of states. Due to the size of this system, the sensitivity analysis method is not applicable. Therefore, the PBH test method is adopted in this work. To summarize, the research contributions can be encapsulated as follows:

- To accurately model the agro-hydrological systems is challenging due to the difficulty in determining soil property parameters and the significant impact of external environmental factors on the system. This study is dedicated to the resolution of the state estimation challenge of 3D agro-hydrological systems which is a high dimension system afflicted by model mismatch.
- The expenses incurred in deploying sensors at every node within an irrigation field can present formidable financial constraints. Consequently, the conundrum of sensor placement assumes paramount importance in the realm of state estimation for agro-hydrological systems. In addressing this challenge, a sensor selection approach grounded in the modal degree of observability is employed, ensuring that the chosen set of sensors attains the observability criterion across the entire system.
- The conventional EM algorithm is modified into a recursive EM algorithm in order to estimate states and model mismatch in real-time with computational efficiency. The recursive EM algorithm is particularly well-suited for high-dimensional systems, such as agro-hydrological systems, making it a practical and efficient choice for these types of applications.

2. AGRO-HYDROLOGICAL SYSTEM AND PROBLEM FORMULATION

2.1 System Model Description

The scope of this study is focused on an agro-hydrological system, which entails the transfer of water among plants, soil, and the external environment. An illustration of this system is presented in Figure 1. The movement of water in the soil can be represented through the utilization of the Richards equation Richards (1931):

$$\frac{\partial \theta_v}{\partial t} = C(h)\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right] - S\left(h,z\right) \quad (1)$$

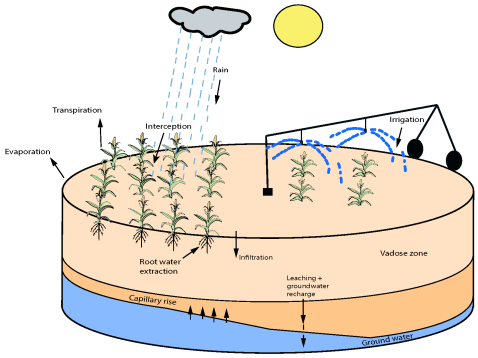


Fig. 1. Schematic diagram of a polar agro-hydrological system Agyeman et al. (2021).

where $K(h)$ and $C(h)$ are formulated in Van Genuchten (1980). The explanations of the variables are listed below.

Table 1. Elements of Richards equation

Parameter	Illustration	Unit
θ_v	moisture content	m/m
z	vertical position	m
h	capillary pressure head	m
$K(h)$	unsaturated hydraulic conductivity	m/s
$C(h)$	capillary capacity	m^{-1}
$S(h,z)$	root water extraction rate	$m^3m^{-3}s^{-1}$

2.2 3D Coordinate Richards Equation

To incorporate the rotational motion of the center pivot into the model, we adopt the cylindrical coordinate version of the Richards equation to simulate a field that includes a center pivot irrigation system. The Richards equation can be expressed in cylindrical coordinates as Agyeman et al. (2021):

$$\begin{aligned} C(h)\frac{\partial h}{\partial t} =& \frac{1}{r}\frac{\partial}{\partial r}\left[rK(h)\frac{\partial h}{\partial r}\right] + \frac{1}{r}\frac{\partial}{\partial \theta}\left[\frac{K(h)}{r}\frac{\partial h}{\partial \theta}\right] \\ &+ \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right] - S(h,z) \end{aligned} \quad (2)$$

The right-hand side of equation (2) represents the spatial (r,θ,z) derivative of the capillary pressure head, while the left-hand side represents its temporal derivative.

The correlation between θ_v and h is expressed as follows:

$$\theta_v(h) = \theta_r + (\theta_s - \theta_r)\left[\frac{1}{1+(-\alpha h)^n}\right]^{1-\frac{1}{n}} \quad (3)$$

where θ_s and θ_r are the saturated volumetric and residual moisture content, respectively. n and α are the curve-fitting soil hydraulic properties.

2.3 Finite Difference Model

Due to the nonlinearity of equation (2), it is challenging to derive an analytical solution, so we turn to a numerical

technique represented in Agyeman et al. (2021) to solve the partial differential equation. The specific approximation procedure is concluded by the following.

Firstly, the finite difference method is utilized to discretize the derivatives of equation (2) with respect to the spatial properties (r, θ, z). Subsequently, the computational technique of the backward Euler method was applied for the determination of the temporal derivatives.

