Chin. Geogra. Sci. 2013 Vol. 23 No. 6 pp. 666-679

doi: 10.1007/s11769-013-0623-8

Assimilating ASAR Data for Estimating Soil Moisture Profile Using an Ensemble Kalman Filter

YU Fan, LI Haitao, GU Haiyan, HAN Yanshun

(Key Laboratory of Geo-Informatics of State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, Beijing 100830, China)

Abstract: Active microwave remote sensing data were used to calculate the near-surface soil moisture in the vegetated areas. In this study, Advanced Synthetic Aperture Radar (ASAR) observations of surface soil moisture content were used in a data assimilation framework to improve the estimation of the soil moisture profile at the middle reaches of the Heihe River Basin, Northwest China. A one-dimensional soil moisture assimilation system based on the ensemble Kalman filter (EnKF), the forward radiative transfer model, crop model, and the Distributed Hydrology-Soil-Vegetation Model (DHSVM) was developed. The crop model, as a semi-empirical model, was used to estimate the surface backscattering of vegetated areas. The DHSVM is a distributed hydrology-vegetation model that explicitly represents the effects of topography and vegetation on water fluxes through the landscape. Numerical experiments were conducted to assimilate the ASAR data into the DHSVM and in situ soil moisture at the middle reaches of the Heihe River Basin from June 20 to July 15, 2008. The results indicated that EnKF is effective for assimilating ASAR observations into the hydrological model. Compared with the simulation and in situ observations, the assimilated results were significantly improved in the surface layer and root layer, and the soil moisture varied slightly in the deep layer. Additionally, EnKF is an efficient approach to handle the strongly nonlinear problem which is practical and effective for soil moisture estimation by assimilation of remote sensing data. Moreover, to improve the assimilation results, further studies on obtaining more reliable forcing data and model parameters and increasing the efficiency and accuracy of the remote sensing observations are needed, also improving estimation accuracy of model operator is important.

Keywords: assimilation; ensemble Kalman filter (EnKF); soil moisture; hydrological model; Advanced Synthetic Aperture Radar (ASAR)

Citation: Yu Fan, Li Haitao, Gu Haiyan, Han Yanshun, 2013. Assimilating ASAR data for estimating soil moisture profile using an ensemble Kalman filter. *Chinese Geographical Science*, 23(6): 666–679. doi: 10.1007/s11769-013-0623-8

1 Introduction

Soil moisture is one of the most important variables in describing the water and energy exchange at the land surface/atmosphere interface. Soil moisture is widely recognized as a key parameter in numerous disciplines, including meteorology, hydrology, ecology, and agriculture (Delworth and Manabe, 1988; Sellers and Schimel, 1993; Brubaker and Entekhabi, 1996). Hence,

soil moisture can be determined from point measurements, hydrologic models, and remote sensing techniques. Traditional point measurements can only provide information about that specific point, and impractical for building a dense network of point observations. Hydrologic models simulate the continuous variation of the spatial distribution and temporal evolution of soil moisture, but the uncertainties of the model's initial parameters, the nonlinear nature of land-atmosphere inter-

Received date: 2012-07-23; accepted data: 2013-01-09

Foundation item: Under the auspices of National Natural Science Foundation for Young Scientists of China (No. 41101321), Major State Basic Research Development Program of China (No. 2007CB714407), Key Projects in the National Science & Technology Pillar Program (No. 2009BAG18B01, 2012BAH28B03)

Corresponding author: YU Fan. E-mail: yufan@casm.ac.cn

© Science Press, Northeast Institute of Geography and Agroecology, CAS and Springer-Verlag Berlin Heidelberg 2013

actions, and the heterogeneity of the land surface cause deviations in the results. Moreover, the errors in hydrologic models gradually accumulate when the model goes forward. The remote sensing technology captures surface soil moisture over large areas and periods. Deriving soil moisture by remote sensing data has been explored in many studies, such as those by Ulaby *et al.* (1978), Dobson and Ulaby (1986), Kogan (1990), England *et al.* (1992), and Sandholt *et al.* (2002). However, the sensors only detect signals above the soil surface, and remote sensing observation is conducted periodically, not continuously.

The data assimilation theory provides an effective method for merging observations and model simulations. Multi-temporal large-area observations by using remote sensing efficiently reduce the uncertainty of initial state and parameters of hydrologic models. Different assimilation methods have been used to estimate the soil moisture profile based on near-surface soil moisture observations in several previous studies (Entekhabi et al., 1994; Houser et al., 1998; Galantowicz et al., 1999; Walker and Willgoose, 2001; Heathman et al., 2003; Yang and Shen, 2009). Currently, variation filtering and Kalman filtering are the most popular algorithms for data assimilation. Both methods require simplified or linearized model operators and observation operators when they deal with the high nonlinearity and discontinuity of model operators and observation operators, causing inaccurate assimilation results. For nonlinear dynamic problem, an extended Kalman filter was developed (Miller et al., 1994), but it is greatly unstable if the nonlinearities are strong, and the method needs huge computation for large-scale land surface systems. To overcome this limitation, Evensen (1994) proposed the ensemble Kalman filter (EnKF), which was a sequential data assimilation method that predicts the state error covariance via the Monte Carlo approach. Verlaan and Heemink (2001) compared the performance of two Kalman filters, and pointed out that the EnKF was more effective for strong nonlinear models. Moreover, EnKF directly predicts the state error covariance matrix by applying an ensemble of model states. As an efficient algorithm that handles strongly nonlinear dynamics and large state spaces, EnKF has been widely applied in many disciplines (Li et al., 2004; Zhang, 2006; Huang et al., 2008; Reichle et al., 2008; Crow and Berg, 2010; Li et al., 2010).

Passive microwave brightness data were often used to estimate soil moisture in assimilation experiments (Reichle *et al.*, 2002a; 2002b; Huang *et al.*, 2008) mainly because the passive microwave signal is less dependent on the soil roughness and surface characteristics that highly influence the active microwave signal. However, passive microwave data provide a low resolution of the Earth's surface, which limits the application in earth observations. The problem can be overcome by using active microwave remote sensing data, such as synthetic aperture radar observations. Previous research showed that the error of soil moisture estimation by active microwave for bare soil at C (3.75–7.50 cm) and L band (15–30 cm) is within 5%, which is comparable to most in situ measurement techniques (Mancini *et al.*, 1999).

