# Sequential Ensemble Based Optimal Design for Parameter Estimation in Unsaturated Flow Model


Jun Man[1], Lingzao Zeng[1], Laosheng Wu[2]

[1]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China
[2]Department of Environmental Sciences, University of California, Riverside, CA 92521, USA


## Introduction

- Ensemble Kalman Filter (EnKF) is becoming more popular in estimating soil hydraulic parameters. But most of the previous studies focused mainly on the estimation, while sampling strategies have rarely been systematically investigated.
- Well-designed sampling strategies can improve the estimation accuracy and subsequently, the model prediction.

## System Model

In this work, we consider the one dimensional unsaturated water flow simulated by the Richards equation,

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

The van Genuchten-Mualem (VGM) model is used to characterize the soil water characteristic curves and unsaturated hydraulic conductivity function,

$$S_e = \frac{\theta-\theta_r}{\theta_s-\theta_r} = \begin{cases} \dfrac{1}{(1+|\alpha h|^n)^{1-1/n}}, & h<0 \\ 1, & h \ge 0 \end{cases}$$

$$K(h) = K_s S_e^{0.5}[1-(1-S_e^{1/m})^m]^2, m = 1-1/n$$

## Methodology

### Ensemble Kalman Filter

Forecast Step:

$$\mathbf{x}_{i,j+1}^f = G(\mathbf{x}_{i,j}^a), \quad i = 1,2,\ldots,N_e$$

Analysis Step:

$$\mathbf{x}_{i,j+1}^a = \mathbf{x}_{i,j+1}^f + \mathbf{K}(\mathbf{d}_{i,j+1} - \mathbf{H}\mathbf{x}_{i,j+1}^f), \quad i = 1,2,\ldots,N_e$$

### Information metrics based on the first two-order statistics

Prior mean: b | Prior covariance: B

Posterior mean: a | Posterior covariance: A

1) Shannon entropy difference

$$SD = \ln\det(\mathbf{B})/2 - \ln\det(\mathbf{A})/2 = \ln\det(\mathbf{BA}^{-1})/2$$

2) Relative entropy

$$RE = (\mathbf{a}-\mathbf{b})^T\mathbf{B}^{-1}(\mathbf{a}-\mathbf{b})/2 + [\ln\det(\mathbf{BA}^{-1}) + Tr(\mathbf{AB}^{-1} - \mathbf{I}_n)]/2$$

3) Degrees of freedom for signal

$$DFS = \langle 2J_b \rangle = Tr(\mathbf{I_n} - \mathbf{AB}^{-1}), \; J_b = (\mathbf{a}-\mathbf{b})^T\mathbf{B}^{-1}(\mathbf{a}-\mathbf{b})/2$$

The sampling location with the maximum mean information metric value is chosen as the optimal one.