Equation (2) is solved using numerical methods with the following boundary conditions, which are suitable for crop fields with a center pivot irrigation system:

$$\begin{aligned}
\left.\frac{\partial h(r,\theta,z,t)}{\partial r}\right|_{(r=0,\theta,z)} &= 0 \\
\left.\frac{\partial (h(r,\theta,z,t))}{\partial z}\right|_{(r,\theta,z=H_z)} &= -1 - \frac{u_{\text{irr}}}{K(h)} \\
\left.\frac{\partial h(r,\theta,z,t)}{\partial r}\right|_{(r=H_r,\theta,z)} &= 0 \\
\left.\frac{\partial (h(r,\theta,z,t)+z)}{\partial z}\right|_{(r,\theta,z=0)} &= 1 \\
\left.\frac{\partial h(r,\theta,z,t)}{\partial \theta}\right|_{(r=0,\theta,z)} &= 0 \\
h(r,\theta=0,z,t) &= h(r,\theta=2\pi,z,t)
\end{aligned} \tag{4}$$

where H_r denotes the overall radius of the field, H_z represents the height of the field column, and u_{irr} (m/s) is the manipulation variable as the irrigation rate.

The nonlinear state space model for the agro-hydrological system can be expressed as follows:

$$\dot{x} = f_c(x(t), u(t)) + \omega(t) \tag{5}$$

where $x(t) \in \mathbb{R}^{N_x}$ is the state vector, $u(t)$ represents the irrigation amount which is the manipulation input of this system, $\omega(t)$ denotes the process noise to the system.

2.4 Problem formulation

In the scope of this study, we recognize the existence of model mismatch, which may be construed as the incorporation of additional UIs into the dynamics of the system. Fundamentally, the authentic representation of the system can be expounded as follows:

$$\dot{x} = f_c(x(t), u(t)) + M_c a(t) + \omega(t) \tag{6}$$

where $a(t)$ comprises the vector of UIs that signifies the presence of model mismatch. The measurement function is detailed as:

$$y(t) = Cx(t) + v(t) \tag{7}$$

where C is the matrix which relates to the sensor location. $v(t)$ is the measurement noise. Then we can obtain the discrete model by discretizing equations (6) and (7):

$$\begin{aligned}
x_{k+1} &= f(x_k, u_k) + Ma_k + \omega_k \\
y_{k+1} &= Hx_{k+1} + v_{k+1}
\end{aligned} \tag{8}$$

The aim of this research is to formulate a computationally efficient method for estimating the state, specifically capillary pressure head, even when there is a presence of model mismatch represented by unidentified additive inputs.

3. PROPOSED EKF BASED RECURSIVE EM ALGORITHM

Within the framework of the model expounded in the preceding section, the overarching objective is to jointly infer the states and UIs in order to address the challenge of state estimation in agro-hydrological systems characterized by model mismatch. In light of the intricacies inherent to high-dimensional agro-hydrological systems, we have introduced a tailored adaptation of the classical EM algorithm, imbuing it with a recursive essence to navigate the inherent complexities. The ensuing section delineates an elaborate exposition of this methodology, accompanied by a set of notations to be employed throughout this discourse: $x_0, \cdots, x_N$ denotes the array of states, while $y_1, \cdots, y_N$ signifies the collection of measurements, and $a_{0:N}$ constitutes the ensemble of UIs. In this context, $\mathbb{E}$ represents the mathematical expectation, Tr signifies the matrix trace, and $\mathbb{N}(\mu, P)$ denotes the Gaussian probability function featuring the mean μ and covariance P. The superscripts '$\wedge$' and '$\vee$' are indicative of the estimation and prediction aspects, respectively.

Initially, we shall provide a concise overview of the conventional EM algorithm, often denoted as the batch EM algorithm. Rooted in the principles of maximum likelihood, the EM algorithm stands as an iterative optimization methodology employed for the purpose of estimating parameter values and latent variable expectations. The batch EM algorithm encompasses the subsequent procedural stages:

- Giving the UIs an initial guess $a^{(0)}$
- For time instants $k = 0, 1, 2, \ldots$, perform:
  - E-step: compute $Q\left(a, a^{(k)}\right)$.
  - M-step: compute $a^{(k+1)} = \arg\max_a Q\left(a, a^{(k)}\right)$.

