ISEIS

Journal of Environmental Informatics 39(2) 81-96 (2022)

Journal of Environmental Informatics

www.iseis.org/jei

Improving Soil Salinity Simulation by Assimilating Electromagnetic Induction Data into HYDRUS Model Using Ensemble Kalman Filter

R. J. Yao[1], J. S. Yang[1 *], X. P. Wang[1], Y. Zhao[3], H. Q. Li[1, 2], P. Gao[4], W. P. Xie[1], and X. Zhang[1]

[1] *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *College of Resources and Environmental Engineering, Ludong University, Yantai 264025, China*
[4] *Department of Earth and Ocean Sciences, University of North Carolina, 601 South College Road, Wilmington, NC 28403, USA*

Received 23 April 2020; revised 7 July 2020; accepted 12 January 2021; published online 24 March 2021

ABSTRACT. Assimilation of proximally and remotely sensed information on soil salinization-related attributes into a hydrological model is essential to improve the forecast performance of the profiled soil salinity dynamics for developing appropriate soil amendment practices. Although the family of ensemble Kalman filters (EnKF) is widely used in data assimilation, their applicability and reliability for soil salinization estimation requires further experimental validation. Here, we evaluated the assimilation performance of apparent electrical conductivity (EC_a) data obtained from an electromagnetic induction meter (EM38) into the HYDRUS hydrological model. Results showed that the EnKF method improved the simulation accuracy of soil salinity at 0 ~ 100 cm soil depths, as indicated by the decreased root-mean-square error of 32.6 ~ 76.7% and increased Nash–Sutcliffe efficiency of 9.6 ~ 71.2%. The HYDRUS-simulated values with EnKF were closer to the measured values than the values simulated by the HYDRUS model, and this benefitted from updating the running trajectory of the HYDRUS model. The EnKF values derived from measured EC_a data were better than HYDRUS-simulated values with EnKF. Soil salinity simulation was sensitive to ensemble size, error level, and EC_a data depth. Considering the ensemble representativeness and computational efficiency, the optimal ensemble size was judged to be 50. The maximum acceptable observation error was 10%, and observation data to a depth of 100 cm was suggested in EnKF assimilation to minimize the root-mean-square error. It was concluded that proximally sensed EM38 data coupled with the EnKF algorithm is promising for improving the simulation performance and providing a prospective method for simulating large-scale ecological and hydrological processes by coupling multi-source data and hydrological models.

Keywords: soil salinity, proximally sensed data, ensemble Kalman filter, data assimilation, HYDRUS model

1. Introduction

Understanding the spatio-temporal dynamics of soil salinity in the root zone is essential for hydrological, agricultural, and ecological research, as soil salinization is a constant threat to soil-sustainable development. This is particularly the case in the coastal area of Eastern China, where soil salinization is a key limitation to crop production. Additionally, increasing population and cropland reduction have drawn growing public concerns to the amendment and efficient utilization of salt-affected soils as well as cropland expansion (Li et al., 2014). In agriculture, obtaining real-time and accurate information on the root-zone soil salinity is crucial for determining optimal irrigation and drainage practices to minimize salinization hazards and maximize crop yields (Metternicht, 2017). It is widely reported that soil salinity influences most important processes of hydrological cycles, such as root water-nutrient uptake and growth and energy exchange between the land surface and atmosphere (e.g., Zhou et al., 2017; Wu et al., 2018). Therefore, the development of measurement and estimation approaches for real-time and accurate characterization of root-zone soil salinity is urgently needed.

During the past two decades, proximal soil sensing, such as electromagnetic induction meter (EM) use, time-domain reflectometry (TDR), frequency-domain reflectometry (FDR), and ground-penetrating radar (GPR) have become popular non-invasive techniques and have been extensively used in many fields, e.g., soil characterization, hydrological research, and precision agriculture (Aldabaa et al., 2015; Wallor et al., 2018). Among them, electromagnetic induction (EMI) instruments, including EM31, EM38, EM38-DD, and EM38-MK2 meters (Geonics Limited, Mississauga, Canada), DUALEM-1 and DUALEM-2 meters (DUALEM Inc., Milton, Ontario, Canada), and Veris 3100 (Veris Technologies, Salina, Kansas, USA) (Doolittle and Brevik, 2014), have been widely used to assess the nature, origin, and evolution of soil salinization at multiple temporal and spatial scales. The EM38 meter, by far the most widely used EMI sensor in soil science, has been commonly employed (Ding and Yu, 2014; Huang et al., 2016; Narjary et al., 2019). Further-

* Corresponding author. Tel.: +86 25 86881222; fax: +86 25 86881222.
E-mail address: jsyang@issas.ac.cn (J. S. Yang).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.202100451

more, active research has been undertaken to improve profile soil salinity estimation using various combinations of operation modes and different measurement heights above the ground, including layered-earth modeling, multiple linear regression, linear mixed-effects modeling, non-linear Tikhonov regularization, and the joint inversion method (Triantafilis and Monteiro Santos, 2010; Huang et al., 2015). However, the limitation of EM38 application is that it does not work well in areas where soil salinity is mostly homogeneous and other soil attributes, such as soil moisture, texture, bulk density, and clay mineralogy, are heterogeneous (Corwin and Lesch, 2014).

With the development of data assimilation approaches, remote and proximal soil sensing data have been continuously adapted to improve soil and crop parameter prediction (Tran et al., 2013; Huang et al., 2016). Data assimilation methods were initially developed to improve weather forecasts and ocean dynamics prediction in meteorology and oceanography and have been used in hydrology and soil science for over two decades (Houser et al., 1998; Bauser et al., 2018). Assimilation methods are ascribed to two categories: variational data assimilation and sequential data assimilation. Variational data assimilation aims to find the most likely state and mode of the analysis probability density function by minimizing the non-linear cost function. In contrast, sequential data assimilation is based on an approximation of the mean of the analysis distribution (Rawlins et al., 2007). Currently, Kalman filter methods are the most widely used assimilation approaches, including the ensemble Kalman filter (EnKF) (Wang et al., 2018), extended Kalman filter (Sun et al., 2016), adaptive ensemble Kalman filter (Reichle et al., 2008), and maximum likelihood ensemble filter (Tran et al., 2013). In soil science, Kalman filters have been extensively used to assimilate remotely and proximally sensed data into models to improve the estimation of the soil moisture profile and crop yield (Wu et al., 2012). For instance, Tran et al. (2013) assimilated proximally sensed ground-penetrating radar (GPR) data into a hydrodynamic model to improve soil moisture profile reconstruction. Brandhorst et al. (2017) integrated time-domain reflectometry (TDR) data with unsaturated zone models to improve soil moisture prediction and handle soil hydraulic conductivity uncertainty. In addition to soil moisture, de Wit and van Diepen (2007) assimilated satellite-derived variables (such as leaf area index and soil water index) into the crop model WOFOST to improve regional crop yield forecasts.

Despite the success of EMI in soil salinity delineation and extensive applications of the ensemble Kalman filter method in hydrological processes, few studies have coupled EMI measurements and ensemble Kalman filters to improve soil salinity estimation (Huang et al., 2017). In the current study, linear mixed-effects models relating soil salinity profiles with the bulk electrical conductivity (EC_a) data (Yao et al. 2016) were used as observation operators. First, we simulated the soil-salt dynamics in the root zone using the HYDRUS model. Then, we assimilated the EC_a data into the HYDRUS model using the ensemble Kalman filter and evaluated the influence of ensemble Kalman filter on soil salinity simulation performance. The objectives of this study were to: (*i*) investigate the effect of EnKF assimilation on the simulation accuracy of soil salinity at the 0 ~ 100 cm depth, and (*ii*) examine the sensitivity of the simulation performance to the assimilation parameters, e.g., ensemble number, error level of observation data, and the depth of the EC_a data used in the assimilation.

2. Materials and Methods

2.1. Experimental Site Description

The experimental site of the Huanghai Raw Seed Growing Farm is situated in the marine-terrestrial interlaced area in the Dongtai Prefecture (32°38′ ~ 32°39′ N, 120°52′ ~ 120°54′ E), North Jiangsu Province, China. The land on this farm was enclosed and reclaimed from coastal mudflats in 1999 and 2004, respectively, and divided by dikes in the north-south direction which were built at different ages (Figure 1(a)). The experimental site is 7 km away from the China Yellow Sea coastline and has a nearly flat topography with an elevation of 1.1 ~ 1.5 m above sea level (Figure 1(b)). The climate is characterized by a subtropical zone with a cold, dry season from November to March and a hot, wet season from June to September. The mean annual rainfall is 1,048.5 mm, with over 70% occurring from May to September. Soils are developed on Yangtze alluvial sediments and marine sediments, and the predominant soil type is classified as a loamy, mixed Typic Halaquepts group of Aquepts in Inceptisols based on soil taxonomy (Soil Survey Staff, 2010). A shallow saline water table (average groundwater EC_g of 8.01 dS/m and water table at 1.44 m) results in large areas of salt-affected land and poor crop growth and soil productivity.

A section of the field between Dike 1999 and Dike 2004, approximately 60 m in width and 160 m in length, was chosen for the experiment (Figure 1(b)). This selected field had no documented history of cultivation until April 2007, and afterward, rice (*Oryza sativa* L.)/barley (*Hordeum vulgare* L.) rotation was initially implanted. Due to freshwater scarcity, a rainfed corn (*Zea mays* L.)/barley (*Hordeum vulgare* L.) rotation was adopted in 2012. Because of the low-lying terrain and freshwater shortage for salt leaching, soil salinization has been recognized as the most significant constraint to agricultural production in this field.