In this study, an assimilation system based on the EnKF was presented by assimilating active microwave observations into two-dimensional hydrologic model-Distributed Hydrology-Soil-Vegetation Model (DHSVM), for estimating the soil moisture profile on vegetated areas. The semi-empirical radiative transform modelcrop model (Roo and Duetal, 2001) was adopted as the observation operator to describe the surface backscattering of vegetated areas. Assimilation experiments were carried out by using the Advanced Synthetic Aperture Radar (ASAR) observations to test the assimilation system in the Heihe River Basin of Gansu Province in Northwest China from June 20, 2008 to July 15, 2008. This research aims to improve soil moisture estimation through active microwave observations of the surface soil moisture content in a data assimilation framework by using the EnKF.

2 Material and Methods

2.1 Study area

The Heihe River Basin, located in the northwestern China, is the second largest inland river basin with an area of 116 000 km². The region has three major regional distinctions, namely, the upper reaches of the Heihe River Basin in the Qinhai Province, the middle reaches of the Heihe River Basin in Gansu Province, and the lower reaches of the Heihe River Basin in Gansu Province and the Inner Mongolia Autonomous Region. The middle reaches of the Heihe River Basin, located in the Zhangye City of Gansu Province (38°45′–39°15′N, 100°00′–100°45′E) (Fig. 1) was selected as the study

area of this paper, wherein a remote sensing assimilation experiment was conducted over an agriculture site located near Yingke meterological station (38°51′55″N, $100^{\circ}25'11''E$). The elevation of the study area varies between 1000 m and 2000 m above sea level, with average annual temperature of 6°C.

As the Heihe River Basin is surrounded by high mountains, and located in the middle of the Eurasian continent and far from the sea, it has a typical temperate continental climate with scarce precipitation and strong evaporation. Annual mean precipitation is 121.5 mm, and the average annual potential evaporation exceeds 2340 mm, which is 20 times as the average precipitation in the middle reaches of Heihe River Basin. Up to 95% of the water in the Heihe River Basin is replenished by precipitation and snowmelt. Up to 35 rivers originate from the Qilian Mountain in the Heihe River Basin.

Among them, there are 18 rivers with a covering area exceeding 100 km² and seven rivers with the runoff exceeding 1.0×10^8 m³/yr. The Heihe River is the only source of surface runoff in the Zhangye City. The artificial oases were dominated by irrigated farmlands in the middle reaches of Heihe River Basin, whereas the lower part has natural oases dominated by meadows and salinized meadows that are relatively small, fragmented, and easily disturbed. Moreover, the decrease in water resources in this region has caused the forest to diminish quickly and the lakes in the downstream area have disappeared some years ago, and turned into marsh-salinized and meadow-salt deserts (Li et al., 2001). The study site mainly consists of irrigated farmland, and it is a part of the low plains in northwestern China. Yingke Town is an oasis with soil consisting of 16.7% sand, 74.8% silt, and 8.5% clay. The vegetative cover of study area is corn.

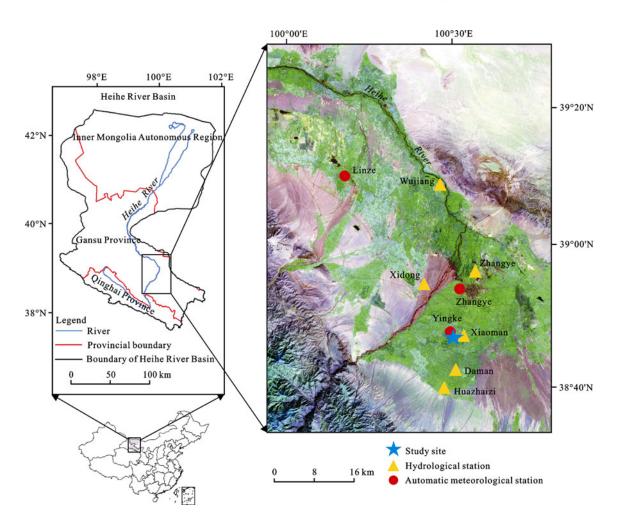


Fig. 1 Location of study area in China and observation stations. Color image on right is a false color picture composited by band 1, 3, and 4 of Landsat TM, which was obtained on July 7, 2008

2.2 Data and processing

2.2.1 Ground observation data

Three automatic meteorological observation stations were established for long-term observations (Fig. 1), recording six meteorological variables required for the DHSVM: relative humidity, air temperature, wind speed, precipitation, incoming shortwave radiation, and incoming longwave radiation. Time domain reflectometers were used in each station to measure the soil moisture at depths of 10 cm, 20 cm, 40 cm, 80 cm, 120 cm, and 160 cm. In this study, the soil volumetric water content at depths of 10 cm, 40 cm, and 80 cm were applied in the assimilation experiment. All ground observation data were obtained from the Watershed Airborne Telemetry Experiment Research of the Chinese Academy of Sciences Action Plan for the West Development Program. The web page (http://westdc.westgis.ac.cn/ water) provides more details about the instrumentation and data collection. The vegetation and soil type data were obtained from the Data Center for Resources and Environment Sciences (DCRES), Chinese Academy of Sciences. The input parameters required by the observation operator are listed in Table 1.

2.2.2 ASAR data

Satellites over the study area provided five ASAR images during the period from June 20, 2008 (day of year,

DOY: 172) to July 15, 2008 (DOY: 197). ASAR is a synthetic aperture radar carried by the ENVISAT-1 satellite and operates in the C-band (central wavelength of 5.63 cm), with multi-polarization, seven observation angles, and five operating modes. In this study, the ASAR data is in the alternating polarization operating mode (product code ASAR_APP_1P), corresponding to two kinds of polarization with resampled high spatial resolution of 12.5 m. The characteristics of the ASAR data in this experiment are given in Table 2.

The ASAR data were preprocessed by using EnviView software. The ASAR images were filtered by using the Lee filter (Lee *et al.*, 1994) to suppress the speckle noise. The backscattering coefficient of ASAR was obtained after radiometric calibration. In the geometric registration, the ASAR images were corrected by using the panchromatic orthophoto SPOT with a 2.5-meter spatial resolution. Registration error was within a pixel.