The Q-function denotes the expectation of the log-likelihood and can be elucidated within the context of the Markov-chain structure inherent to the state-space model. To be precise, the log-likelihood function can be calculated as:

$$\begin{aligned}
L_{0:N} &= \log p\left(x_0, \ldots, x_N, y_1, \ldots, y_N \mid a_{0:N}\right) \\
&= \log p\left(x_0 \mid a_0\right) + \sum_{k=1}^{N} \log p\left(x_k \mid x_{k-1}, a_{k-1}\right) \\
&\quad + \sum_{k=1}^{N} \log p\left(y_k \mid x_k\right)
\end{aligned} \tag{9}$$

The batch EM algorithm is not ideal for real-time applications and large scale systems due to its computational complexity, as it requires processing all historical data in each optimization implementation Liu et al. (2022). This impracticality is particularly evident when applying the batch EM algorithm to estimate the state and UIs in high-dimensional agro-hydrological processes. However, the complexity of the EM algorithm can be reduced by using a recursive approach to calculate the Q-function.

The focus of this paper is on proposing a recursive EM algorithm that utilizes an EKF-based strategy to achieve online state estimation and identification of UIs. By significantly reducing computational complexity, the proposed approach becomes practical for real-time applications and large-scale systems.

3.1 Recursive Q-function

Within this section, we shall expound upon the process of deducing a recursive manifestation of the Q-function. In circumstances where data accrues sequentially over time, one can harness the collected samples and the pre-established parameters to undertake an E-step with each individual measurement. This inherent characteristic serves as the fundamental underpinning for the ensuing derivation. The Q-function, employed for the computation of the expectation in equation (9), can be expressed as follows:

$$\begin{aligned} Q_N\left(a, a_N^{\text{old}}\right) =& \mathbb{E}(\log p\left(x_0 \mid a_0\right) \\ &+ \sum_{k=1}^{N-1} \log p\left(x_k \mid x_{k-1}, a_{k-1}\right) \\ &+ \sum_{k=1}^{N-1} \log p\left(y_k \mid x_k\right) \\ &+ \log p\left(x_N \mid x_{N-1}, a_{N-1}\right) \\ &+ \log p\left(y_N \mid x_N\right)) \end{aligned} \tag{10}$$

In this equation, the subscript N signifies the current temporal moment, whereas a represents the set of UIs slated for estimation at the present time step. The variable a_N^{old} pertains to the solution derived at the time index $N-1$, which plays a pivotal role in computing the posterior expectation at the current temporal instance, denoted as N. The superscript 'old' denotes that the estimation of UIs was performed during last time step. Consequently, by calculating the item of the last time step separately, we can obtain:

$$\begin{aligned} Q_N\left(a, a_N^{\text{old}}\right) =& Q_{N-1}\left(a, a_{N-1}^{\text{old}}\right) \\ &+ \mathbb{E}(\log p\left(x_N \mid x_{N-1}, a_{N-1}\right) \\ &+ \log p\left(y_N \mid x_N\right)) \end{aligned} \tag{11}$$

Equation (11) can be classified as quasi-recursive, given that the E-step exhibits consistency across various temporal indices. By encapsulating the temporal mean of Q_N in equation (11), we can obtain:

$$\tilde{Q}_N\left(a, a_N^{\text{old}}\right) = \frac{1}{N} Q_N\left(a, a_N^{\text{old}}\right) \tag{12}$$

Upon substituting equation (12) into equation (11), we derive the subsequent expression for the recursive likelihood of the complete dataset:

$$\begin{aligned} \tilde{Q}_N\left(a, a_N^{\text{old}}\right) =& (1-\frac{1}{N})\tilde{Q}_{N-1}\left(a, a_{N-1}^{\text{old}}\right) \\ &+ \frac{1}{N}\mathbb{E}(\log p\left(x_N \mid x_{N-1}, a_{N-1}\right) \\ &+ \log p\left(y_N \mid x_N\right)) \end{aligned} \tag{13}$$

While the inherent step size naturally diminishes over time, it is not the optimal strategy. To counter this issue, we introduce a fixed step-size, analogous to a learning rate, as a hyperparameter that can be tailored to replace the diminishing factor of $1/N$. Drawing upon principles of stochastic approximation, as expounded in Chen et al. (2020), we incorporate an artificial step-size denominated as γ_N to supplant $1/N$ in equation (13). Subsequently, the recursive computation of the expectation can be articulated as follows:

$$\begin{aligned} \tilde{Q}_N\left(a, a_N^{\text{old}}\right) =& (1-\gamma_N)\tilde{Q}_{N-1}\left(a, a_{N-1}^{\text{old}}\right) \\ &+ \gamma_N\mathbb{E}(\log p\left(x_N \mid x_{N-1}, a_{N-1}\right) \\ &+ \log p\left(y_N \mid x_N\right)) \end{aligned} \tag{14}$$