2.2. Field Observations and Data Collection

The experiment was conducted from November 2015 to October 2016. Meanwhile, the observation data, including field EC_a data, weather, soil, and groundwater data, were periodically collected for EnKF assimilation (Table 1). The data of three EMI measurement sites and three soil sampling locations near the central observation well were selected for EnKF assimilation (Figure 1(b)), considering that a larger EnKF data area resulted in higher spatial heterogeneity of soil and groundwater properties. Weather data were continuously recorded from the automatic weather station installed at the experimental site (Figures 2(a) and 2(b)). Daily reference evapotranspiration (ET_0) was calculated based on daily meteorological data (Raes, 2009). Daily crop evapotranspiration (ET_c) was determined from ET_0 using the single-crop coefficient approach (Allen et al.,

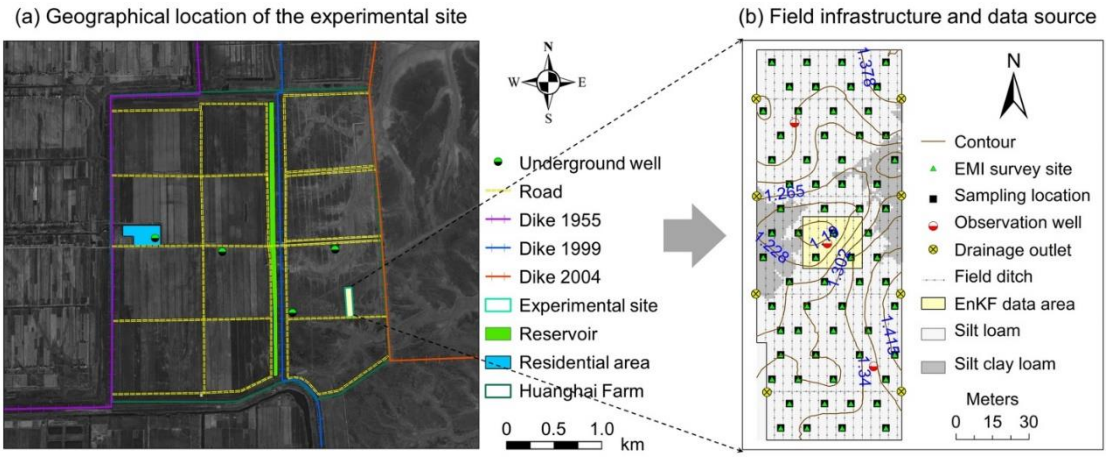

Figure 1. Geographical location of the experimental site and spatial distribution of field infrastructure, contour of ground elevation, sampling sites, EMI survey sites as well as the data area selected for EnKF assimilation. (a) Geographical location of the experimental site; (b) field infrastructure and data source. Dike 1955, Dike 1999, and Dike 2004 mean the reclamation dike built in 1955, 1999, and 2004, respectively.

Table 1. Overview of the Observation Data and Their Sources in This Study

Data type	Main indices	Observation period/time	Source/method
Meteorological data	Precipitation, global radiation, relative humidity, wind speed, barometric pressure, air temperature, evaporation, sunshine hours	1st Nov. 2015 ~ 31st Oct. 2016	Automatic Weather Station
Groundwater properties	Water table, groundwater electrical conductivity	1st Nov. 2015 ~ 31st Oct. 2016	CTD-Diver (type DI263)
Soil moisture and salinity*	Soil EC_e, total dissolved salts TDS_e and water content on the profile (0 ~ 20, 20 ~ 40, 40 ~ 60, 60 ~ 80, and 80 ~ 100 cm)	1st Nov. 2015 ~ 31st Oct. 2016	Soil sampling and lab analysis
Soil physical properties	Bulk density, texture, field capacity, saturated hydraulic conductivity, saturated soil moisture on the profile (0 ~ 20, 20 ~ 40, 40 ~ 60, 60 ~ 80, and 80 ~ 100 cm)	Aug. 2015	Core method, the Bouyoucos Hydrometer method, the Wilcox method, constant head method
Proximally sensed EMI data*	Apparent electrical conductivity (EC_a) obtained at 0, 20, 40, 60, and 80 cm above ground in the horizontal mode	1st Nov. 2015 ~ 31st Oct. 2016	Electromagnetic induction meter (type EM38)

* Soil moisture and salinity data on 11 dates from 1st Nov. 2015 to 31st Oct. 2016 were determined using field soil sampling and lab analysis, and proximally sensed EMI data was simultaneously collected at each soil sampling site. The 11 soil sampling and EMI survey dates were Nov. 3, 2015, Dec. 9, 2015, Jan. 9, 2016, Feb. 9, 2016, Mar. 24, 2016, Apr. 27, 2016, Jun. 3, 2016, Jul. 6, 2016, Aug. 11, 2016, Sep. 15, 2016, and Oct. 27, 2016.

1998). The water table and groundwater salinity were automatically recorded using CTD-Diver sensors (type DI263) installed in the observation well. The groundwater data were collected at hourly intervals, and the daily groundwater data were averaged (Figure 2(c)).

2.3. Soil Sampling and Lab Analysis

Core soil samples were collected at 0 ~ 20, 20 ~ 40, 40 ~ 60, 60 ~ 80, and 80 ~ 100 cm layers for lab analysis to determine the bulk density, texture, field capacity, saturated hydraulic conductivity, and saturated soil moisture. Disturbed soil samples at the same depths were also obtained by hand auger for soil salinity and water content measurement. The measured basic soil properties are given in Table 2. Soil salinity and moisture data of disturbed soil samples, collected in late October 2015, were used as initial conditions of the HYDRUS model. Moreover, soil profile samples were repeatedly collected on 11 dates during the experimental period (see Table 1) for model calibration. The van Genuchten-Mualem (V-G) model parameters of the soil water retention curve (SWRC) were estimated from the bulk density and sand, silt, and clay percentage values using the Rosetta pedo-transfer functions (Schaap et al., 2001). These parameters were used as the initial values of the HYDRUS model.

2.4. Proximal Soil Sensing

An electromagnetic induction (EM38) meter was used to collect the EMI measurements (EC_a data). EMI measurements at the same locations were repeatedly made on 11 dates from November 2015 to October 2016 (see Table 1). At each location, the EM38 meter in the horizontal operation dipole was positioned on the soil surface, and at heights of 20, 40, 60, and 80

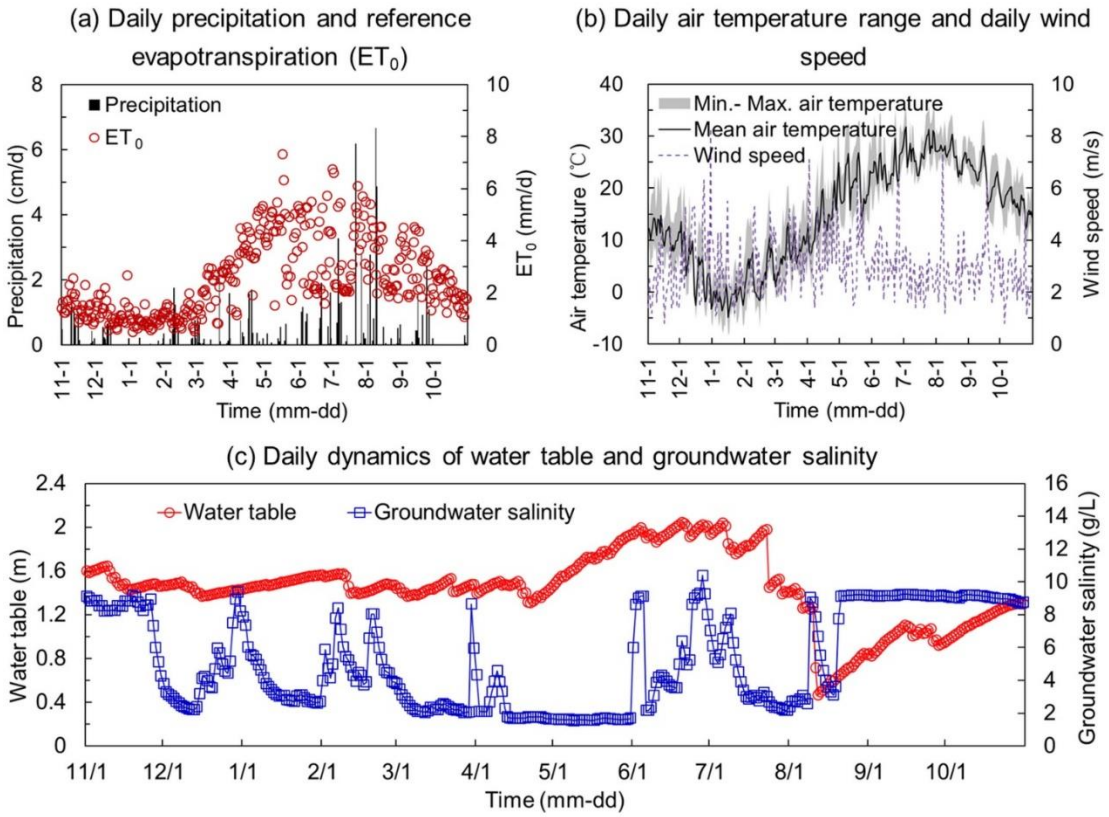

Figure 2. Daily weather and groundwater data from Nov. 2015 to Oct. 2016. (a) Daily precipitation and reference evapotranspiration (ET_0); (b) daily air temperature range and daily wind speed; (c) daily dynamics of water table and groundwater salinity.

Table 2. Soil Property Data, the Calibrated van Genuchten–Mualem Model Parameters, and Solute Transport Parameters at the Experimental Site

Soil parameters		Soil layers (cm)				
		0 ~ 20	20 ~ 40	40 ~ 60	60 ~ 80	80 ~ 100
Texture class		Silt loam	Silt loam	Silt loam	Silt loam	Silt loam
Sand (%)		17.8	15.5	15.1	14.7	14.2
Silt (%)		70.6	71.8	74.0	73.8	72.5
Clay (%)		11.6	12.7	10.9	11.5	13.3
Soil salt content TDS_e (g/kg)*		2.49	2.26	3.25	2.84	3.45
Soil moisture (cm^3/cm^3)*		0.253	0.303	0.288	0.278	0.320
Bulk density (g/cm^3)		1.39	1.50	1.44	1.45	1.43
Field capacity (cm^3/cm^3)		0.28	0.24	0.26	0.27	0.26
Saturated hydraulic conductivity (cm/d)		18.55	9.25	18.16	9.73	9.53
The van Genuchten–Mualem model parameters	θ_r (cm^3/cm^3)	0.068	0.067	0.054	0.073	0.072
	θ_s (cm^3/cm^3)	0.415	0.404	0.380	0.433	0.437
	α (1/cm)	0.0053	0.0062	0.0059	0.0063	0.0063
	n	1.657	1.601	1.632	1.604	1.602
Solute transport parameters	Longitudinal dispersivity D_L (cm)	2.46	1.12	3.99	0.74	0.68
	Adsorption isotherm coefficient K_d (cm^3/g)	0.35	0.39	0.29	0.41	0.38

* Soil moisture and salinity data measured on 26th Oct. 2015 were used as initial conditions of the HYDRUS-1D model.

cm above the ground, and the EMI data at the different heights were obtained. Each EMI survey was completed within two consecutive days to ensure homogeneous soil conditions, and minimize the influence of soil properties other than soil salinity on EMI measurements. This work was simultaneously performed with soil sampling. During each EMI survey, an electronic thermometer was used to measure soil temperature in the 0 ~ 40 cm soil layer at hourly intervals for temperature cali-

bration of EMI measurements (Ma et al., 2011).

2.5. EnKF Assimilation of EMI Measurements into the HYDRUS Model

In the present study, the assimilation system used the ensemble Kalman filter (EnKF) to integrate proximally sensed EMI measurements (EC_a data) and observation operators into a water flow and solute transport model (HYDRUS).