2.3 Methods

This section presents the strategy for soil moisture estimation using data assimilation algorithm. As mentioned in the introduction, DHSVM was used as the system model operator to describe the moisture in the unsaturated zone of homogeneous isotropic soil. The observation equation aimed to build the relationship between

 Table 1
 Input parameters required by observation operator

Parameter	Definition	Value	Unit
λ	Wavelength	5.63	cm
θ	Incident angle	14.2-42.7	deg
W_{can}	Vegetation water content	0.30-0.45	kg/m ³
k	Extinction coefficient of vegetation canopy	0.10-0.35	Np/m
h	Vegetation canopy height	0.5-1.6	m
S	Root mean square (RMS) height of surface roughness	1.3-4.2	cm
cl	Correlation length of surface roughness	8.77	cm
$ ho_b$	Soil bulk density	1.52	kg/m ³
$ ho_s$	Soil specific density	2.8	kg/m^3
S	Sand percentage	17	%
C	Clay Percentage	75	%

Table 2 Overview of acquired ASAR data from Yingke station

Acquisition date	Day of year (DOY)	Polarization mode	Spatial resolution (m)	Incidence angle (°)
06/25/2008	177	HV/HH	12.5	25.7–31.2
06/28/2008	180	HV/HH	12.5	14.2–22.3
07/05/2008	187	VV/VH	12.5	42.4–45.2
07/08/2008	190	VV/VH	12.5	38.9-42.7
07/11/2008	193	VV/VH	12.5	30.8–36.2

simulated state variables and observations; thus, the crop model and the advanced integral equations model (AIEM) were used to determine the relationship between the satellite observations and soil moisture in the surface layer. EnKF was applied to integrating the simulation and observation by utilizing observation information to update the state variables. The schematic description of the assimilation of the ASAR data with EnKF is given in Fig. 2. Pseudorandom noise with prescribed statistics was added to the first-guess initial soil moisture to generate the ensemble of initial soil moisture. Then, DHSVM was driven by the initial soil moisture profile, model parameters, and atmospheric forcing data. Therefore, the ensemble of forecast soil moisture

profile can be obtained. At each time when ASAR observations were available, the simulated backscattering coefficient was calculated by using the crop model and AIEM based on the forecasted surface soil moisture and other parameters required by observation operators. Hence, the simulated and observed backscattering coefficients were then fed into the EnKF to update the soil moisture profile. The updated soil moisture profiles were then used to reinitialize the model in subsequent runs until microwave observations were available again.

2.3.1 Distributed Hydrology-Soil-Vegetation Model

Distributed Hydrology-Soil-Vegetation Model explicitly represents the effects of topography and vegetation on

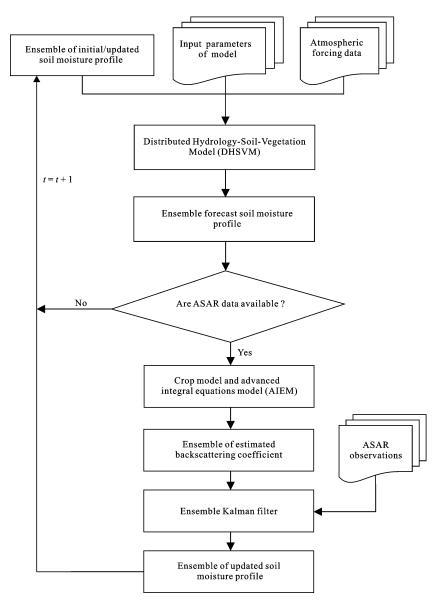


Fig. 2 Flowchart of assimilating Advanced Synthetic Aperture Radar (ASAR) observations with ensemble Kalman filter

water fluxes through the landscape. The model includes canopy interception, evaporation, transpiration, and snow accumulation and melt, as well as runoff generation. DHSVM, as an open source hydrologic model, provides a dynamic representation of watershed processes at the spatial scale described by digital elevation model (DEM) data. The modeled landscape is divided into computational grid cells centered on DEM nodes. Topography characterization is used to model topographic controls on absorbed shortwave radiation, precipitation, air temperature, and downslope water movement. Vegetation characteristics and soil properties are assigned to each model grid cell. These properties may vary spatially throughout the basin. At each time step, the model provided simultaneous solutions to energy and water balance equations for each grid cell in the watershed. The individual grid cells were hydrologically linked through surface and subsurface flow routing (Wigmosta et al., 1994).

Canopy snow interception and release was simulated by using a one-layer mass and energy balance model. Snow accumulation was simulated by using a two-layer energy and mass balance model. Evaporation and transpiration followed the Penman-Monteith equation. Unsaturated moisture movement through multiple soil rooting zone layers was calculated by using Darcy's Law. The water discharge from the lower rooting zone recharges the local water table. Each grid cell exchanges water with its adjacent neighbors. Return flow and saturation overland flow are generated in locations where grid cell water tables intersect the ground surface (Wigmosta *et al.*, 2002).

The dynamics of unsaturated moisture movement in the DHSVM were estimated by using a two-layer model, and the mass balance equations for the surface soil layer and the root zone are given by Wigmosta *et al.* (1994):

$$d_{1}(m_{v1}^{t+\Delta t} - m_{v1}^{t}) = I_{\Delta t} - Q_{v}(\theta_{1}) - \sum_{j=1}^{2} f_{rj1} E_{tj} - E_{s} + V_{ex2} - V_{ex1}$$

$$(1)$$

$$d_{k}(m_{vk}^{t+\Delta t} - m_{vk}^{t}) = Q_{v}(m_{v(k-1)}) - Q_{v}(m_{vk}) - \sum_{j=1}^{2} f_{tjk} E_{tj} + V_{ex(k+1)}$$
(2)

where d_k (k = 1, 2, 3) is the soil layer thickness in the soil layer k; m'_{vk} (k = 1, 2, 3) is the average soil moisture in the soil layer k at time t; $I_{\Delta t}$ is the volume of water infiltrated during the time step Δt ; Q_v is the volume of water discharged downward to the next layer; f_{tjk} is the fraction of roots from vegetation layer j in soil layer k; V_{exk} is the volume of water supplied by a rising water table in soil layer k; E_{tj} is the volume of evaporation from vegetation layer j during the time step Δt ; and E_s is the volume of evaporated soil moisture from the surface layer. The model first calculates infiltration into the upper layer, and then the downward vertical moisture transfermation moving from top to bottom.

2.3.2 Ensemble Kalman filter

In this study, an improved EnKF proposed by Burgers *et al.* (1998) was adopted as the assimilation algorithm. It was based on Monte Carlo method and formulated with nonlinear dynamics as it can properly handle the error covariance evolution in nonlinear models. The process of the EnKF algorithm mainly includes forecast and analysis. The steps of procedure of EnKF are probably as follows:

(1) First, The initial ensemble of state variables should be generated. X_0^a is given as the first-guest value by the Monte Carlo method. Then each of the initial state variables in the ensemble $X_{i,0}^a$ can be determined by adding random noise to X_0^a :

$$X_{i,0}^{a} = X_{0}^{a} + \xi_{i} \qquad \xi_{i} \sim N(0,P)$$
 (3)

where ξ_i is the background error vector, a Gaussian distribution with a mean of zero and covariance matrix P. The superscript 'a' represents state variables of analysis.