Equation (14) can be further elaborated by considering all the components in the calculation as:

$$\begin{aligned} \tilde{Q}_N\left(a, a_N^{\text{old}}\right) =& \prod_{t=2}^{N}\left(1-\gamma_t\right)\mathbb{E}(G_1) \\ &\sum_{k=2}^{N-1}\left[\prod_{t=k+1}^{N}\left(1-\gamma_t\right)\right]\gamma_k\mathbb{E}(G_k) \\ &+ \gamma_N\mathbb{E}(G_N) \end{aligned} \tag{15}$$

where

$$\begin{aligned} \mathbb{E}\left(G_1\right) =& \mathbb{E}\left(\log p\left(x_0 \mid a_0\right) + \log p\left(x_1 \mid x_0, a_0\right)\right. \\ &\left. + \log p\left(y_1 \mid x_1\right)\right) \end{aligned} \tag{16}$$

$$\mathbb{E}\left(G_k\right) = \mathbb{E}\left(\log p\left(x_k \mid x_{k-1}, a_{k-1}\right) + \log p\left(y_k \mid x_k\right)\right) \tag{17}$$

$$\begin{aligned} \mathbb{E}\left(G_N\right) =& \mathbb{E}\left(\log p\left(x_N \mid x_{N-1}, a_{N-1}\right)\right. \\ &\left. + \log p\left(y_N \mid x_N\right)\right) \end{aligned} \tag{18}$$

To perform the calculations outlined in equations (16)-(18), it is imperative to possess the conditional probability density functions of $p\left(x_0 \mid a_0\right)$, $p\left(x_k \mid x_{k-1}, a_{k-1}\right)$, and $p\left(y_k \mid x_k\right)$. These functions can be expressed as follows, with the underlying assumption of Gaussian-distributed system noise and measurement noise:

$$p\left(x_0 \mid a_0\right) = \mathbb{N}\left(\hat{x}_0, P_0\right) \tag{19}$$

$$\begin{aligned} p\left(x_k \mid x_{k-1}, a_{k-1}\right) =& \mathbb{N}(f(x_{k-1}, u_{k-1}) + Ma_{k-1}, \\ & Q_k) \end{aligned} \tag{20}$$

$$p\left(y_k \mid x_k\right) = \mathbb{N}\left(Hx_k, R_k\right) \tag{21}$$

Subsequently, we can proceed to deduce the log-likelihood for the aforementioned equations (19)-(21) in the following manner:

$$\begin{aligned} \log p\left(x_0 \mid a_0\right) =& -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left|P_0\right| \\ &- \frac{1}{2}\mathscr{D}\left(x_0 - \hat{x}_0, P_0\right) \end{aligned} \tag{22}$$

$$\begin{aligned}\log p\left(x_k \mid x_{k-1}, a_{k-1}\right) = &-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|Q_k| \\ &-\frac{1}{2}\mathscr{D}\left(x_k - f\left(x_{k-1}, u_{k-1}\right)\right. \\ &\left.-Ma_{k-1}, Q_k\right)\end{aligned} \tag{23}$$

$$\begin{aligned}\log p\left(y_k \mid x_k\right) = &-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|R_k| \\ &-\frac{1}{2}\mathscr{D}\left(y_k - Hx_k, R_k\right)\end{aligned} \tag{24}$$

here the notation $\mathscr{D}$ is utilized as a shorthand to represent the matrix operation defined as $\mathscr{D}(x, P) = x^{\mathrm{T}}P^{-1}x$.

The ensuing statements denote the expected value of the likelihood function:

$$\mathbb{E}\left(\log p\left(x_0 \mid a_0\right)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|P_0| - \frac{n}{2} \tag{25}$$

$$\begin{aligned}\mathbb{E}\left(\log p\left(x_k \mid x_{k-1}, a_{k-1}\right)\right) = &-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|Q_k| \\ &-\frac{1}{2}\operatorname{Tr}\left\{Q_k^{-1}\left\{\mathscr{C}\left(\hat{x}_k\right.\right.\right. \\ &-f\left(\hat{x}_{k-1}, u_{k-1}\right) - Ma_{k-1}\Big) \\ &+\left(\hat{P}_k - F_k\check{P}_k\right) \\ &\left.\left.\left.-\left(\check{P}_k{F_k}^{\mathrm{T}} - F_k\hat{P}_{k-1}{F_k}^{\mathrm{T}}\right)\right\}\right\}\end{aligned} \tag{26}$$

$$\begin{aligned}\mathbb{E}\left(\log p\left(y_k \mid x_k\right)\right) = &-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|R_k| \\ &-\frac{1}{2}\operatorname{Tr}\left\{R_k^{-1}\left[\mathscr{C}\left(y_k - H\hat{x}_k\right)\right.\right. \\ &\left.\left.+\left(H\hat{P}_kH^{\mathrm{T}}\right)\right]\right\}\end{aligned} \tag{27}$$

In this context, F_k signifies the Jacobian matrix associated with $f(\cdot)$ at the time step denoted by k. Additionally, the symbol $\mathscr{C}$ is employed to abbreviate a matrix operation $\mathscr{C}(x) = xx^{\mathrm{T}}$.