2.5.1. Ensemble Kalman Filter (EnKF)

In a data assimilation system, model-generated data are corrected toward observational estimates, and the extent of correction is determined by the error levels between the simulations and observations. The standard Kalman filter (KF) is an optimal sequential data assimilation method for linear dynamics and measurement processes based on a Gaussian error distribution. For non-linear dynamics, Evensen (1994) and Burgers et al. (1998) developed the ensemble Kalman filter (EnKF) to overcome the shortcomings of the KF, including instability of the state error covariance matrix and error information propagation for large-scale environmental progress. The idea behind the EnKF is that an appropriate ensemble of model trajectories captures the relevant parts of the error structure. The EnKF, representative of a Monte Carlo variant of the Kalman filter, uses an ensemble of the non-linear model's trajectories to capture the error covaryances and propagate the ensemble states. It has found wide applications in land data assimilation for modestly non-linear problems in recent years (Thiboult and Anctil, 2015; Wang et al., 2018). For the basic principles and detailed descriptions of the EnKF algorithm, refer to Houtekamer and Mitchell (1998) and Reichle et al. (2008).

2.5.2. HYDRUS-1D Model

The HYDRUS-1D model is a one-dimensional finite element model that incorporates physically-based modules to simulate the movement of soil water, heat, vapor, multiple solute transport, and major ion movement in variably saturated porous media (Šimůnek et al., 2008; Šimůnek et al., 2016). The HYDRUS-1D model was widely adopted to simulate vertical soil water flow and salt transport in unsaturated zones (Šimůnek et al., 2012). In the HYDRUS model, water flow and solute transport modules are included, and soil water flow and salt transport in the 0 ~ 1.0 m layer were simulated at 0 ~ 20, 20 ~ 40, 40 ~ 60, 60 ~ 80, and 80 ~ 100 cm layers. Root water uptake and root growth were negligible due to excessively high soil salinity and poor crop growth. For water flow simulation, the upper boundary condition was "atmospheric boundary condition with a surface runoff", and the lower boundary condition was "variable pressure head" due to the shallow groundwater. No irrigation occurred during the experimental period, as the experimental site was rainfed. For salt transport calculation, the salt concentration of precipitation was used as the upper boundary condition, and the measured time-variable groundwater salinity (EC_g, dS/m) was used as the lower solute boundary condition.

2.5.3. Observation Operators

The observation operators, relating the soil salinity (TDS_e, g/kg) profile with the EMI data (EC_a, mS/m) obtained at different heights above the ground, were established using methods of the linear mixed-effects model (LME) and restricted maximum likelihood (REML) (Yao et al., 2015). For the observation operator in each soil layer, the EMI data collected at different heights were used as independent variables to estimate soil salinity. REML was adopted to determine the optimal independent variables by removing some less important independent variables, and LME was used to relate the selected independent variables with soil salinity. This procedure was conducted for each soil layer, and the established mixed-effects models were employed as observation operators in the EnKF assimilation system. Using observation operators, the measured EMI data were assimilated into the HYDRUS-1D model to update the state of the soil salinity profile on different dates. More details of the observation operators are described in Yao et al. (2016).

2.5.4. Assimilation Scheme of HYDRUS-1D Model and Proximally Sensed Data Using EnKF

The flowchart in Figure 3 shows the assimilation process of the HYDRUS-1D model and the proximally sensed EC_a data (EMI measurements) using the EnKF algorithm. The assimilation procedure is briefly outlined below:

1. Use *a priori* knowledge on the soil conditions to specify the initial soil salinity state and initial state ensemble. These values will work as the analysis state $\overline{X}_0^a$ and its ensemble $\overline{X}_{i,0}^a$ at time $t = 0$.
2. Based on the $\overline{X}_t^a$ and $X_{i,t}^a$ values at time t, run the HYDRUS-1D model to simulate the temporal dynamics of soil salinity from time t to $t + 1$ and obtain the forecast state $\overline{X}_{t+1}^f$ and its ensemble $X_{i,t+1}^f$ at time $t + 1$.
3. Examine whether the EMI measurements are available. If unavailable, assign $\overline{X}_{t+1}^f$ and $X_{i,t+1}^f$ to $\overline{X}_{t+1}^a$ and $X_{i,t+1}^a$, namely $\overline{X}_{t+1}^f = \overline{X}_{t+1}^a$ and $X_{i,t+1}^f = X_{i,t+1}^a$ at time $t + 1$. If available, employ the EMI measurements, observation operators, and EnKF algorithm to estimate $\overline{X}_{t+1}^a$ and $X_{i,t+1}^a$, and then replace the soil salinity state with those estimates. The Kalman gain K_{t+1} is used here to update the state of the soil salinity profile.
4. With $\overline{X}_{t+1}^a$ and $X_{i,t+1}^a$obtained from step 3 as initial conditions run the HYDRUS-1D model for the forecast or EnKF update at the next time.

During the assimilation procedure, four different soil salinity values were obtained when soil sampling and the EMI survey were conducted, including 1) 'measured value', 2) 'HYDRUS-simulated value', 3) 'EnKF value', and 4) 'HYDRUS-simulated value with EnKF'. Here, we considered soil samples' salinity determined from lab analysis as the 'measured value'. The soil salinity simulated by the HYDRUS model was determined as the 'HYDRUS-simulated value,' which was derived from HYDRUS simulations without state variable updates. The 'EnKF value' was defined as the simulated soil salinity by assimilating mea-

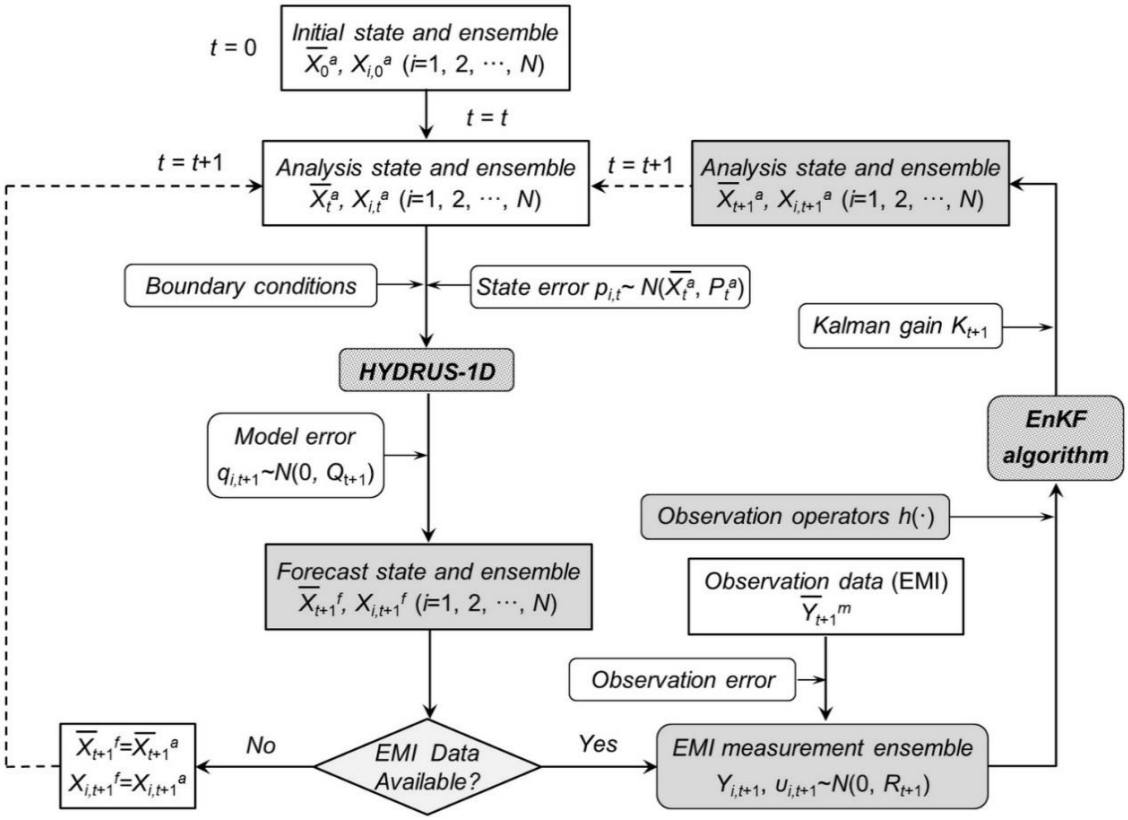

Figure 3. Flowchart of the EnKF algorithm assimilating proximally-sensed EC_a data into the HYDRUS model. The salt content in the soil solution, calculated from soil salt content (TDS_e, g/kg), soil moisture (θ, cm^3/cm^3), and bulk density (ρ_b, g/cm^3), using the equation: $TDS_e \times \rho_b / \theta$, was used as soil salinity in the assimilation procedure. $\overline{X}_0^a$ is initial value of soil salinity, g/L; $X_{i,0}^a$ is initial state ensemble of soil salinity in the i^{th} layer, g/L; $\overline{X}_t^a$ and $\overline{X}_{t+1}^a$ are analysis values of soil salinity at time t and $t + 1$, respectively, g/L; $X_{i,t}^a$ and $X_{i,t+1}^a$ are analysis state ensembles of soil salinity in the i^{th} layer at time t and $t + 1$, respectively, g/L; $p_{i,t}$ is the state error of soil salinity at time t, which is normally distributed with mean of $\overline{X}_t^a$ and variance of P_t^a ; $q_{i,t+1}$ is the HYDRUS model error at time $t + 1$, which is normally distributed with mean of 0 and variance of Q_{t+1}; $\overline{X}_{t+1}^f$ is the forecast soil salinity at time $t + 1$, g/L; $X_{i,t+1}^f$ is the forecast state ensemble of soil salinity at time $t + 1$, g/L; Y_{t+1}^m is the observation of EC_a data at time $t + 1$, mS/m; $Y_{i,t+1}$ is the observation state ensemble of apparent electrical conductivity, mS/m; $\upsilon_{i,t+1}$ is the observation error disturbance, following a normal distribution with mean of 0 and standard deviation of R_{t+1}; $h(\cdot)$ is the observation operator relating apparent electrical conductivity to soil salinity profiles; K_{t+1} is the Kalman filter gain.

sured EMI measurements and observation operators into the 'HYDRUS-simulated value', and the 'EnKF value' was available only on the date when EMI measurements were made. The 'HYDRUS-simulated value with EnKF' was considered as the simulated soil salinity by the HYDRUS model, which used the updated state variable, i.e., 'EnKF value' on the previous date as input data. Thus, the 'HYDRUS-simulated value with EnKF' incorporated the state error and model error, which propagated forward in time. The same calibrated parameter system and boundary conditions were used in the HYDRUS model for the 'HYDRUS-simulated value', 'EnKF value', and 'HYDRUS-simulated value with EnKF'.