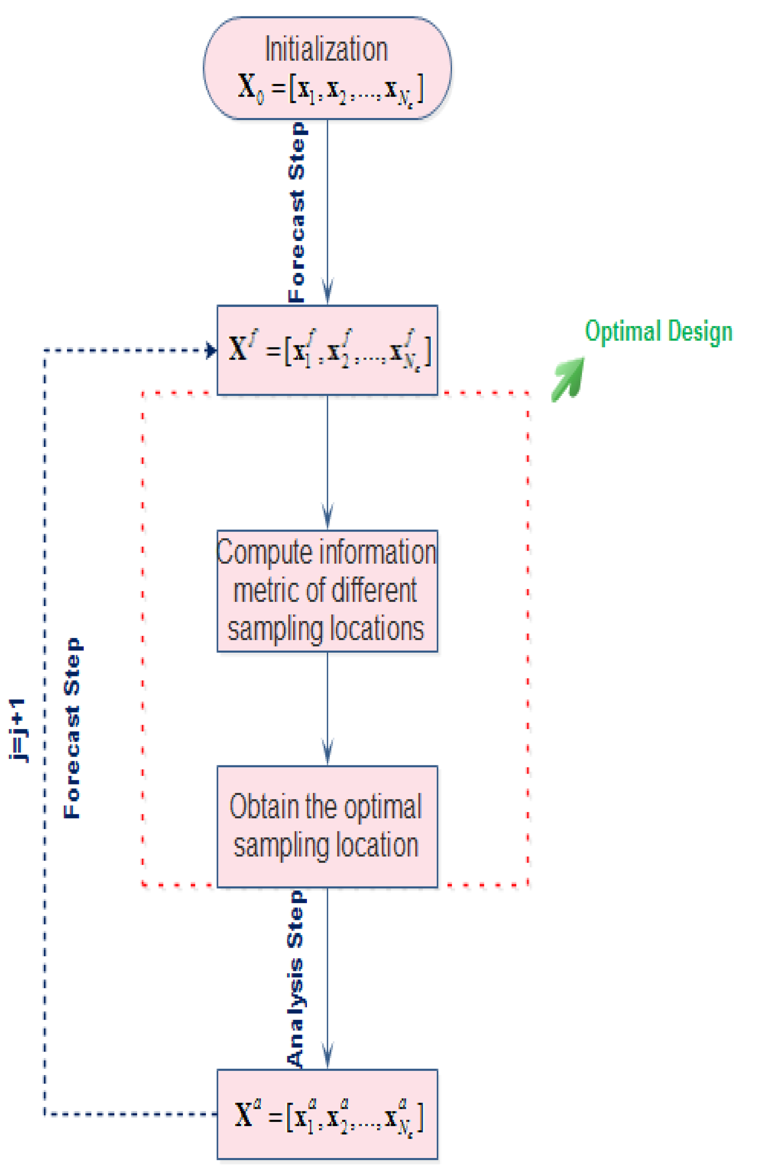
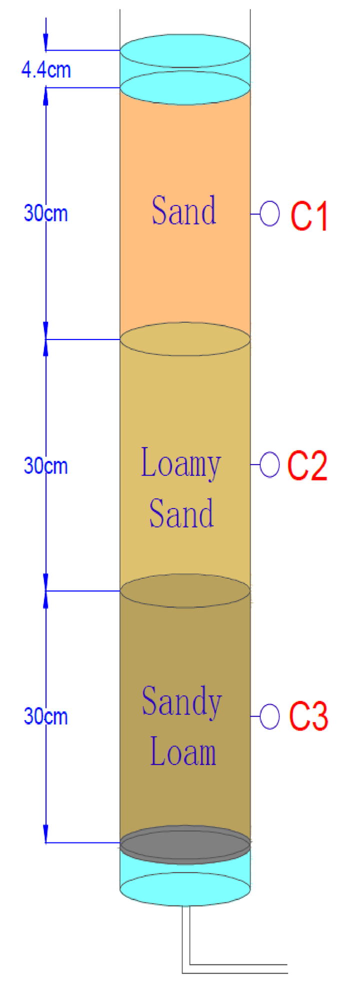



Figure 1. Flowchart of the sequential ensemble based optimal design method.

## Illustrative Example




Initially, the column was saturated with a surface ponding of 4.4 cm. The boundary condition at the bottom of the domain was a constant pressure head of -80 cm. The experiment lasted for 64 hours, measurements were taken every 2 hours.

In this work, $K_s$, $\theta_r$, $\alpha$, and $n$ were assumed unknown and to be estimated from the pressure head at only one location each time. Transformed by Johnson systems, the distributions of transformed parameters were assumed to be Gaussian. Three transformation forms were lognormal (LN), log ratio (SB) and hyperbolic (SU).