Upon replacing the expressions in equations (25)-(27) with those in equations (16)-(18), we derive the following:

$$\begin{aligned}\mathbb{E}\left(G_k\right) = &-\frac{m+n}{2}\log(2\pi) - \frac{1}{2}(\log|Q_k| + \log|R_k|) \\ &-\frac{1}{2}\log\left|\hat{P}_k\right| - \frac{1}{2}\operatorname{Tr}\left\{Q_k^{-1}\right. \\ &\left\{\mathscr{C}\left(\hat{x}_k - f(\hat{x}_{k-1}, u_{k-1}) - Ma_{k-1}\right)\right. \\ &+\left(\hat{P}_k - F\check{P}_k\right) \\ &\left.-\left(\check{P}_kF^{\mathrm{T}} - F\hat{P}_{k-1}F^{\mathrm{T}}\right)\right\}\Big\} \\ &-\frac{1}{2}\operatorname{Tr}\left\{R_k^{-1}\left\{\left(H\hat{P}_kH^{\mathrm{T}}\right)\right.\right. \\ &\left.\left.+\mathscr{C}\left(y_k - H\hat{x}_k\right)\right\}\right\}\end{aligned} \tag{28}$$

The computation of $\mathbb{E}(G_1)$ and $\mathbb{E}(G_N)$ is akin to that of $\mathbb{E}(G_k)$, and the specific methodology is omitted for brevity. Within equation (15), the term $\prod_{t=k+1}^{N}(1-\gamma_t)$ represents the cumulative product of step sizes ranging from time $k+1$ to N, where t designates the respective time index. This concludes the derivation process for the recursive Q-function.

Algorithm 1 EKF based REM Algorithm

Input: y_N
Output: $\hat{x}_N$, a_N
1: Initialize the guess of states and unknown inputs: a_1
2: **for** $N = 2 : k$ **do**
3: **E-step** Update the recursive Q-function:
$\tilde{Q}_N\left(a, a_N^{\text{old}}\right) = (1-\gamma_N)\tilde{Q}_{N-1}\left(a, a_{N-1}^{\text{old}}\right) + \gamma_N\mathbb{E}(\log p\left(x_N \mid x_{N-1}, a_{N-1}\right) + \log p\left(y_N \mid x_N\right))$
4: **Prediction**
$\check{x}_N = f\left(\hat{x}_{N-1}, u_{N-1}\right) + Ma_{N-1}$
$\check{P}_N = F\hat{P}_{N-1}F^{\mathrm{T}} + Q_{N-1}$
5: **Update**
$K_N = \check{P}_NH^{\mathrm{T}}\left(H\check{P}_NH^{\mathrm{T}} + R_N\right)^{-1}$
$\hat{x}_N = \check{x}_N + K_N\left(y_N - H\check{x}_N\right)$
$\hat{P}_N = \left(I - K_NH\right)\check{P}_N$
6: **M-step** Maximize the expectaion: $\frac{\partial\tilde{Q}_N\left(a, a_N^{\text{old}}\right)}{\partial a_N} = 0$
$a_N = (1-\gamma_N)\,a_{N-1} + \gamma_N\left(M^{-1}\left(\hat{x}_N - f\left(\hat{x}_{N-1}, u_{N-1}\right)\right)\right)$
7: **end for**

3.2 State and Covariance Prediction and Update

In order to compute equation (15), an evaluation of the posterior probability and prediction for x_N is required, along with its corresponding P_N. Previous research has employed smoothers to determine the posterior distribution within the Q-function, as exemplified in citations such as Lan et al. (2013) and Khan et al. (2019). However, while smoothers offer an effective method for batch processing in state estimation, their practicality for real-time computation is limited. In light of this, our study adopts the EKF for state estimation, given its recursive computational capacity. The EKF demonstrates proficiency in providing suboptimal state estimates, particularly for nonlinear stochastic systems. The EKF's inherent recursive calculation ability aligns it with seamless integration into the REM algorithm, and its representation is as follows:

Prediction:

$$\check{x}_N = f\left(\hat{x}_{N-1}, u_{N-1}\right) + Ma_{N-1} \tag{29}$$

$$\check{P}_N = F\hat{P}_{N-1}F^{\mathrm{T}} + Q_{N-1} \tag{30}$$

Update:

$$K_N = \check{P}_NH^{\mathrm{T}}\left(H\check{P}_NH^{\mathrm{T}} + R_N\right)^{-1} \tag{31}$$

$$\hat{x}_N = \check{x}_N + K_N\left(y_N - H\check{x}_N\right) \tag{32}$$

$$\hat{P}_N = \left(I - K_NH\right)\check{P}_N \tag{33}$$

From equation (29) to (33), we can calculate the value of posterior of the states given the UIs estimated at preceding time instant.

3.3 Maximization of recursive EM

In pursuit of the optimal estimation solution, we undertake the computation of partial derivatives of the recursive Q-function in relation to the UIs. Setting this partial derivative to zero facilitates the determination of the local maximum within the UIs vector, denoted as a_k within the model.

$$\frac{\partial \tilde{Q}_N\left(a, a_N^{\text{old}}\right)}{\partial a_N} = 0 \tag{34}$$

Subsequently, through the substitution of equation (15) into equation (34), we can derive

$$\begin{aligned}\frac{\partial \tilde{Q}_N\left(a, a_N^{\text{old}}\right)}{\partial a_N} =& \frac{\partial}{\partial a_N}\Bigg\{\prod_{t=2}^{N}(1-\gamma_t)\,\mathbb{E}(G_1) \\ &+ \sum_{k=2}^{N-1}\left[\prod_{t=k+1}^{N}(1-\gamma_t)\right]\gamma_k \mathbb{E}(G_k) \\ &+ \gamma_N \mathbb{E}(G_N)\Bigg\} = 0\end{aligned} \tag{35}$$

Finally, by calculating the above equation, we can get

$$\begin{aligned}a_N =& (1-\gamma_N)\, a_{N-1} \\ &+ \gamma_N\left(M^{-1}\left(\hat{x}_N - f\left(\hat{x}_{N-1}, u_{N-1}\right)\right)\right)\end{aligned} \tag{36}$$

Hence, the derivation of the M-step within the EKF-based REM algorithm is concluded. The outlined methodology is encapsulated in the pseudocode, as delineated in Algorithm 1.

4. SIMULATIONS

In the present study, we examine a field with the overarching radius, denoted as H_r, measured at 50 meters, and the vertical extent, represented as H_z, height of 0.3 meters. The discretization of this field into a total of $3,840$ nodes is systematically achieved through the application of the finite difference method. This discretization is implemented with a distribution of 6 nodes in the radial direction, 40 nodes in the azimuthal direction, and 16 nodes in the vertical direction, respectively.

In order to identify suitable locations for sensor installation in agricultural fields, we employ the algorithm detailed in Sahoo et al. (2019), which relies on the concept of modal degree of observability. The Popov-Belevitch-Hautus (PBH) test forms the foundation for the modal degree of observability, which assesses a sensor node's ability to estimate other nodes within a system. Figure 2 visually represents the ranked average modal observability. To make sure the whole system for states and UIs are observable, we select 2000 nodes with the highest rank to place sensors.

The efficacy of the EKF-based REM algorithm is empirically substantiated through simulation. The simulation encompasses instances of model mismatch, arising from erroneous estimations of critical parameters such as the crop coefficient (Kc) and evapotranspiration values (Et),

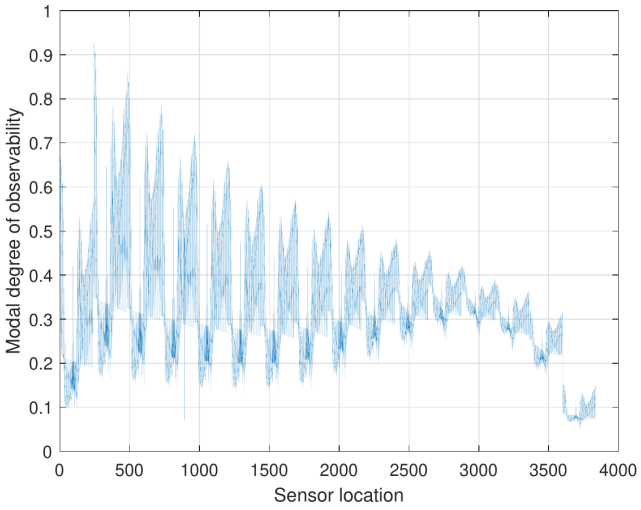


Fig. 2. The average degree of observability calculated for different nodes in the system.