It must be noted that the assimilation procedure shown in Figure 3 is based on the following prerequisites: 1) the soil water flow process is fitted well using the HYDRUS-1D model with the calibrated soil hydraulic parameters, 2) the specified parameter system is static during the assimilation procedure, and 3) soil water flow is independent of the simulation and EnKF updates of the temporal soil salinity state.

2.6. Statistical Analysis

The performance of EnKF assimilation during the simulation period was statistically evaluated using the following four quality criteria (Feng et al., 2017): (1) the mean relative error, MRE; (2) the root-mean-square error, RMSE; (3) the determination coefficient, r^2; and (4) the Nash–Sutcliffe efficiency coefficient, NSE.

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{(P_i - M_i)}{M_i} \times 100\% \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - M_i)^2} \quad (2)$$

$$r^2 = 1 - \frac{\sum(P_i - M_i)^2}{\sum(P_i - \overline{P})^2 \sum(M_i - \overline{M})^2} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(M_i - P_i)^2}{\sum_{i=1}^{N}(M_i - \overline{M})^2} \quad (4)$$

where N is the number of observed values of soil salinity in the profile; M_i is the individual measured soil salinity value at the soil depth i, g/L; P_i is the individual simulated soil salinity at the soil depth i, g/L; $\overline{P}$ is the average of simulated soil salinity in the profile, g/L; M is the average of measured soil salinity in the profile, g/L. For a perfect prediction, the values of MRE and RMSE should be close to 0 and r^2, NSE should approximate to 1, and NSE values ranging between 0.5 and 1.0 are considered preferable.

3. Results

3.1. Field Observations and HYDRUS Model Calibration

Soil water content fluctuated with the rainfall, evaporation, and redistribution processes during the simulation period. Large rainfall events caused an increase in soil water content (saturated water content) at the soil surface for several days, such as from days 284 ~ 287, with 155.3 mm of rainfall recorded during this period (Figure 4(a)). Contrary to soil moisture dynamics, soil salinity content in the upper soil layers decreased sharply with rainfall as the soil solution was diluted and salts were leached to deeper layers, and then gradually increased since most of the leached salts could not be discharged in time and remained in the deeper soil layers (Figure 4(b)). A possible explanation is that the topsoil had a relatively higher water holding capacity, whereas the 20 ~ 40 cm soil layer was more compact and had lower permeability (Table 2). Meanwhile, the 40 ~ 60 cm soil layer had significantly higher saturated hydraulic conductivity (K_{sat}) and permeability than the 60 ~ 80 and 80 ~ 100 cm layers, resulting in salt accumulation primarily in the 20 ~ 60 cm soil layer. After that, soluble salts transported upward via capillary rise due to high evapotranspiration rates and a shallow groundwater table, leading to salt accumulation in the upper layers until the next effective rainfall event.

The HYDRUS model sufficiently captured the trends in soil moisture and salinity concentrations in different soil layers from November 2015 to October 2016. The hydraulic parameters of HYDRUS model were calibrated using an inverse algorithm. The V-G model parameters α, n, and θ_r were the main parameters to be calibrated, whereas K_s and θ_s were only slightly adjusted, as K_s and θ_s were determined from field soil sampling and lab analysis. Based on the goodness-of-fit criteria, the solute transport parameters, includeing the longitudinal dispersivity (D_L) and adsorption isotherm coefficient (K_d) for each soil layer, were also optimized using the inverse algorithm. Determination of the dimensionless fraction, immobile water content (θ_{im}), and molecular diffusion coefficient (D_w) at the experimental site were based on Wang et al. (2014). The calibrated soil hydraulic and solute transport parameters are presented in Table 2. For soil moisture simulation, the MRE values were all within ±5%, the RMSE values ranged between 0.018 and 0.024 cm^3/cm^3, and the r^2 values varied from 0.73 to 0.94, whereas the NSE values were all > 0.7 (Table 3). Concerning soil salinity simulation, some discrepancies were observed in deeper soil layers of the root zone, especially in the 40 ~ 60 and 60 ~ 80 cm layers, in which 1 ~ 1.5 cm thin clay layers appeared due to different alluvial sediments produced during the soil-formation process. Moreover, the groundwater table can sometimes rise into these layers after considerable precipitation, i.e., the water table rose to 0.47 m after continuous rainfall on Aug. 12, 2016 (day 286), which was also shown by the simulated soil moisture at the corresponding time. Generally, the goodness-of-fit criteria for salinity simulation were satisfactory with −0.24% < MRE < 6.97%, 1.124 < RMSE < 2.088 g/L, r^2 > 0.56, and NSE > 0.52 (Table 3).

Table 3. Goodness-of-Fit Test Criteria for Soil Water and Salinity Simulation during Model Calibration

	Soil layers (cm)	MRE (%)	RMSE (cm^3/cm^3 or g/L)	r^2	NSE
Soil moisture	0 ~ 20	4.94	0.021	0.94	0.91
	20 ~ 40	3.78	0.024	0.89	0.87
	40 ~ 60	−3.73	0.022	0.87	0.84
	60 ~ 80	−2.59	0.018	0.85	0.82
	80 ~ 100	−2.32	0.019	0.73	0.71
Soil salinity	0 ~ 20	6.97	2.015	0.56	0.52
	20 ~ 40	1.07	2.088	0.80	0.69
	40 ~ 60	0.13	1.124	0.60	0.57
	60 ~ 80	−0.24	1.702	0.75	0.59
	80 ~ 100	4.67	1.192	0.76	0.73

3.2. EnKF Assimilation Results

EnKF assimilation performance was evaluated by comparing the four different soil salinity values (Figure 5). As indicated by the error analysis (Table 4), the EnKF values were closer to the measured values than either the HYDRUS-simulated values or HYDRUS-simulated values with EnKF. In contrast, the HYDRUS-simulated values with EnKF had better performance than the HYDRUS-simulated values. For HYDRUS-simulated values with EnKF, the RMSE ranged between 0.708 and 1.394 g/L, and NSE varied from 0.67 to 0.86, and the corresponding criteria for EnKF values ranged from 0.414 to 0.930 g/L and from 0.84 to 0.94, respectively. This result showed that the simulation performance for EnKF values was better than HYDRUS-simulated values with EnKF. It was not unexpected considering that EnKF values were derived from EC_a data and observation operators, whereas HYDRUS-simulated values with EnKF were developed from HYDRUS simulation using the EnKF values as updated state variables. The positive effects

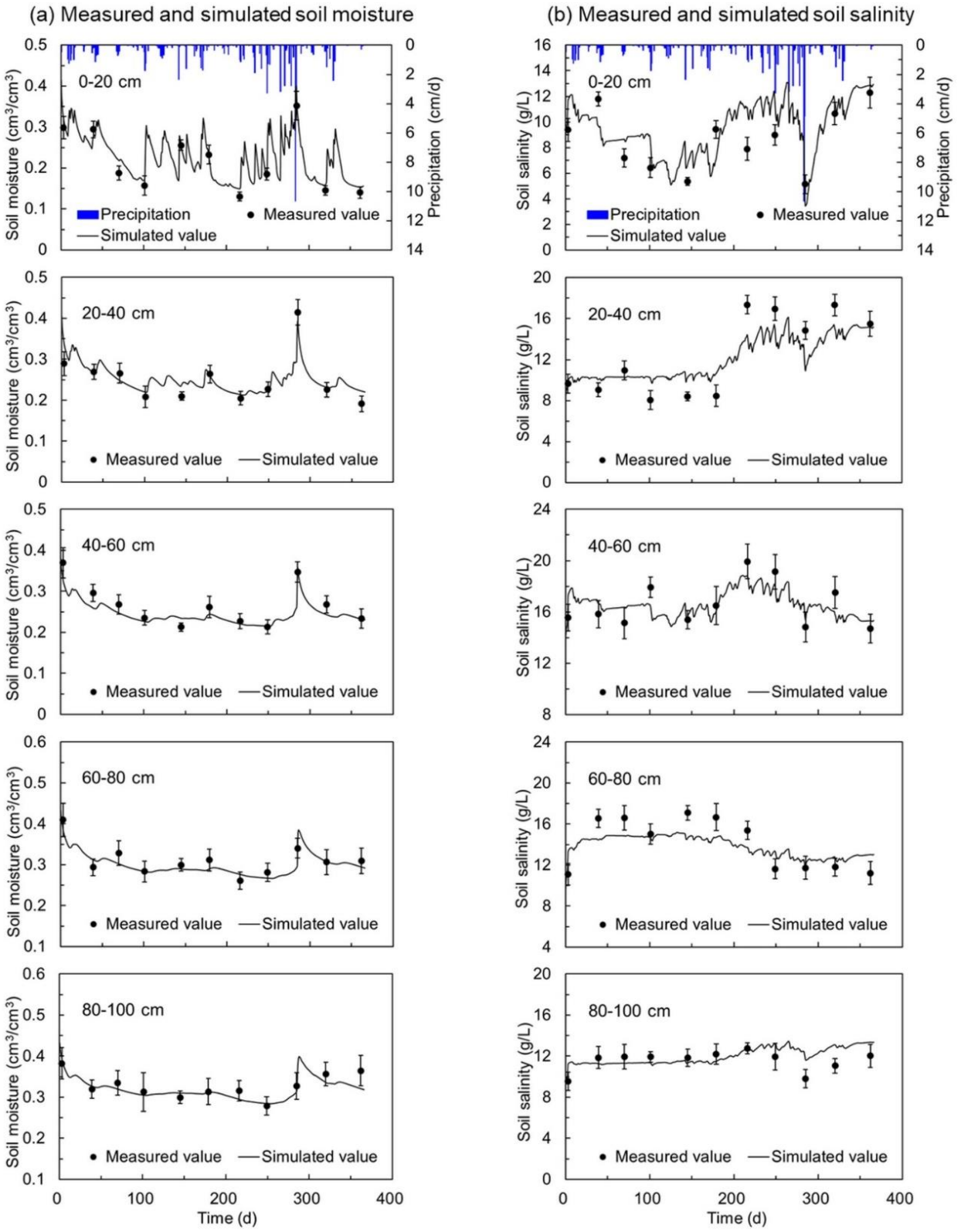

Figure 4. The HYDRUS-simulated versus measured soil moisture and soil salinity in different layers. (a) Measured and HYDRUS-simulated soil moisture in the 0 ~ 100 cm layers; (b) measured and HYDRUS-simulated soil salinity in the 0 ~ 100 cm layers. Vertical bars correspond to the standard deviation of observations.