(2) In the forecast step, a model operator was used to update each ensemble member, $X_{i,t+1}^{f}$. The forecasted state variable at time t+1 can be calculated by the analyzed state variable $X_{i,t}^{a}$ at former time t with the equation:

$$X_{i,t+1}^{f} = M(X_{i,t}^{a}, a_{t+1}, \beta_{t+1}) + v_{i} \quad v_{i} \sim (0, U)$$
 (4)

where superscript 'f' refers to state variable of forecast; M() refers to the model operator; a_{t+1} is the atmospheric forcing data; β_{t+1} is model parameter and v_i is the model error vector, which is supposed to be a Gaussian distribution with a mean of zero and covariance matrix U.

(3) In the analysis step, the mean of $\{X_{i,t+1}^f\}$ at time t+1 was calculated for its covariance matrix. The observations are related to the true state $(X_{i,t+1}^f)$ through:

$$Y_{i,t+1} = H(X_{i,t+1}^{f}) + w_i \quad v_i \sim (0,R)$$
 (5)

where H() is the observation operator; $Y_{i,t+1}$ is the observation at time t+1; and w_i is the model error vector, which is a (0, R) Gaussian distribution. Then the analyzed state variable is updated with following equations.

$$X_{i,t+1}^{a} = X_{i,t+1}^{f} + K_{t+1}[(Y_{t+1} - H(X_{i,t+1}^{f}) + W_{i}]$$
(6)

$$K_{t+1} = P_{t+1}^{f} H^{T} (H P_{t+1}^{f} H^{T} + R)^{-1}$$
 (7)

$$P_{t+1}^{f} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{i,t+1}^{f} - \overline{X_{t+1}^{f}})$$

$$(X_{i,t+1}^{f} - \overline{X_{t+1}^{f}})^{T}$$
(8)

where K_{t+1} is the Kalman gain matrix at time t+1; Y_{t+1} is the observation data at time t+1; P_{t+1}^f and X_{t+1}^f are the covariance matrix and expectation of forecasted state vector of ensemble members $\{X_{i,t+1}^f\}$ at time t+1, respectively; $H^T()$ is the transpose of H(); N is the number of state variable X_i .

(4) Ensemble members were integrated independently and updated in accordance with the Kalman filter method when new observations become available, and the process was repeated.

2.3.3 Observation operator

The fundamental basis of microwave remote sensing for soil moisture is the relationship between soil dielectric properties and volumetric soil moisture content. Many approaches have been developed for modeling the microwave backscattering from vegetated areas (Ulaby *et al.*, 1984; 1990; Roo and Duetal, 2001). The crop model, which is based on the simplification of the Michigan microwave canopy scattering model, was chosen to calculate the vegetation backscattering. The crop model concluded the backscattering contribution from the interactions between vegetation canopy and ground, unlike the water-cloud model. Thus, the crop model ignores the scattering component associated with ground-trunk scattering by the following assumptions: 1) the stalk and crown of crop are not quite different and can

be treated as one layer, which are distributed uniformly throughout the volume; 2) vegetation canopy has a physical vertical thickness. For a given incident angle θ , the backscattering coefficient of the crop model (σ_{pq}^0) could be generally presented as follows:

$$\sigma_{pq}^{0} = \sigma_{pq1}^{0} + \sigma_{pq2}^{0} + \sigma_{pq3}^{0} + \sigma_{pq4}^{0}$$
 (9)

where p, q are the polarized configuration; σ_{pq1}^0 is the direct backscattering contribution from the canopy; σ_{pq2}^0 is the sum of the ground-canopy and canopyground forward scattering contribution; σ_{pq3}^0 is the ground-canopy-ground scattering contribution; and σ_{pq4}^0 is the direct backscattering contribution of the underlying soil surface (including two-way attenuation by the canopy). These components are as follows:

$$\sigma_{pq1}^{0} = \frac{\sigma_{pq1} \cos \theta}{k_{p} + k_{q}} (1 - T_{p} T_{q})$$
 (10)

$$\sigma_{pq2}^{0} = 2T_{p}T_{q}(\Gamma_{p} + \Gamma_{q})h\sigma_{pq2}$$
 (11)

$$\sigma_{pq3}^{0} = \sigma_{pq1}^{0} T_{p} T_{q} \Gamma_{p} \Gamma_{q}$$

$$\tag{12}$$

$$\sigma_{\text{po4}}^0 = \sigma_{\text{pos}}^0 T_{\text{p}} T_{\text{q}} \tag{13}$$

$$T_{\rm p} = \exp(-k_{\rm p} h \sec \theta) \tag{14}$$

$$\Gamma_{\rm p} = \Gamma_{\rm p0} \exp[-(2ks\cos\theta)^2] \tag{15}$$

where σ_{pq1} is the backscatter cross section per unit volume of the leaves and stem (m²/m³); σ_{pq2} is the bistatic cross section per unit volume of the leaves and stems (m²/m³); k_p (k_q) is the p-polarized (q-polarized) extinction coefficient of vegetation canopy (Np/m); T_p (T_q) is the p-polarized (q-polarized) one-way transmissivity of the canopy; h is the canopy height (m); Γ_p (Γ_q) is the p-polarized (q-polarized) Fresnel reflectivity of ground surface; Γ_{p0} is the Fresnel reflectivity of a specular surface; k is the wave number ($k = 2\pi / \lambda$); s is the root mean square (RMS) height of ground surface (m), and σ_{pqs}^0 is the backscattering coefficient of soil surface in the absence of vegetation cover.