both integral components within the sink term $S(h, z)$. The true values for Kc range from 0.75 to 0.88 for the 7-day simulation period, while the true values for Et range from $0.57mm/day$ to $1.4mm/day$. The guessed values for Kc and Et are fixed 1.08 and $1.4mm/day$, respectively. It is assumed that the parameters for soil properties, for example, the saturated volumetric moisture content θ_s, the saturated hydraulic conductivity K_s, vary across different soil columns. The velocity of the center pivot is $0.022m/s$ and supplies an irrigation amount of $4mm/day$. Figure 3 demonstrates the effectiveness of the proposed algorithm by displaying certain states. The true trajectories of the simulated pressure head are depicted using a red solid line. In this simulation, the openloop states are the response with an incorrect initial guess of the states and error parameters of Kc and Et. As shown in the figure, the discrepancy between the true states and the openloop response is significant. The blue dot-dashed lines in the figure depict the results of the estimation using the proposed EKF-based REM algorithm. The estimates demonstrate excellent performance in tracking the true values of the states, converging to the states within a few days of simulation. The figure also includes estimations obtained solely through the use of the EKF method. It is apparent that the EKF method converges at a slower rate than the proposed EKF-based REM algorithm. When model mismatch occurs, the EKF method exhibits a steady error. In other words, the proposed algorithm provides better estimation performance than the EKF method in the presence of model mismatch.

Given the inherent complexity of the high-dimensional 3D polar agro-hydrological system, presenting simulation results can be somewhat intricate. To create soil moisture maps, we transform pressure head data into moisture content using the equation (3). As showing from Figure 4 to Figure 7, the errors between the actual states and the estimated states decrease as the simulation progresses. This demonstrates that the proposed EKF-based REM algorithm can effectively estimate both the top and bottom layers with good performance.

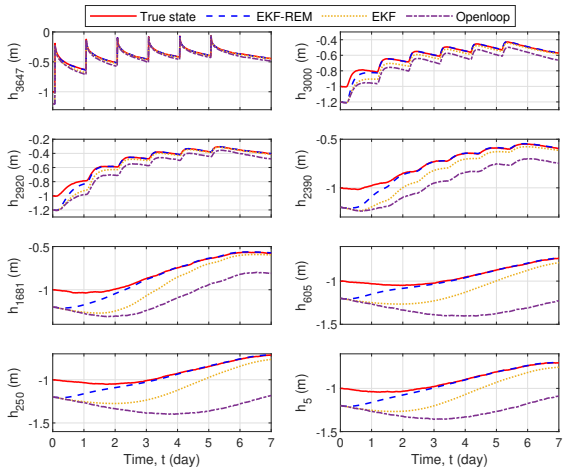
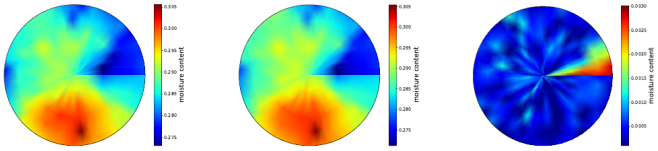
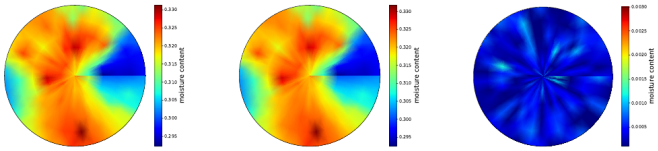


Fig. 3. Trajectories for chosen process states, open-loop state responses incorporating error parameters, and state estimates with proposed algorithm.



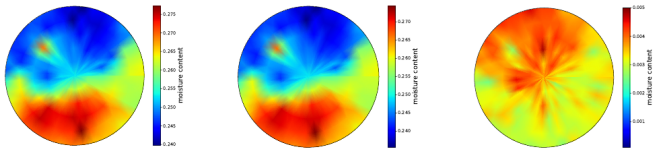
(a) Actual (b) Estimated (c) Error

Fig. 4. Soil moisture maps of the top layer after 1 day simulation



(a) Actual (b) Estimated (c) Error

Fig. 5. Soil moisture maps of the top layer after 3 days simulation

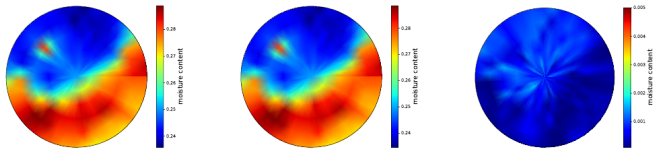


(a) Actual (b) Estimated (c) Error

Fig. 6. Soil moisture maps of root zone (depth of $0.3m$) after 2 days simulation