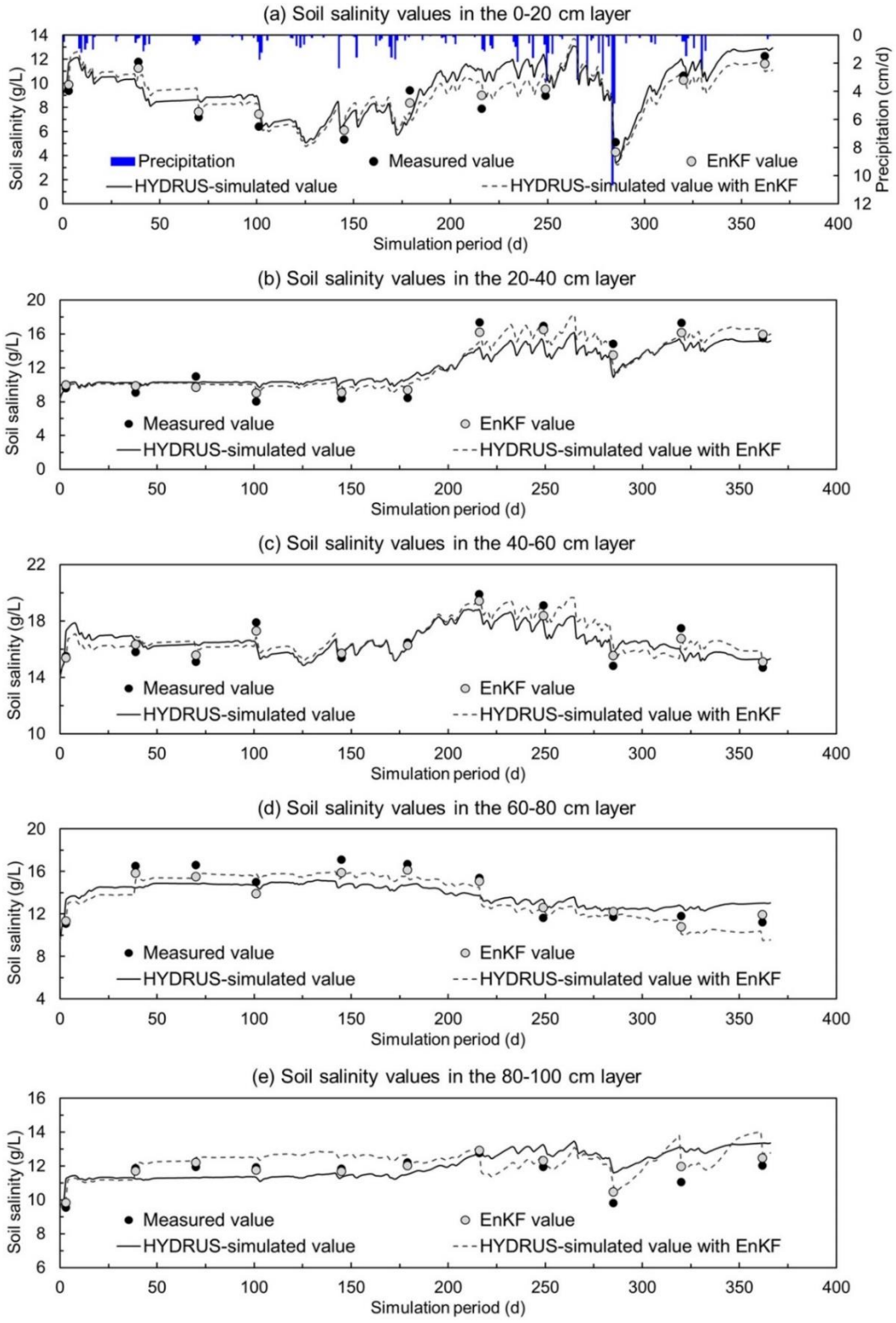

Figure 5. Soil salinity of HYDRUS-simulated values, HYDRUS-simulated values with EnKF, EnKF values, and measured values in different layers. (a) Soil salinity values in the 0 ~ 20 cm layer; (b) soil salinity values in the 20 ~ 40 cm layer; (c) soil salinity values in the 40 ~ 60 cm layer; (d) soil salinity values in the 60 ~ 80 cm layer; (e) soil salinity values in the 80 ~ 100 cm layer. Where the solid line represents the dynamics of the HYDRUS-simulated value, the dashed line represents the dynamics of the HYDRUS-simulated values with EnKF, solid dots represent the measured values, and hollow dots represent the EnKF values.

Table 4. Error Analyses on Soil Salinity of HYDRUS-Simulated Values, HYDRUS-Simulated Values with EnKF and EnKF Values

Soil layers (cm)	HYDRUS-simulated values				HYDRUS-simulated values with EnKF				EnKF values			
	MRE (%)	RMSE (g/L)	r^2	NSE	MRE (%)	RMSE (g/L)	r^2	NSE	MRE (%)	RMSE (g/L)	r^2	NSE
0 ~ 20	6.97	2.015	0.56	0.52	2.57	1.335	0.69	0.67	2.13	0.757	0.90	0.89
20 ~ 40	1.07	2.088	0.80	0.69	1.54	1.394	0.89	0.86	0.96	0.930	0.96	0.94
40 ~ 60	0.13	1.124	0.60	0.57	0.94	0.808	0.80	0.78	0.03	0.532	0.95	0.90
60 ~ 80	−0.24	1.702	0.75	0.59	−4.30	1.127	0.85	0.78	−1.66	0.831	0.93	0.88
80 ~ 100	4.67	1.192	0.76	0.73	5.87	0.708	0.94	0.80	2.11	0.414	0.87	0.84

of EnKF assimilation on simulation accuracy were observed in the whole soil profile, and this was confirmed by the decrease in RMSE and increase of r^2 and NSE in all soil layers.

3.3. Effect of Ensemble Size on Simulation Accuracy

The optimal ensemble size of the EnKF algorithm was experimentally determined by comparing the root-mean-square error (RMSE) between EnKF values and measured values. For all soil layers, a sharp decreasing trend was observed for the RMSE values of soil salinity when the ensemble size was less than 30. The improvement was slight when the ensemble size varied between 30 and 80, and no apparent improvement was observed when the ensemble size was set to more than 80 (Figure 6(a)). For instance, the RMSE of soil salinity in the 0 ~ 20 cm layer decreased from 1.441 to 0.911 g/L when the ensemble size increased from 5 to 30, whereas this value varied between 0.819 and 0.911 g/L with the ensemble size ranging between 30 and 80. The average RMSE at 0 ~ 100 cm also indicated that the improvement to simulation accuracy was negligible when the ensemble size exceeded 80 (Figure 6(a)). The RMSE reduction rate in Figure 6(b) exhibited the EnKF algorithm's efficiency, and a high RMSE reduction rate meant a considerable improvement in the simulation accuracy. Similar to the RMSE value, the RMSE reduction rate decreased with the ensemble size, indicating that an excessively high ensemble size lowered EnKF efficiency. It was further observed from the average RMSE reduction rate that high EnKF efficiency occurred when the ensemble size was below 30, and the RMSE reduction rate varied little when the ensemble size exceeded 50. Generally, the simulation accuracy was sensitive to ensemble sizes of less than 30. An ensemble size of 50 was quite satisfactory when considering both the RMSE value and its reduction rate.

3.4. Effect of Observation Data Error on Simulation Accuracy

In the EnKF assimilation procedure, the EC_a data is assimilated into the HYDRUS-simulated value, and the blended information of EC_a data and HYDRUS-simulated value is used for state updates. Thus, the observation data quality directly impacts the accuracy and reliability of the EnKF assimilation results. The observation data error, defined as the weight of each observation data's deviation from the background, is an important topic in data assimilation. In this study, error levels of 1.0, 2.5, 5, 10, 15, 20, and 30% were considered for the observed EC_a data. The RMSE of soil salinity between the measured and EnKF values increased with the observation error level (Table 5). With the RMSE between measured values and HYDRUS-simulated values as a reference, significant improvement was observed when the observation error level was below 5%, and the improvement was negligible when the observation error level exceeded 10%. This result showed that

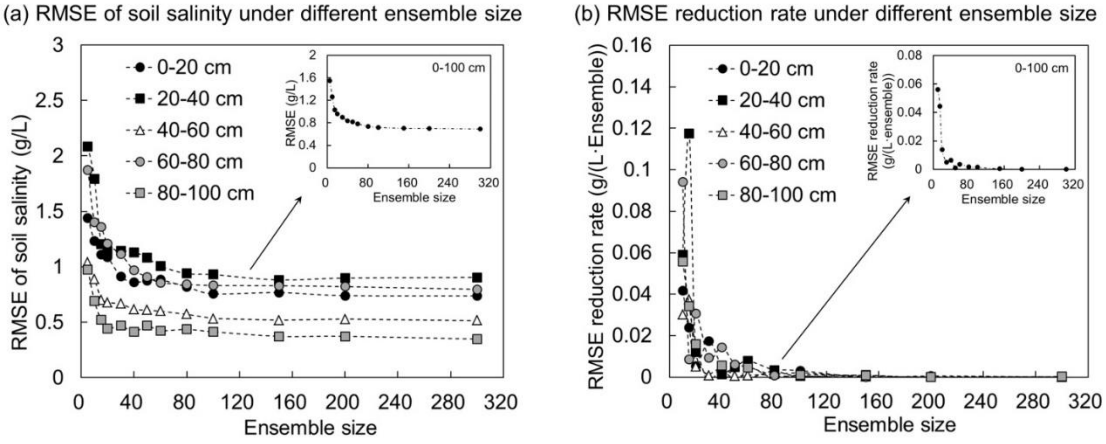

Figure 6. Root-mean-square error (RMSE) and RMSE reduction rate of soil salinity as functions of different ensemble sizes for the 0 ~ 100 cm layers. (a) RMSE of soil salinity under different ensemble size; (b) RMSE reduction rate under different ensemble size. RMSE reduction rate is defined as the reduction in RMSE value with each increment of ensemble size, and the RMSE value under ensemble size 5 is set as the reference.