As the most popular method for calculating electromagnetic scattering on the real natural surface (Fung *et al.*, 1992; Wu *et al.*, 2008), AIEM was used to estimate

 σ_{pqs}^{0} , its single scattering term is given by the following equations:

$$\sigma_{pqs}^{0} = \frac{k_{1}^{2}}{2} \exp(-s^{2}(k_{z}^{2} + k_{sz}^{2}))$$

$$\sum_{n=1}^{\infty} \frac{s^{2n}}{n!} |I_{pq}^{n}|^{2} W^{n}(k_{sx} - k_{x}, k_{sy} - k_{y})$$
(16)

$$I_{pq}^{n} = (k_{sz} + k_{z})^{n} f_{pq} \exp(-s^{2}k_{z}k_{sz}) + \frac{(k_{sz})^{n} F_{pq}(-k_{x}, -k_{y}) + (k_{z})^{n} F_{pq}(-k_{sx}, -k_{sy})}{2}$$

where k_1 is the wave number in medium 1; $W^n(k_{sx} - k_x, k_{sy} - k_y)$ is the roughness spectrum of the surface related to the nth power of the surface correlation function by the Fourier transforms; f_{pq} and F_{pq} are the Kirchhoff coefficient and the complementary field coefficient, respectively; $k_z = k\cos\theta$, $k_{sz} = k\cos\theta_s$, $k_x = k\sin\theta\cos\varphi$, $k_{sx} = k\sin\theta\cos\varphi$, $k_y = k\sin\theta\sin\varphi$, and $k_{sy} = k\sin\theta\sin\varphi$, $k_{sy} = k\sin\theta\sin\varphi$, and $k_{sy} = k\sin\theta\sin\varphi$, and $k_{sy} = k\sin\theta\sin\varphi$, are the scattering angle and scattering azimuth angle, respectively; and $k_{sy} = k\sin\theta\sin\varphi$, and $k_{sy} = k\sin\theta\sin\varphi$, are the scattering angle and scattering azimuth angle, respectively; and $k_{sy} = k\sin\theta\sin\varphi$, and $k_{sy} = k\sin\theta\sin\varphi$, and soil dielectric constant ($k_{sy} = k\sin\theta\sin\varphi$).

 $\varepsilon_{\rm m}$ can be transformed into soil moisture by using the formula proposed by Dobson *et al.* (1985).

$$(\varepsilon_{\rm m})^a \cong 1 + \frac{\rho_{\rm b}}{\rho_{\rm s}} ((\varepsilon_{\rm s})^a - 1) + (m_{\rm v})^\beta (\varepsilon_{\rm fw})^a - m_{\rm v}$$

$$(18)$$

$$\beta = 1.09 - 0.11S + 0.18C \tag{19}$$

where ρ_b is the soil bulk density; ρ_s is the soil specific density; ε_s is the dielectric constant of soil with extremely low moisture content [$\varepsilon_s \simeq (4.7, 0)$]; ε_{fw} is the dielectric constant of free water; m_v is the soil moisture content. In this study, a = 0.65 and β is determined by soil texture; S and C are the percentages of sand and clay (In this paper, S = 16.7%, C = 8.5%), respectively.

3 Experiments and Results

3.1 Assimilation of ASAR observation

3.1.1 Size of ensemble

The Monte Carlo method was used to solve nonlinear problems in the EnKF; thus, more samples can describe the spatial distribution of the state variables accurately.

However, only increasing the size of the ensemble increases the number of calculation. Previous studies have shown that when the size of ensemble exceeds the limit, the assimilation results stabilize with no substantial improvements. In fact, the appropriate ensemble size for data assimilation has not been established. Evensen (1994) chose an ensemble size of 100 in his assimilation experiment and pointed out that this value met the general application. Based on previous studies, the proper ensemble size should balance the precision and computation time of assimilation. In this study, the ensemble size was adjusted, whereas the assimilation cycle, the background error model, and the observation error were unchanged. The soil moisture was assimulated with different numbers of samples, and the root mean square error (RMSE) values of the assimilated soil moisture are shown in Fig. 3. At small ensemble sizes, the RMSE converged quickly with increasing ensemble size in the surface layer, root layer, and deep layer. As the number reached 120, the results stabilized and fluctuated slightly. Further increase in ensemble size did not improve the assimilation results, but the computational burden greatly increased. The processing time with 500 ensemble samples was twice as that with 100 samples. In accordance with the tests above, the size of the ensemble was set to 120, which was also close to those of other published ensemble experiments (Haugen and Evensen, 2002; Etienne and Dombrowsky, 2003; Huang et al., 2008).

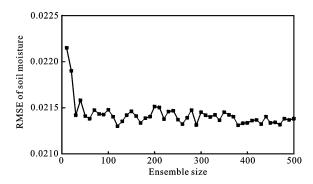


Fig. 3 Error analysis of assimilated soil moisture at different ensemble sizes. RMSE is root mean square error

3.1.2 Determination of observation error and model error

Before conducting the assimilation experiments, the surface soil moisture was inversed by the observation operator and ASAR data. The results were compared with in situ soil moisture, as shown in Table 3.

Table 3 Results of inversed soil moisture content by observation operator and ASAR data

Soil moisture	06/25/2008	06/28/2008	07/05/2008	07/08/2008	07/11/2008
Inversed value	0.254	0.248	0.232	0.230	0.201
In situ	0.231	0.219	0.223	0.221	0.223

Table 3 shows that the inversion soil moisture was very accurate, and the ASAR data are suitable for soil moisture assimilation. Implementing an EnKF requires information on both observation and model errors, which is crucial for correctly assimilating observations into models. The observation error and model error are considered mutative at different times; thus, quantifying the errors for both model and observation errors is very difficult. The in situ observations were assumed perfect, a month of data from May 21 to June 20, 2008 when the assimilation experiment began, was used to run the DHSVM. Then, the model background and its error variance were calculated. A Gaussian noise of the initial soil moisture was introduced in all the ensemble members. In this study, the observation variances were given as 5% of the observed backscattering coefficient, which was calibrated by using the measured data set. The model noise was given as 10% of the change in state variables for that particular time step. Driven by forcing data and model parameters, the DHSVM was run with ASAR observations at a three-hour time step during the experiment period.

To illustrate the effectiveness of the assimilation scheme, the data assimilation experiments were conducted from June 20 (DOY: 172) to July 15 (DOY: 197), 2008. During the assimilation step, a simulation by using DHSVM was performed at the same time, with no assimilation of ASAR data.

3.2 Comparison of assimilation and simulation results

The soil moisture in the surface layer, root layer, and deep layer from the assimilation and simulation are given in Fig. 4. This scheme accurately estimates the surface soil moisture through data assimilation. The simulated soil moisture was overestimated, while the assimilation results are closer to the in situ observations at the surface layer. In most cases, the soil moisture estimation was pulled closely to the in situ observations by data assimilation when ASAR observations are available. Compared with the in situ observations, the assimilation results at the root layer were much better

than simulation results, which relatively overestimated the soil moisture (Fig. 4b). Unlike the significantly improved estimation of soil moisture in the surface layer and root layer, soil moisture in the deep layer was stable and the assimilation results did not have any improvement (Fig. 4c).