$$LN: Y = \ln(p)$$

$$SB: Y = \ln[(p-u)(v-p)]$$

$$SU: Y = \sinh^{-1}[w] = \ln[w+(1+w^2)^{1/2}], w = (p-u)/(v-u)$$

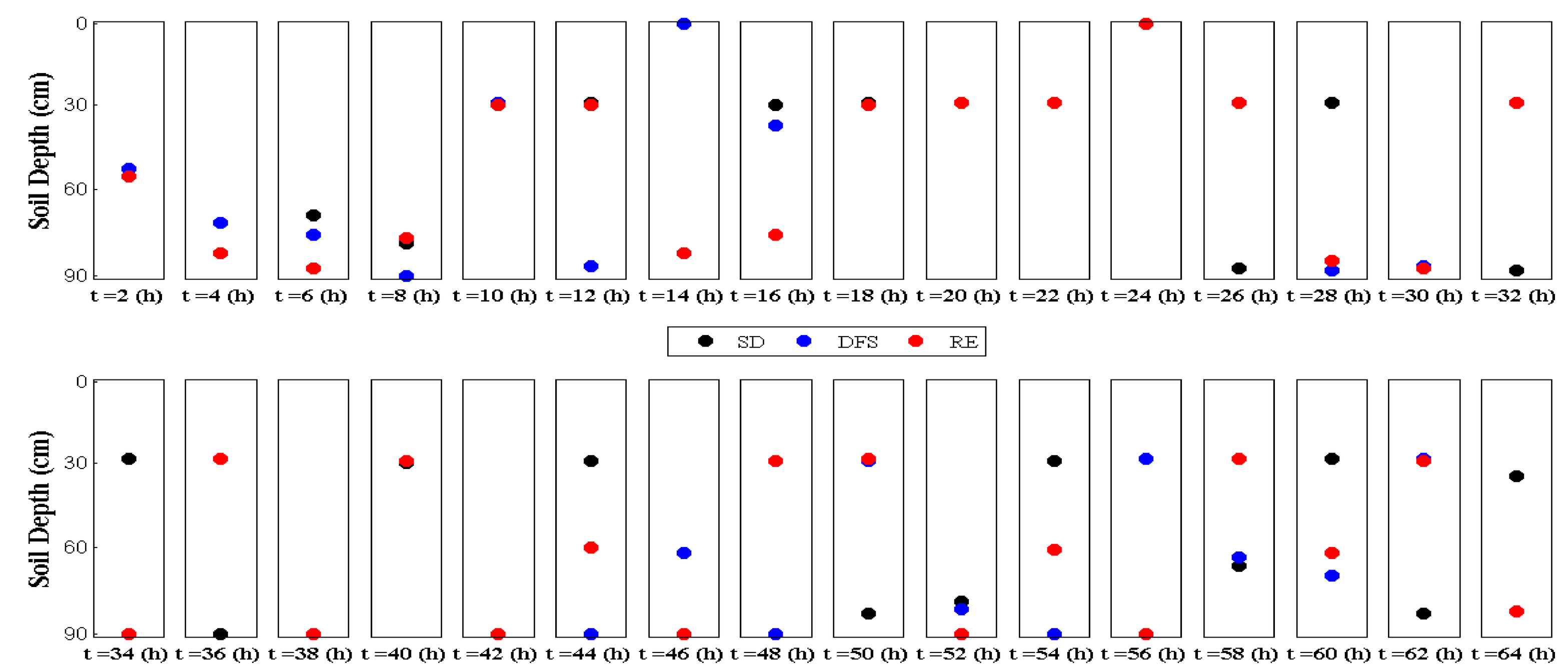
| Soil Texture | Hydraulic Parameter | Transformation Type | Limits of Variation u | Limits of Variation v | Mean | Standard Deviation |
|---|---|---|---|---|---|---|
| Sand | $K_s$ | SB | 0.00 | 70.00 | -0.394 | 1.150 |
| | $\theta_r$ | LN | 0.00 | 0.10 | -3.120 | 0.224 |
| | $\alpha$ | SB | 0.00 | 0.25 | 0.378 | 0.439 |
| | $n$ | LN | 1.50 | 4.00 | 0.978 | 0.100 |
| Loamy Sand | $K_s$ | SB | 0.00 | 51.00 | -1.270 | 1.400 |
| | $\theta_r$ | SB | 0.00 | 0.11 | 0.075 | 0.567 |
| | $\alpha$ | LN | 0.00 | 0.25 | 0.124 | 0.043 |
| | $n$ | SB | 1.35 | 5.00 | -1.110 | 0.307 |
| Sandy Loam | $K_s$ | SB | 0.00 | 30.00 | 2.490 | 1.530 |
| | $\theta_r$ | SB | 0.00 | 0.11 | 0.384 | 0.700 |
| | $\alpha$ | SB | 0.00 | 0.25 | -0.937 | 0.764 |
| | $n$ | LN | 1.35 | 3.00 | 0.634 | 0.082 |

Table 1. Statistics for parameters to be estimated

| | | $K_s$ | $\theta_r$ | $\alpha$ | $n$ |
|---|---|---|---|---|---|
| Sand | $K_s$ | 1.040 | -0.109 | 0.328 | 0.081 |
| | $\theta_r$ | | 0.182 | 0.258 | -0.047 |
| | $\alpha$ | | | 0.143 | -0.011 |
| | $n$ | | | | 0.017 |
| Loamy Sand | $K_s$ | 1.480 | -0.201 | 0.037 | 0.211 |
| | $\theta_r$ | | 0.522 | 0.017 | -0.194 |
| | $\alpha$ | | | 0.014 | 0.019 |
| | $n$ | | | | 0.108 |
| Sandy Loam | $K_s$ | 1.600 | -0.201 | 0.037 | 0.211 |
| | $\theta_r$ | | 0.538 | 0.017 | -0.194 |
| | $\alpha$ | | | 0.014 | 0.019 |
| | $n$ | | | | 0.108 |

Table 2. Correlations among transformed parameters

## Results and Discussion




C1, C2 and C3 are fixed sampling strategies, while SD, DFS and RE are designed strategies based on different information metrics.

Figure 2. The optimal sampling location at each time step based on the three information metrics: Shannon entropy difference (SD), degrees of freedom for signal (DFS) and relative entropy (RE).