Table 5. RMSE between Measured Values and EnKF Values under Different Error Levels of Observation Data

Soil layers (cm)	RMSE of soil salinity under different error levels of observation data (g/L)						
	1.0%	2.5%	5%	10%	15%	20%	30%
0 ~ 20	1.030	1.691	1.897	1.957	1.970	1.975	1.980
20 ~ 40	1.151	1.996	2.222	2.306	2.319	2.323	2.328
40 ~ 60	1.072	1.945	2.189	2.255	2.269	2.274	2.279
60 ~ 80	1.188	2.184	2.569	2.690	2.719	2.732	2.740
80 ~ 100	0.582	1.047	1.198	1.254	1.263	1.269	1.273

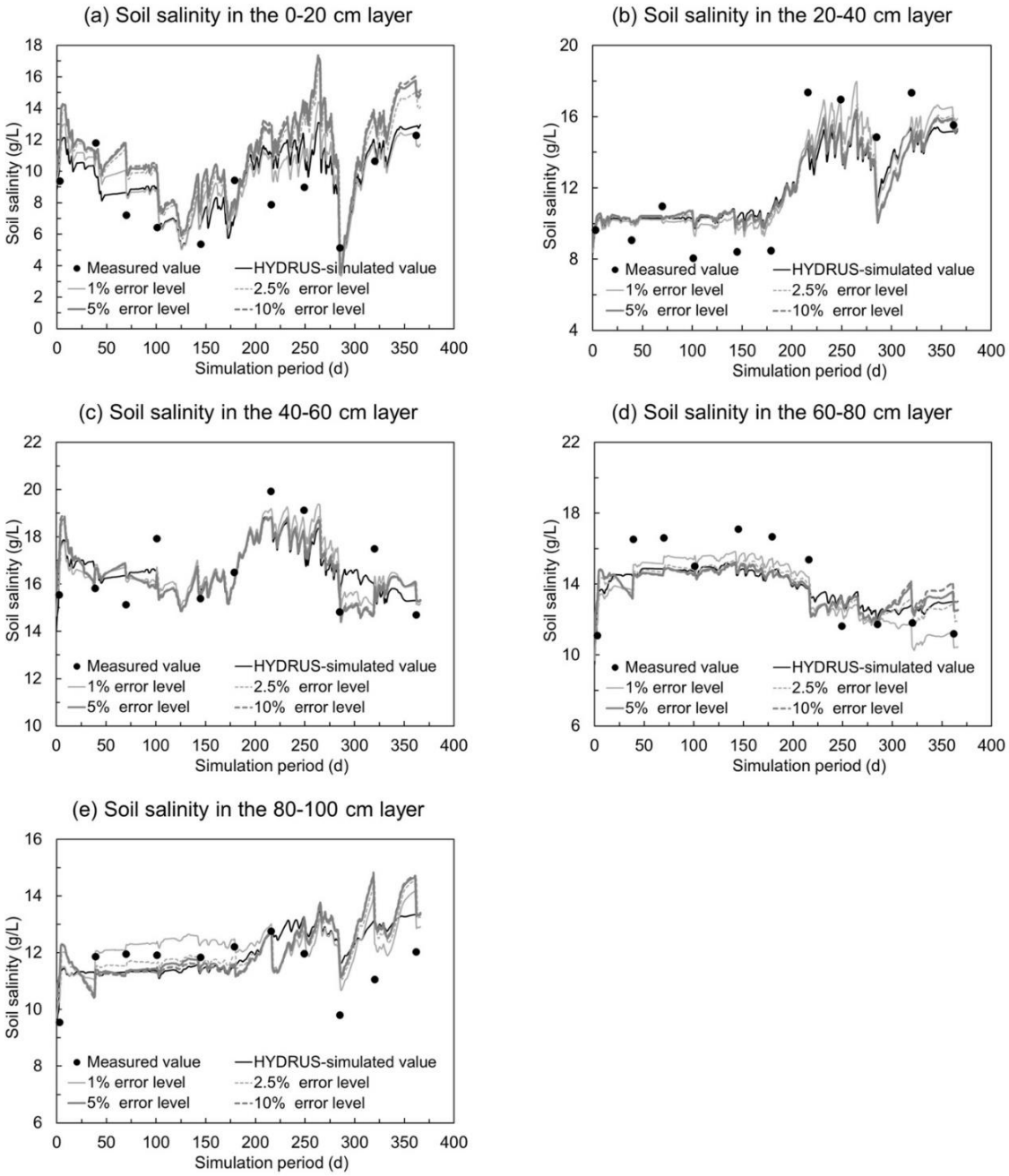

Figure 7. HYDRUS-simulated values vs. HYDRUS-simulated values with EnKF at different error levels of observation data. (a) Soil salinity in the 0 ~ 20 cm layer; (b) soil salinity in the 20 ~ 40 cm layer; (c) soil salinity in the 40 ~ 60 cm layer; (d) soil salinity in the 60 ~80 cm layer; (e) soil salinity in the 80 ~ 100 cm layer.

EnKF assimilation had a positive effect on simulation accuracy under low observation error levels. On the other hand, the simulation performance of EnKF values was close to HYDRUS-simulated values, with an increase in the observation error level for all soil layers, also evident in the HYDRUS-simulated values with EnKF under different observation error levels (Figure 7). Obviously, the HYDRUS-simulated values with EnKF approached the HYDRUS-simulated values when the observation error level increased. The reason was that a high observation error level generally introduced more uncertainty of soil salinity information into the EnKF assimilation system; namely, more errors were incorporated into the background state of the HYDRUS model and propagated forward into the HYDRUS-simulated values with EnKF.

3.5. Effect of the Depth of the EC_a Data on Simulation Accuracy

The depth of the EC_a data used in the EnKF assimilation system also influences soil salinity simulation performance, and EC_a data obtained from lower depths means that more information on soil profile salinity is introduced into the EnKF assimilation. It is clear that soil soluble salts are transported with soil water movement across the profile, and vertical exchange of soil salts exists among different soil layers. The RMSE of soil salinity between the measured and EnKF values decreased when EC_a data from lower depths was employed in the EnKF assimilation (Table 6). For instance, the RMSE of soil salinity in the 80 ~ 100 cm layer increased from 0.414 to 0.617 g/L when the EC_a data from this layer were not included. The RMSE value increased from 0.617 to 0.835 g/L when the EC_a data in the 60 ~ 100 cm layer was not used, and this value increased to 1.156 g/L when the EC_a data in the 20 ~ 100 cm layer were not employed. This indicated that the simulation accuracy of soil salinity at deep layers was sensitive to the reduction of the EC_a data. However, the effect of the EC_a data depth on the simulation accuracy of surface soil salinity was not significant. Considering that the EC_a data in the 0 ~ 20 cm layer was used for all the scenarios, the number of observation data influenced the HYDRUS-simulated values with EnKF in the 0 ~ 20 cm layer, but with little difference (Figure 8). This was also observed for the HYDRUS-simulated values with EnKF in the 20 ~ 40 and 40 ~ 60 cm layers. The possible reason was the correlation among soil salinity in different layers of the profile. Large differences were observed between HYDRUS-simulated values with EnKF in the 60 ~ 80 and 80 ~ 100 cm layers. The HYDRUS-simulated values with EnKF were close to the HYDRUS-simulated values only when the EC_a data in 0 ~ 20 and 0 ~ 40 cm layers were used.

4. Discussions

4.1. Effect of EnKF Assimilation on Soil Salinity Simulation

The present study uses the EnKF algorithm to update the HYDRUS model states and investigates EnKF assimilation influence on soil salinity dynamics simulations. Most previous studies reported that the introduction of EnKF assimilation could improve the simulation performance of HYDRUS model with static and variable soil parameter systems. Huang et al. (2009) used an EnKF assimilation method to update the hydraulic conductivity and state variables and improve solute transport prediction with unknown initial solute source conditions. In the present study, soil water flow simulation was regarded as independent of soil salinity simulation; specifically, soil water flow had little influence on the EnKF update of the temporal soil salinity state. This was done as soil moisture dynamics were adequately simulated using the HYDRUS model, and the present hydraulic parameter system. Another reason is that the state update of hydraulic parameters might result in model divergence in the short assimilation period, although model convergence was not necessarily required in EnKF assimilation of soil salinity.

The flow and transport model coupled with the EnKF assimilation method can provide better prediction of temporal soil moisture (Tran et al., 2013; Yin et al., 2015), hydraulic conductivity (Crestani et al., 2015), and crop leaf area index (LAI) (Zhu et al., 2013). In this study, the EnKF method improved the simulation accuracy of soil salinity by assimilating the proximally sensed EC_a data into the HYDRUS model. The error sources of soil salinity simulation were extensive. First, the error originated from HYDRUS simulation, i.e., model error, including the uncertainty of hydraulic and solute transport parameters, as well as the uncertainty of boundary conditions. As shown in Figure 5, the difference in soil salinity between EnKF values and HYDRUS-simulated values existed. Second, the error was from the observation operators, i.e., observation error. The conversion models relating measured apparent electrical conductivity with soil salinity in different layers were used as the observation operators in the EnKF assimilation system. The success lies in the fact that EM38 measurements are readily correlated with soil salinity. However, the response of EM38 measurements to soil salinity is also influenced by a wide range of indirect factors, such as soil moisture, clay content, bulk density, and mineralogy. Soil moisture varied with precipitation and evaporation during the observation period,

Table 6. RMSE between Measured Values and EnKF Values under Different Depths of Observation Data

Soil layers (cm)	RMSE of soil salinity under different depths of observation data (g/L)				
	0 ~ 20 cm	0 ~ 40 cm	0 ~ 60 cm	0 ~ 80 cm	0 ~ 100 cm
0 ~ 20	1.196	0.960	0.873	0.806	0.757
20 ~ 40	1.448	1.111	1.059	0.999	0.930
40 ~ 60	0.899	0.762	0.606	0.586	0.532
60 ~ 80	1.531	1.303	1.041	0.922	0.831
80 ~ 100	1.156	0.991	0.834	0.617	0.414