Figure 4a shows that DOY of 185 is a turning point of the simulation results and assimilation results from DOY of 184 to 186. After the turning point, the assimilated soil moisture and simulated soil moisture begin to increase, while the measured soil moisture keeps decreasing during these period. Maybe it is caused by the estimation error of DHSVM, including the atmospheric forcing data error and the uncertainty of model parameters. Some of the atmospheric forcing data are obtained from automatic meterological stations, such as air temperature, wind speed, relative humidity, shortwave radiation, long wave radiation, precipitation. These data are observed continuously. Other atmospheric forcing data were manual measurements; thus, they can not be conducted continuously. Maybe only one measurement for a dynamic parameter was performed for a long time. Parts of the input parameters were substituted with empirical parameters, including hydrological and ecological variables, as well as vegetation and soil parameters. The parameter substitution increased the uncertainty of the simulation results. In addition, the assimilation system exhibited the model error, which would accumulate gradually with the running time, which will cause the inconsistency in simulated and measured soil moisture.

The assimilation and simulation at the three soil layers were validated by calculating the root mean square error (RMSE) and mean bias errors (MBE), and the results are shown in Table 4. At the surface layer, the RMSE and MBE of the simulation were 0.031 and 0.029, respectively. When the satellite data were assimilated, the RMSE and MBE were 0.021 and 0.018, respectively. Therefore, the assimilation results are more accurate than the simulation results and our assimilation method is effective. The soil moisture estimation was significantly improved. At the root layer, the assimilation results were significantly improved compared with

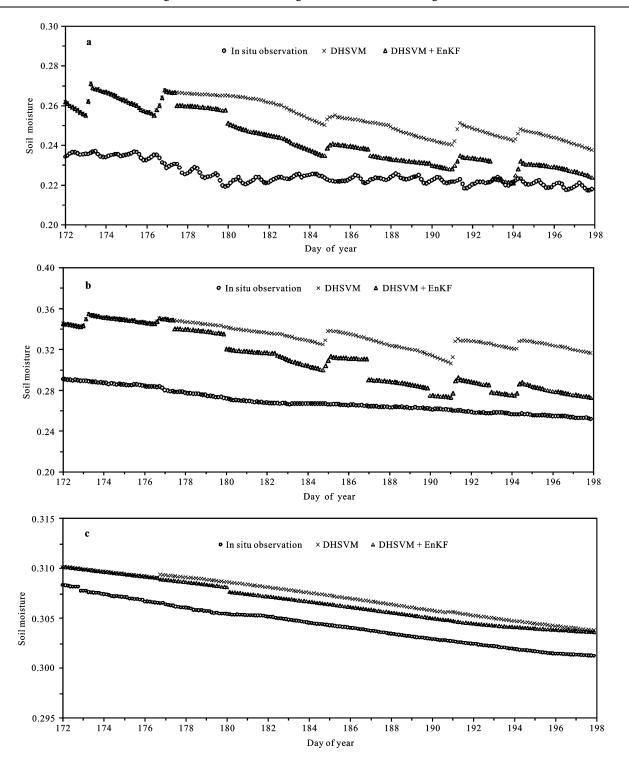


Fig. 4 Soil moisture comparison of in situ observation, Distributed Hydrology-Soil-Vegetation Model (DHSVM) simulation and assimilation results (DHSVM + EnKF) at surface layer (a), root layer (b) and deep layer (c) for Yingke station from June 20 (DOY: 172) to July 15 (DOY: 197) in 2008

simulation results. At the deep layer, the assimilation and simulation results slightly differed because the soil moisture in the deep layer was stable. On the other hand, the given small model errors in the deep layer weakened the information propagation from the surface layer and the soil moisture was weakly influenced by the surface layer.

The assimilation results were influenced by the un-

Table 4 Error statistics of assimilated and simulated soil moisture at Yingke station from June 20, 2008 (DOY: 172) to July 15, 2008 (DOY: 197)

Soil layer	Simulation by DHSVM		Assimilation of ASAR	
Son layer	RMSE	MBE	RMSE	MBE
Surface layer	0.0310	0.0290	0.0210	0.0180
Root layer	0.0650	0.0640	0.0440	0.0410
Deep layer	0.0027	0.0025	0.0025	0.0019

Notes: RMSE is root mean square error; MBE is mean bias error

certainty of the DHSVM and observation errors. DHSVM is the ideal description of the hydrological processes on land surfaces, and some assumptions were made to simplify the actual hydrological processes. For example, the vegetation was divided into two layers, the overstory and the understory, the soil was classified into the surface layer, the root layer, and the saturated layer. The understory layer and surface soil were assumed to not contribute to evapotranspiration. The DHSVM did not include thawing, but soil moisture changed partly by thawing in our assimilation experiment in the Yingke station. Therefore, further studies are needed to improve the accuracy of the model because of the inherent uncertainties in the DHSVM. Although the crop model and the AIEM describe the scattering of the land surface and inversion of the soil moisture in vegetated area, the vegetation canopy greatly affected the inversion results. When the vegetation cover was intensive and the wavelength of microwave was short, the microwave signal could not penetrate the vegetation canopy and detect the ground information. In addition, remote sensing data is influenced easily by weather conditions and sensor errors, which may increase the observation errors from remote sensing.

4 Discussion

Similar studies have recently used the same dataset from Zhangye City of Gansu Province in Northwest China. Liu *et al.* (2010) assimilated the ASAR data with hydrologic model and empirical backscattering model to

estimate soil moisture. In addition, the study area was located near the Yingke station in Zhangye City. Moreover, the study site was a bare area, with soil moisture content less than 5%; thus, the assimilation results were much better and closer to the in situ observations than the simulation results. However, compared with the algorithm mentioned in this paper, Liu's model can not be used in the vegetated areas, only in the bare soil. And it can not get good assimilation results at a moist soil surface $(m_v > 5\%)$. These deficiencies limited its application greatly. Wang et al. (2011) also determined soil moisture by using the multi-temporal ASAR data near Yingke station, and the roughness parameters and soil moisture were obtained simultaneously by using the proposed method. RMSE of soil moisture calculated by Wang's method was about 6%, which is higher than the results derived by the data assimilation in this paper. Hence, the soil moisture estimation was close to the in situ observations in most cases by using the data assimilation. Table 5 shows the detail of soil moisture estimation by the methods mentioned above.