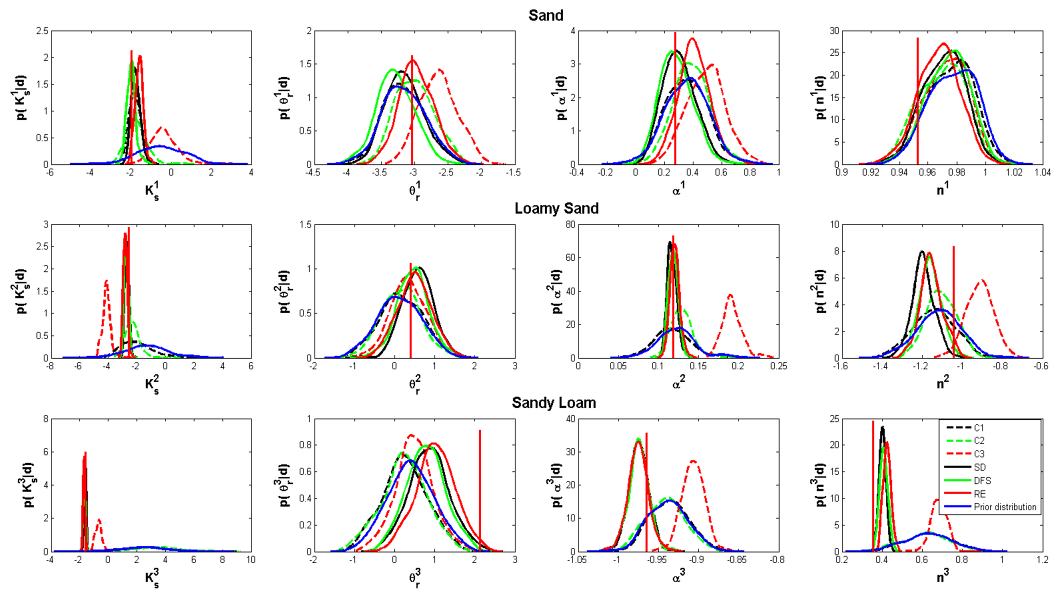



Figure 3. The posterior probability distributions of parameters with different sampling strategies. The red vertical lines stand for the true values, while the blue solid curves stand for the prior probability distributions. Results show that the designed strategies provided more accurate estimations (more reasonable confidence intervals and more accurate maximum-a-posterior estimations).

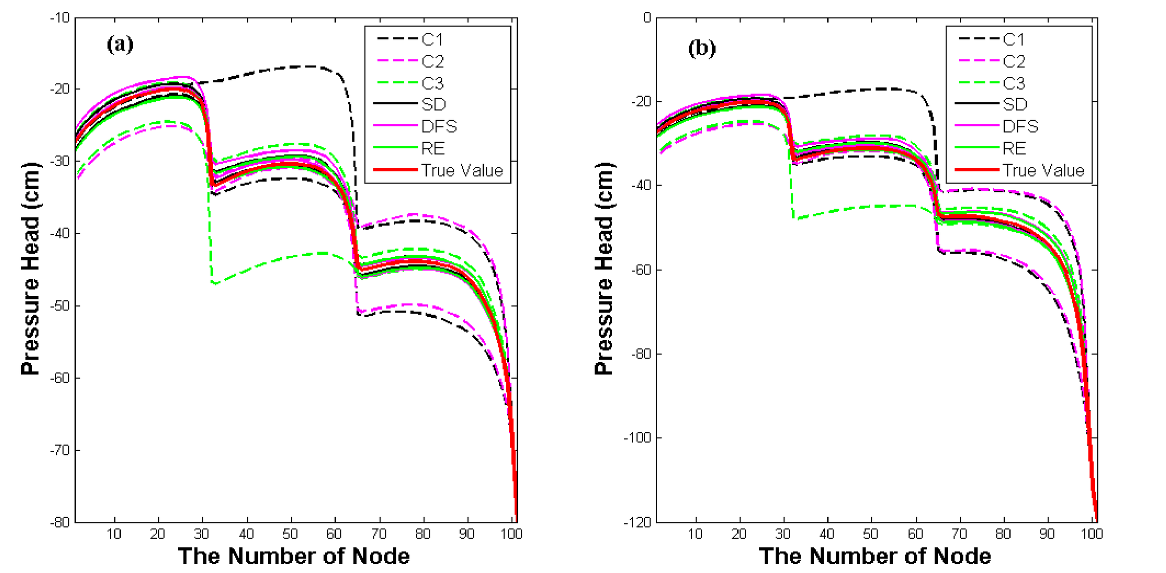



Figure 4. The predicted pressure heads on all nodes at the final time step (t = 64 h). The lines with the same color stand for the bounds of 80% confidence intervals with different sampling strategies. (a) The lower boundary condition of study domain was a constant pressure head of -80 cm, (b) The lower boundary condition of study domain was changed to -120 cm. Results show that the designed strategies predicted pressure heads more accurately.

## Conclusions

✓ Compared with conventional sampling strategies, the optimal sampling strategies provided more accurate parameter estimation and state prediction;

✓ The optimal sampling designs based on various information metrics (Shannon entropy, degree of freedom for signal and relative entropy) performed similarly.