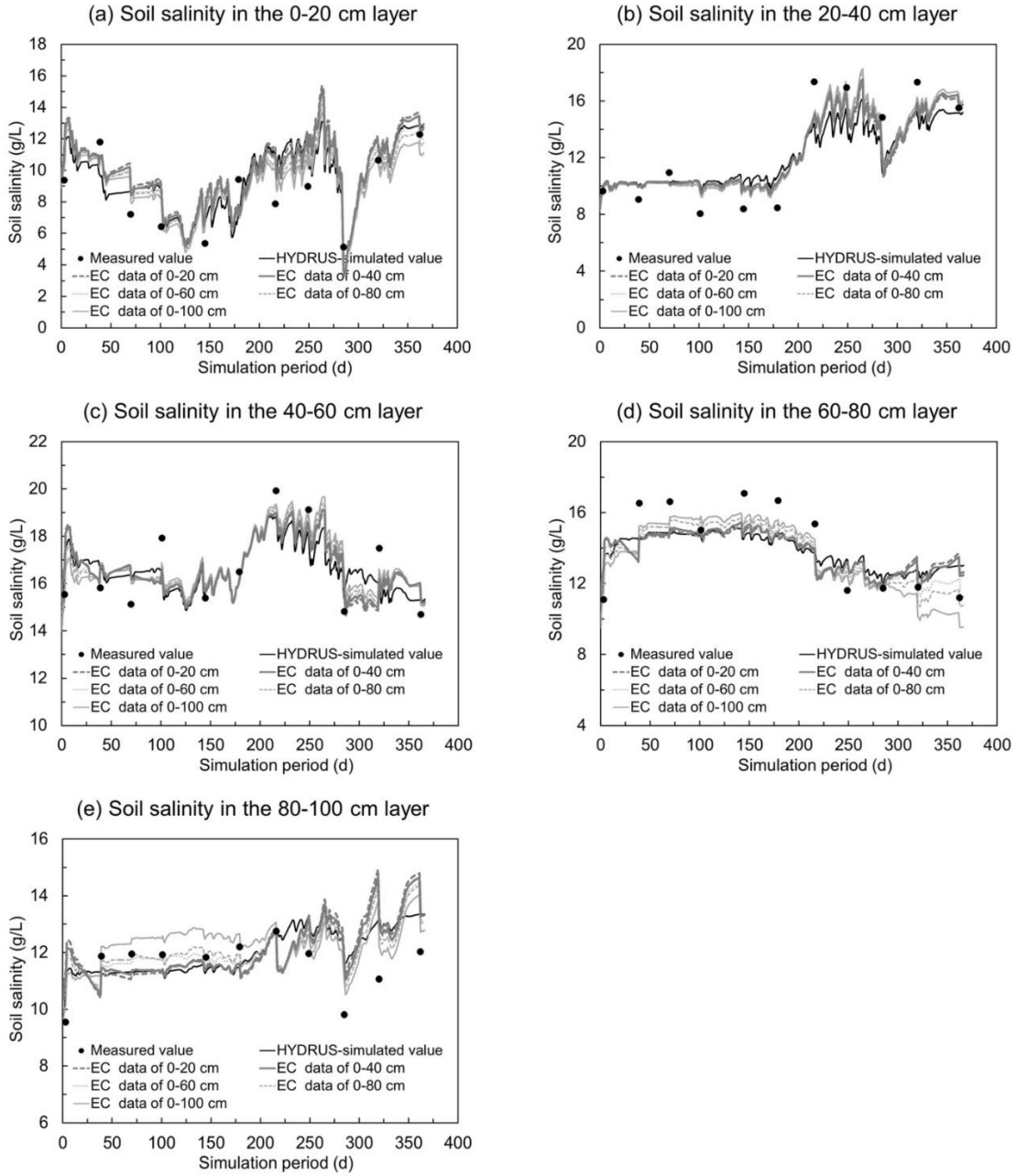

Figure 8. HYDRUS-simulated values vs. HYDRUS-simulated values with EnKF using different observation data depths. (a) Soil salinity in the 0 ~ 20 cm layer; (b) soil salinity in the 20 ~ 40 cm layer; (c) soil salinity in the 40 ~ 60 cm layer; (d) soil salinity in the 60 ~ 80 cm layer; (e) soil salinity in the 80 ~ 100 cm layer.

whereas the same observation operators were used in the assimilation period; that is, the influence of soil moisture on EMI measurements was not considered, and this imported uncertainty into the assimilation system. Third, the error came from the EnKF assimilation system, i.e., assimilation error. The state variable of soil salinity was updated when observation data were available. However, the update of hydraulic conductivity was not considered in this study, resulting in the propagation and accumulation of the simulation error when the HYDRUS model was run. Nevertheless, the EnKF algorithm reduced some of the error by updating the temporal soil salinity state during the simulation procedure.

4.2. Sensitive Parameters of the EnKF Assimilation

The computational cost of the EnKF algorithm scaled with the ensemble size, but excessively reducing the ensemble size resulted in reduced predictive performance. Ensemble size determination often depends on the size of the input data and the available computational power. However, the optimal ensem-

ble size has never been reached as the approach to determine the ensemble size is usually limited, and the selected optimal ensemble size for one case is inapplicable to other cases (Kumar et al., 2008). Houtekamer and Mitchell (1998) stated that a modest number of ensemble sizes (N = 100) could maintain the ensemble representativeness and computational efficiency. Yin et al. (2015) reported that EnKF with ensemble sizes ranging between 12 and 40 had the best computational efficiency in surface water and heat flux assimilation. In the current study, the accuracy of soil salinity simulation was sensitive to an ensemble size of less than 30, and the improvement of the simulation performance was negligible when the ensemble size exceeded 80 (Figure 6). Thus, the maximum efficiency of EnKF could be reached when the ensemble size was set to 50. This is consistent with Mitchell et al. (2002), who used EnKF with an ensemble size of 64 to assimilate radiosonde, satellite, and aircraft data into a dry, global, primitive-equation model. Over the past decade, the boost in computational power has opened the door to new approaches for improving simulation performance, including bagging ensemble filters, entropy ensemble filters, and hierarchical ensemble filters (Foroozand and Weijs, 2017; Wang et al., 2020).

Observation error also played an essential role in soil salinity assimilation, as the merit of data assimilation depended on the accuracy of the observation data used in assimilation. Nonetheless, observation data often contains errors or is applied in a simplified way. We observed that the soil salinity simulation accuracy declined with an increase in observation error level, and no noticeable improvement was observed when the observation error level exceeded 10% (Figure 7). This was in line with Verhoest et al. (2007), who revealed that under considerable uncertainty of the observation data (soil roughness condition), field-averaged soil moisture values with an acceptable accuracy (standard deviation less than 5%) could still be obtained. In another study by Liu et al. (2011), observation errors were found to be sensitive to soil moisture assimilation during the winter wheat season. Some improvements to the simulation performance were still observed when the uncertainty of observation data reached 10%. To overcome the uncertainty of observation data, Verhoest et al. (2007) developed a possibility theory to model inaccurate observation data utilizing possibility distributions. However, success depended on the assumption that the observation data did not change rapidly over time. Moreover, additional field measurements, calibration procedures, and the definition of possibility distributions were required.

The simulation performance of soil salinity was found to be sensitive to the depth of observation data used in the assimilation, especially for soil salinity simulation at deeper layers. Observation data at a depth of 1.0 m was suggested in EnKF assimilation to minimize the RMSE in this study (Figure 8), coinciding with most existing literature. Liu et al. (2011) found that soil moisture assimilation at lower depths of EC_a data had better performance, while the assimilation performance with surface EC_a data was barely satisfactory. Heathman et al. (2003) used surface soil moisture (0 ~ 5 cm) to estimate soil profile moisture (0 ~ 60 cm) and found that assimilation performance was positive at the 0 ~ 30 cm layer, and the improvement in the 30 ~ 60 cm layer was negligible. The reliability of assimilating surface soil attributes to estimate soil profile attributes primarily relied on the connection of hydrological processes in different layers of the soil profile. However, the optimal depth of observation data used for assimilation was associated with several factors, including the selection of the driving model, model parameters, data acquisition quality, observation data frequency, spatial and temporal stability of observation data, and even assimilation methods (Albergel et al., 2017).

It must be noted that the success of our study largely depended on the high accuracy of the EM38 calibration models and relatively homogeneous soil conditions. Whether this method would perform well with another set of data requires further validation. Besides, observation data (EC_a) response to soil salinity was influenced by other factors, such as soil moisture, temperature, bulk density, and texture. The results may be different when this method is applied in other regions and under different hydrogeological conditions. Further efforts will be devoted to evaluating and validating the assimilation effect over a long period and under time-variable soil conditions, and investigating the sensitivity of other factors on assimilation, such as the state update of hydraulic parameters, quality and frequency dependence of observation data, and coupling of multi-source observation data.

5. Conclusions

This study investigated the potential of assimilating proximally-sensed data into the HYDRUS model using the ensemble Kalman filter (EnKF) framework. The EnKF approach directly assimilated the observation data (EC_a) to update the soil salinity status of the HYDRUS model. This is different from most previous studies that only assimilated the surface soil moisture obtained from proximal soil sensing to update the entire soil moisture profile. The assimilation method improved the simulation accuracy of soil salinity in the 0 ~ 100 cm profile. The findings showed that HYDRUS-simulated values with EnKF, promising in the simulation and prediction of soil salinity dynamics, were more accurate than the HYDRUS-simulated values. The sensitivity analysis results showed that soil salinity simulation was sensitive to ensemble size, whereas the error level and depth of the EC_a data were more influential on soil salinity assimilation. An optimal ensemble size of 50, maximum acceptable error level of 10%, and observation data to a depth of 1.0 m are proposed in EnKF assimilation to maximize the ensemble representativeness and simulation performance. This study provides a promising method for simulating large-scale ecological processes using multi-source data and physical hydrological models that is vital for agricultural management and ecological restoration.

Acknowledgments. This study was financially supported by the Natural Science Foundation of China (42077084; U1906221; U1806215; 41701253; 41571223), the Innovation project of the Institute of Soil Science, Chinese Academy of Sciences (CAS) (ISSASIP1633), the National Key Research & Development Program of China

(2016YFC0501300; 2019YFD1002702; 2019YFD0900702) and Taishan Scholars Youth Expert Program of China (201812096). We also thank anonymous reviewers for their valuable comments on this manuscript.

References

Albergel, C., Munier, S., Leroux, D.J., Dewaele, H., Fairbairn, D., Barbu, A.L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Moigne, L.P., Decharme, B., Mahfouf, J.F. and Calvet, J.C. (2017). Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geosci. Model Dev*. 10(10), 3889-3912. https://dx.doi.org/10.5194/gmd-10-3889-2017

Aldabaa, A.A.A., Weindorf, D.C., Chakraborty, S., Sharma, A. and Li, B. (2015). Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geoderma*. 239-240, 34-46. https://dx.doi.org/10.1016/j.geoderma.2014.09.011

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9), D05109. ISBN: 92-5-104219-5

Bauser, H.H., Berg, D., Klein, O. and Roth, K. (2018). Inflation method for ensemble Kalman filter in soil hydrology. *Hydrol. Earth Syst. Sc*. 22(9), 4921-4934. https://dx.doi.org/10.5194/hess-22-4921-2018

Brandhorst, N., Erdal, D. and Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: handling uncertainty of soil hydraulic parameters. *Adv. Water Resour*. 110, 360-370. https://dx.doi.org/10.1016/j.advwatres.2017.10.022

Burgers, G., Peter, J.V.L. and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev*. 126(6), 1719-1724. https://dx.doi.org/10.1175/1520-0493(1998)1262.0.CO;2

Corwin, D.L. and Lesch, S.M. (2014). A simplified regional-scale elec-tromagnetic induction—Salinity calibration model using ANOCOVA modeling techniques. *Geoderma*. 230, 288-295. https://dx.doi.org/ 10.1016/j.geoderma.2014.03.019

Crestani, E., Camporese, M. and Salandin, P. (2015). Assessment of hydraulic conductivity distributions through assimilation of travel time data from ERT-monitored tracer tests. *Adv. Water Resour*. 84, 23-36. https://dx.doi.org/10.1016/j.advwatres.2015.07.022

De Wit, A.D. and Van Diepen, C.A. (2007). Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts. *Agr. Forest Meteorol*. 146(1-2), 38-56. https://dx.doi.org/10.1016/j.agrformet.2007.05.004

Ding, J.L. and Yu, D.L. (2014). Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments. *Geoderma*. 235, 316-322. https://dx.doi.org/10.1016/j.geoderma.2014.07.028

Doolittle, J.A. and Brevik, E.C. (2014). The use of electromagnetic induction techniques in soils studies. *Geoderma*. 223, 33-45. https://dx.doi.org/10.1016/j.geoderma.2014.01.027