Although the dynamic response of the modeled output was improved by assimilation, the assimilated soil moisture was overestimated in this paper. Considering the simulation results are always greater than the in situ observations, the data assimilation pulls the estimation results close to the measured soil moisture rather than eliminate the effects of the excessively high simulation results. The deviation of the simulated results may be caused in the uncertainty of model parameters (such as vegetation parameters and soil parameters) and atmospheric forcing data (such as air humidity, air temperature, and radiation). For example, soil physical parameters such as the saturated hydraulic conductivity and the porosity are considered homogeneous in the three soil layers by using the DHSVM. The assumption may cause significant errors in the simulation results. Thus, the current model needs to be adjusted to cope with heterogeneous soil characteristics. However, the limited experimental conditions prevented us from obtaining all

Table 5 Comparison of estimated soil moisture with similar works

Method	Study area	RMSE of estimated soil moisture at surface layer	RMSE of estimated soil moisture at root layer	Application region
Algorithm in this paper	Yingke	0.0210	0.0440	Bare soil or vegetated area
Liu's model (Liu et al., 2010)	Yingke	0.0112	0.0025	Bare soil
Wang's method (Wang et al., 2011)	Yingke	0.0600	_	Bare soil

Note: '-' means no data because Wang's method can not simulate soil moisture at deep layer

model parameters, some of which were substituted with empirical parameters. The in situ soil moisture fluctuated during one day, whereas the model simulation hardly showed this variation. The physical model inaccurately simulated the surface environment because of the uncertainty of model parameters and atmospheric forcing data and the real surface environment. In addition, the external environment changed with time. However, the empirical parameters were constant in our model during the same time, so the simulation results did not fluctuate with the in situ soil moisture within one day.

The data assimilation algorithm has a potential draw-back even though it is widely used in hydrological and meteorological applications. The proper distribution of observation error covariance plays a critical role in the correct use of observation data, and it is crucial for the data assimilation. However, the true state variables at different times are unknown, and the covariance of the observation error can not be measured directly and it varies with time. Therefore, only appropriate statistical methods were used to describe the observation error covariance under certain reasonable assumptions. A Gaussian distribution was used to describe the distribution of observation error covariance in this paper, which may introduce uncertainty into the assimilation results.

A method for assimilating ASAR data into the DHSVM for soil moisture profile estimation was proposed at the point scale. Nevertheless, the data assimilation research exhibited some limitations; thus, the improvements could be made in the future.

- (1) Numerous data and parameters are presented in the assimilation system, such as long-term continuous meteorological forcing data, soil parameters, vegetation parameters. Due to the limited experimental conditions, some of the parameters were substituted with the empirical parameters of the study. This may cause errors in the data assimilation. Therefore, more reliable forcing data and model parameters are needed in future studies.
- (2) The assimilation results are influenced considerably by the accuracy of satellite observations which is determined by the quality of satellite data and the precision of the observation operators (crop model and AIEM in this paper). Determining the quality of satellite data and increasing the efficiency and accuracy of the observation operator are needed to improve the assimilation results further.

- (3) The key point of data assimilation algorithm is to minimize statistical error, which mainly comes from the observation noise and the uncertainty of the model. Recent studies on error estimation are relatively weak; thus, further studies on decreasing the uncertainty of the model and observation error are needed to improve the assimilation results.
- (4) The main purpose of assimilating remote sensing data is to take advantage of remote sensing observations over large areas. Therefore, hydrological data assimilation system at the catchment scale should be developed, which is useful for hydrological forecast, water resource management, and other practical applications.

5 Conclusions

A one-dimensional soil moisture assimilation system based on DHSVM and EnKF is presented in this research. The crop model, a semi-empirical radiative transformation model, was adopted as observation operator, which accurately describes the scattering of the vegetated area. The performance of this system was successfully tested and validated by the in situ soil moisture observations from the Heihe River Basin reference site in the northwestern China from June 20 to July 15, 2008.

Analysis of the ensemble size showed that when the size of ensemble exceeded the limit, the assimilated soil moisture exhibited no further improvement. The proper ensemble size in this research was 120, which provided accurate assimilation results, and required less time for assimilation calculation. In addition, the assimilation results were significantly influenced by the observation error and uncertainty of model; therefore, more attention should be given to identifying observation errors and decreasing the uncertainty of model. Additionally, EnKF is an effective and easy-to-implement hydrologic data assimilation scheme for estimating soil moisture profile from satellite observations. Compared with the simulation results of DHSVM, the soil moisture estimation by using the assimilation method was significantly better at the surface layer and root layer, with decreasing average errors and RMSE relative to the simulation results. However, for soil moisture in the deep layer, the assimilation results were nearly the same as the simulation results because soil moisture in the deep layer was stable and unaffected by the soil moisture in the surface

layer.

The active microwave remote sensing data were assimilated practically and effectively into hydrologic models of the vegetated area. Compared with similar studies in the same area, better results were achieved by our assimilation algorithm. However, the uncertainty of the model parameters and atmospheric forcing data should be considered because they may influence the assimilation results. This paper presents a preliminary study of the assimilation of active microwave data for estimating the surface soil moisture profile in vegetated area on the point scale. Further research is needed in several aspects such as obtaining more reliable forcing data and model parameters and increasing the efficiency and accuracy of the remote sensing observations. Then, improving estimation precision of model operator and decreasing the observation error are needed, and the assimilation scheme should be optimized from the point scale to the catchment scale.

References

- Brubaker K L, Entekhabi D, 1996. Analysis of feedback mechanisms in land-atmosphere interaction. *Water Resources Research*, 32(5): 1343–1357. doi: 10.1029/96WR00005
- Burgers G, Leeuwen P J, Evensen G, 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6): 1719–1724. doi: 10.1175/1520-0493(1998)126
- Crow W T, Berg M J, 2010. An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research*, 46(12): 12–51. doi: 10.1029/2010WR009402.
- Delworth T, Manabe S, 1988. The influence of potential evaporation on the variabilities of the simulated soil wetness and climate. *Journal of Climate*, 1(5): 523–547. doi: 10.1175/1520-0442(1988)001
- Dobson M C, Ulaby F T, 1986. Active microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, 24(1): 23–36. doi: 10.1109/TGRS.1986.289585
- Dobson M C, Ulaby F T, Hallikainen M T, 1985. Microwave dielectric behavior of wet soil-Part II: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, 23(1): 35–46. doi: 10.1109/TGRS.1985.289498
- England A W, Galantowicz J F, Schretter M S, 1992. The radiobrightness thermal inertia measure of soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, 30(1): 132–139. doi: 10.1109/36.124223
- Entekhabi D, Galantowicz J F, Njoku E G, 1994. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote*