5. CONCLUSIONS

This paper endeavors to confront the intricate problem of state estimation characterized by model mismatch, an inherent challenge in agro-hydrological systems attributable to the intricate nature of soil properties and the influence of external environmental factors. To mitigate this problem, we propose the EKF-based REM algorithm for implementing state estimation in large-scale agro-hydrological



(a) Actual (b) Estimated (c) Error

Fig. 7. Soil moisture maps of root zone (depth of $0.3m$) after 6 days simulation

systems. Our approach involves adapting the batch EM algorithm into a recursive EM algorithm that enables real-time estimation. Additionally, we utilize the modal degree of observability method to select an appropriate sensor set. Lastly, a soil moisture map is generated during the simulation phase to demonstrate the effectiveness of the proposed algorithm in estimating soil water content.

REFERENCES

Agyeman, B.T., Bo, S., Sahoo, S.R., Yin, X., Liu, J., and Shah, S.L. (2021). Soil moisture map construction by sequential data assimilation using an extended kalman filter. *Journal of Hydrology*, 598, 126425.

Bo, S. and Liu, J. (2020). A decentralized framework for parameter and state estimation of infiltration processes. *Mathematics*, 8(5), 681.

Bo, S., Sahoo, S.R., Yin, X., Liu, J., and Shah, S.L. (2020). Parameter and state estimation of one-dimensional infiltration processes: A simultaneous approach. *Mathematics*, 8(1), 134.

Chen, W., Huang, C., Shen, H., and Li, X. (2015). Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation. *Advances in water resources*, 86, 425–438.

Chen, X., Zhao, S., and Liu, F. (2020). Online identification of time-delay jump markov autoregressive exogenous systems with recursive expectation-maximization algorithm. *International Journal of Adaptive Control and Signal Processing*, 34(3), 407–426.

Fischer, G., Tubiello, F.N., Van Velthuizen, H., and Wiberg, D.A. (2007). Climate change impacts on irrigation water requirements: Effects of mitigation, 1990–2080. *Technological Forecasting and Social Change*, 74(7), 1083–1107.

Khan, M.A.I., Imtiaz, S.A., and Khan, F. (2019). Simultaneous estimation of hidden state and unknown input using expectation maximization algorithm. *Industrial & Engineering Chemistry Research*, 58(26), 11553–11565.

Lan, H., Liang, Y., Yang, F., Wang, Z., and Pan, Q. (2013). Joint estimation and identification for stochastic systems with unknown inputs. *IET Control Theory & Applications*, 7(10), 1377–1386.

Liu, S., Zhang, X., Xu, L., and Ding, F. (2022). Expectation–maximization algorithm for bilinear systems by using the rauch–tung–striebel smoother. *Automatica*, 142, 110365.

Liu, Z., Liu, J., Zhao, S., Luan, X., and Liu, F. (2023). State estimation for one-dimensional agro-hydrological processes with model mismatch. *The Canadian Journal of Chemical Engineering*. doi: https://doi.org/10.1002/cjce.25095.

Lü, H., Yu, Z., Zhu, Y., Drake, S., Hao, Z., and 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.

Mao, Y., Liu, S., Nahar, J., Liu, J., and Ding, F. (2018). Soil moisture regulation of agro-hydrological systems using zone model predictive control. *Computers and Electronics in Agriculture*, 154, 239–247.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.J.H., Canty, M., and 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.

Moradkhani, H., Sorooshian, S., Gupta, H.V., and Houser, P.R. (2005). Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in water resources*, 28(2), 135–147.

Nahar, J., Liu, J., and Shah, S.L. (2019). Parameter and state estimation of an agro-hydrological system based on system observability analysis. *Computers & Chemical Engineering*, 121, 450–464.

Pan, M., Wood, E.F., Wójcik, R., and McCabe, M.F. (2008). Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sensing of Environment*, 112(4), 1282–1294.

Richards, L.A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1(5), 318–333.

Sahoo, S.R., Yin, X., and Liu, J. (2019). Optimal sensor placement for agro-hydrological systems. *AIChE Journal*, 65(12), e16795.

Van Genuchten, M.T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil science society of America journal*, 44(5), 892–898.