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans*. 99(C5), 10143-10162. https://dx.doi.org/10.1029/94JC00572

Feng, G., Zhang, Z., Lu, P. and Bakour, A. (2017). Simulation of farmland groundwater table depth and soil salinity under drainage systems in tidal areas, Laizhou Bay of China. *Irrig. Drain*. 67 (S1), 105-118. https://dx.doi.org/10.1002/ird.2175

Foroozand, H. and Weijs, S.V. (2017). Entropy Ensemble Filter: a modified bootstrap aggregating (Bagging) procedure to improve efficiency in ensemble model simulation. *Entrophy*, 19(10), 520. https://dx.doi.org/10.3390/e19100520

Heathman, G.C., Starks, P.J., Ahuja, L.R. and Jackson, T.J. (2003). Assimilation of surface soil moisture to estimate profile soil water content. *J. Hydrol*. 279(1-4), 1-17. https://dx.doi.org/10.1016/S0022-1694(03)00088-X

Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H. and Goodrich, D.C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Re-sour. Res*. 34(12), 3405-3420. https://dx.doi.org/10.1029/1998WR900001

Houtekamer, P.L. and Mitchell, H.L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev*. 126(3), 796-811. https://dx.doi.org/10.1175/1520-0493(1998)1262.0.CO;2

Huang, C., Hu, B.X., Li, X. and Ye, M. (2009). Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source. *Stoch. Environ. Res. Risk Assess*. 23(8), 1155-1167. http://dx.doi.org/10.1007/s00477-008-0289-4

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J. and Wu, W. (2016). Assimilating a synthetic Kalman filter leaf area index series into the Wofost model to improve regional winter wheat yield estimation. *Agr. Forest Meteorol*. 216, 188-202. https://dx.doi.org/10.1016/j.agrformet.2015.12.043

Huang, J., McBratney, A.B., Minasny, B. and Triantafilis, J. (2017). Monitoring and modelling soil water dynamics using electromagnetic conductivity imaging and the ensemble Kalman filter. *Geoderma*, 285, 76-93. https://dx.doi.org/10.1016/j.geoderma.2016.09.027

Huang, J., Mokhtari, A.R., Cohen, D.R., Monteiro, F.A. and Triantafilis, J. (2015). Modelling soil salinity across a gilgai land-scape by inversion of EM38 and EM31 data. *Eur. J. Soil Sci*. 66(5), 951-960. https://dx.doi.org/10.1111/ejss.12278

Kumar, S.V., Riechle, R.H., Peters-Lidard, C.D., Koster, R.D., Zhan, X., Crow, W.T., Eylander, J.B. and Houser, P.R. (2008). A land surface data assimilation framework using the land information system: Description and applications. *Adv. Water Resour*. 31(11), 1419-1432. https://dx.doi.org/10.1016/j.advwatres.2008.01.013

Li, J.G., Pu, L.J., Han, M.F., Zhu, M., Zhang, R.S. and Xiao, Y.Z. (2014). Soil salinization research in china: advances and prospects. *J. Geog. Sci*. 24(5), 943-960. https://dx.doi.org/10.1007/s11442-014-1130-2

Liu, Z., Zhou, Y.L., Ju, W.M. and Gao, P. (2011). Simulation of cropland soil moisture based on an ensemble Kalman filter. *Chin. J. Appl. Ecol*. 22(11), 2943-2953. https://dx.doi.org/10.1700/1018.11081

Ma, R.J., McBratney, A., Whelan, B., Minasny, B. and Short, M. (2011). Comparing temperature correction models for soil electrical conductivity measurement. *Precis. Agric*. 12(1), 55-66. https://dx.doi.org/10.1007/s11119-009-9156-7

Metternicht, G. (2017). Soils: Salinization. In: Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., and Marston, R.A. (Eds.), *The International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Wiley-Blackwell: Hoboken, NJ, USA, pp. 1-10. ISBN: 9780470659632

Mitchell, H.L., Houtekamer, P.L. and Pellerin, G.R. (2002). Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Weather Rev*. 130(11), 2791-2808. https://dx.doi.org/10.1175/1520-0493(2002)1302.0.CO;2

Narjary, B., Meena, M.D., Kumar, S., Kamra, S.K., Sharma, D.K. and Triantafilis, J. (2019). Digital mapping of soil salinity at various depths using an EM38. *Soil Use Manage*. 35(2), 232-244. https://dx.doi.org/10.1111/sum.12468

Raes, D. (2009). ETo Calculator: a Software Program to Calculate Evapotranspiration from a Reference Surface. *FAO Land Water Division, Digital Media Service*. (36).

Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M. and Li, D. (2007). The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc*. 133(623), 347-362. https://

dx.doi.org/10.1002/qj.32

Reichle, R.H., Crow, W.T. and Keppenne, C.L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res*. 44(3), W03423. https://dx.doi.org/10.1029/2007WR006357

Schaap, M.G., Leij, F.J. and Van Genuchten, M.T. (2001). Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol*. 251(3-4), 163-176. https://dx.doi.org/10.1016/S0022-1694(01)00466-8

Šimůnek, J., Šejna, M., Saito, H., Sakai, M. and Van Genuchten, M.T. (2008). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.0, Hydrus Series 3. *University of California-Riverside Research Reports*, 3, 1-240.

Šimůnek, J., Van Genuchten, M.T. and Šejna, M. (2012). HYDRUS: model use, calibration, and validation. *T. ASABE*. 55(4), 1261-1274. https://dx.doi.org/10.13031/2013.42247

Šimůnek, J., van Genuchten, M.T. and Šejna, M. (2016). Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone J*. 15(7), vzj2016.04.0033. https://dx.doi.org/10.2136/vzj2016.04.0033

Soil Survey Staff. (2010). *Keys to soil taxonomy*. Government Printing Office, pp 173-178. ISBN: 978-0-16-085427-9

Sun, L., Seidou, O., Nistor, I., Goïta, K. and Magagi, R. (2016). Simultaneous assimilation of in situ soil moisture and streamflow in the SWAT model using the Extended Kalman Filter. *J. Hydrol*. 543, 671-685. http://dx.doi.org/10.1016/j.jhydrol.2016.10.040

Thiboult, A. and Anctil, F. (2015). On the difficulty to optimally implement the Ensemble Kalman Filter: An experiment based on many hydrological models and catchments. *J. Hydrol*. 529, 1147-1160. http://dx.doi.org/10.1016/j.jhydrol.2015.09.036

Tran, A.P., Vanclooster, M. and Lambot, S. (2013). Improving soil moisture profile reconstruction from ground-penetrating radar data: a maximum likelihood ensemble filter approach. *Hydrol. Earth Syst. Sci*. 17(7), 2543-2556. http://dx.doi.org/10.5194/hess-17-2543-2013

Triantafilis, J. and Santos, F.M. (2010). Resolving the spatial distribution of the true electrical conductivity with depth using EM38 and EM31 signal data and a laterally constrained inversion model. *Soil Res*. 48(5), 434-446. http://dx.doi.org/10.1071/SR09149

Verhoest, N.E.C., De Baets, B., Mattia, F., Satalino, G., Lucau, C. and Defourny, P. (2007). A possibilistic approach to soil moisture retrieval from ERS synthetic aperture radar backscattering under soil roughness uncertainty. *Water Resour. Res*. 43(7), 931-936. http://dx.doi.org/10.1029/2006WR005295

Wallor, E., Kersebaum, K.C., Lorenz, K. and Gebbers, R. (2018). Soil state variables in space and time: first steps towards linking proximal soil sensing and process modelling. *Precis. Agric*. 20(2), 313-334. http://dx.doi.org/10.1007/s11119-018-9617-y

Wang, H.F., Li, J.W. and He, K. (2020). Hierarchical ensemble reduction and learning for resource-constrained computing. *ACM T. Des. Automat. El*. 25(1), 1-21. https://doi.org/10.1145/3365224

Wang, X., Yang, J., Yao, R. and Yu, S. (2014). Irrigation regime and salt dynamics for rice with brackish water irrigation in coastal region of North Jiangsu Province. *Transactions of the CSAE*. 30(7), 54-63. http://dx.doi.org/10.3969/j.issn.1002-6819.2014.07.007

Wang, Y., Shi, L., Zha, Y., Li, X., Zhang, Q. and Ye M. (2018). Sequential data-worth analysis coupled with ensemble Kalman filter for soil water flow: a real-world case study. *J. Hydrol*. 564, 76-88. http://dx.doi.org/10.1016/j.jhydrol.2018.06.059

Wu, D., Zhou, X. and Jiang, X. (2018). Water and salt migration with phase change in saline soil during freezing and thawing processes. *Groundwater*, 56(5), 742-752. http://dx.doi.org/10.1111gwat.12605

Wu S., Huang J., Liu X., Fan J., Ma G. and Zou J. (2012) Assimilating MODIS-LAI into Crop Growth Model with EnKF to Predict Regional Crop Yield. In: Li D., Chen Y. (eds) Computer and Computing Technologies in Agriculture V. CCTA 2011. *IFIP Adv. Inf. Comm. Te*. 370. Springer, Ber-lin, Heidelberg. https://doi.org/10.1007/978-3-642-27275-2_46

Yao, R.J., Yang, J.S., Wu, D.H., Xie, W.P., Cui, S.Y., Wang, X.P., Yu, S. and Zhang, X. (2015). Determining soil salinity and plant biomass response for a farmed coastal cropland using the electromagnetic induction method. *Comput. Electron. Agr*., 119, 241-253. http://dx.doi.org/10.1016/j.compag.2015.10.020

Yao, R.J., Yang, J.S., Wu, D.H., Xie, W.P., Gao, P. and Wang, X.P. (2016). Geostatistical monitoring of soil salinity for precision management using proximally sensed electromagnetic induction (EMI) method. *Environ. Earth Sci*. 75(20), 1-18. http://dx.doi.org/10.1007/s12665-016-6179-z

Yin, J., Zhan, X., Zheng, Y., Hain, C.R., Liu, J. and Fang, L. (2015). Optimal ensemble size of ensemble Kalman filter in sequential soil moisture data assimilation. *Geophys. Res. Lett*. 42(16), 6710-6715. http://dx.doi.org/10.1002/2015GL063366

Zhou, M., Butterbach-Bahl, K., Vereecken, H. and Brüggemann, N. (2017). A meta-analysis of soil salinization effects on nitrogen pools, cycles and fluxes in coastal ecosystems. *Global Change Biol*. 23(3), 1338-1352. http://dx.doi.org/10.1111/gcb.13430

Zhu, X., Zhao, Y. and Feng, X. (2013). A methodology for estimating leaf area index by assimilating remote sensing data into crop model based on temporal and spatial knowledge. *Chin. Geogra. Sci*. 23(5), 550-561. http://dx.doi.org/10.1007/s11769-013-0621-x