- Sensing, 32(2): 438-448. doi: 10.1109/36.295058
- Etienne H, Dombrowsky E, 2003. Estimation of the optimal interpolation parameters in a quasi-geostrophic model of the Northeast Atlantic using ensemble methods. *Journal of Marine System*, 40(4): 317–339. doi.org/10.1016/S0924-7963 (03)00023-X
- Evensen G, 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5): 10143–10162. doi: 10.1029/94JC00572
- Fung A K, Lee Z, Chen K S, 1992. Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2): 356–369. doi: 10.1109/36. 134085
- Galantowicz J F, Entekhabi D, Njoku E G, 1999. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4): 1860–1870. doi: 10.1109/36.774699
- Haugen E J, Evensen G, 2002. Assimilation of SLA and SST data into an OGCM for the Indian Ocean. *Ocean Dynamics*, 52(3): 133–151. doi: 10.1007/s10236-002-0014-7
- Heathman G C, Starks P J, Ahuja L R *et al.*, 2003. Assimilation of surface soil moisture to estimate profile soil water content. *Journal of Hydrology*, 27(1): 1–17. doi: 10.1016/S0022-1694 (03)00088-X
- Houser P R, Shuttleworth W J, Gupta H V, 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resource Research*, 34(12): 3405– 3420. doi: 10.1029/1998WR900001
- Huang C H, Li X, Lu L *et al.*, 2008. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sensing of Environment*, 112(3): 888–900. doi: 10.1016/j.rse.2007.06.026
- Kogan F, 1990. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing*, 11(8): 1405–1419. doi: 10.1080/01431169008955102
- Lee S J, Jurkevich L, Dewaele P *et al.*, 1994. Speckle filtering of synthetic aperture radar images: A review. *Remote Sensing Reviews*, 8(4): 313–340. doi: 10.1080/02757259409532206
- Li F Q, Wade T, William P *et al.*, 2010. Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals. *Advances in Water Resources*, 33(2): 201–214. doi: 10.1016/j.advwatres. 2009.11.007
- Li X, Koike T, Mahadevan P, 2004. A very fast simulated re-annealing (VFSA) approach for land data assimilation. *Computer & Geosciences*, 30(3): 239–248. doi: 10.1016/j. cageo.2003.11.002
- Li X, Lu L, Cheng G D *et al.*, 2001. Quantifying landscape structure of the Heihe River Basin, north-west China using FRAGSTATS. *Journal of Arid Environments*, 48(4): 521–535. doi: org/10.1006/jare.2000.0715
- Liu Qian, Wang Mingyu, Zhao Yingshi, 2010. A weighted average soil moisture assimilation experiment based on ensemble

- Kalman filter. *Geography and Geo-Information Science*, 26(1): 94–97. (in Chinese)
- Mancini M R, Hoeben R, Troch P, 1999. Multifrequency radar observations of bare surface soil moisture content: A laboratory experiment. *Water Resources Research*, 35(6): 1827–1838. doi: 10.1029/1999WR900033
- Miller R N, Ghil M, Ghautiez F, 1994. Advanced data assimilation in strongly nonlinear dynamical system. *Journal of the Atmospheric Sciences*, 51(8): 1037–1055. doi: 10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2
- Reichle R H, Crow W T, Christian L *et al.*, 2008. An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, 44(3): 23–34. doi: 10.1029/2007 WR006357
- Reichle R H, McLaughlin D B, Entekhabi D, 2002a. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1): 103–114. doi: 10.1175/1520-0493 (2002)130
- Reichle R H, Walker J P, Koster R D, 2002b. Extended versus ensemble filtering for land data assimilation. *Journal of Hydrometeorology*, 3(2): 728–740. doi: 10.1175/1525-7541 (2002)003
- Roo R D, Duetal Y, 2001. A semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 39(4): 864–872. doi: 10.1109/36.917912
- Sandholt I, Rasmussen K, Andersen J, 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79(2): 213–224. doi: 10.1016/S0034-4257(01) 00274-7
- Sellers P J, Schimel D S, 1993. Remote sensing of the land biosphere and biogeochemistry in the EOS era: Science priorities, methods and implementation. *Global and Planetary Change*, 7(4): 279–297. doi: 10.1016/0921-8181(93)90002-6
- Ulaby F T, Allen C T, Eger G, 1984. Relating microwave backscattering coefficient to leaf area index. *Remote Sensing of Environment*, 14: 113–133.
- Ulaby F T, Batlivala P P, Dobson M C, 1978. Microwave back-

- scatter dependence on surface roughness, soil moisture, and soil texture. *IEEE Transactions on Geoscience and Remote Sensing*, 16(4): 286–295. doi: 10.1109/TGE.1978.294586
- Ulaby F T, Sarahandi K, Donald M K, 1990. Michigan microwave canopy scattering model. *International Journal of Remote Sensing*, 11(7): 1223–1253. doi: 10.1080/0143116900 8955090
- Verlaan M, Heemink A W, 2001, Nonlinearity in data assimilation applications: A practical method for analysis. *Monthly Weather Review*, 129(6): 1578–1589. doi: 10.1175/1520-0493 (2001)129
- Walker J P, Willgoose G R, 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms. *Advances in Water Resources*, 24(6): 631–650. doi: 10.1016/S0309-1708(00)00043-9
- Wang S G, Li X, Han X J et al., 2011. Estimation of surface soil moisture and roughness from multi-angular ASAR imagery in the Watershed Allied Telemetry Experimental Research (WATER). Hydrology and Earth System Sciences, 15(5): 1415–1426. doi: 10.5194/hess-15-1415-2011
- Wigmosta M S, Nijssen P S, Lettenmaier D P, 2002. The distributed hydrology soil vegetation model. In: Singh V P (eds.). Mathematical Models of Small Watershed Hydrology and Applications. Highlands Ranch: Water Resources Press, 7–42.
- Wigmosta M S, Vail L, Lettenmaier D P, 1994. A distributed hydrology-vegetation model for complex terrain. *Water Resource Research*, 30(6): 1665–1679. doi: 10.1029/94WR00436
- Wu T D, Chen K S, Shi J C, 2008. A study of an AIEM model for bistatic scattering from randomly rough surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 46(9): 2584–2598. doi: 10.1109/TGRS.2008.919822
- Yang Shenbing, Shen Shuanghe, 2009. Mapping rice yield based on assimilation of ASAR data with rice growth model. *Journal of Remote Sensing*, 13(2): 282–290. (in Chinese)
- Zhang S W, Li H R, Zhang W D et al., 2006. Estimating the soil moisture profile by assimilating near-surface observations with the ensemble Kalman filter (EnKF). Advances in Atmosphereic Sciences, 22(6): 936–945. doi: 10.1007/BF02918692