Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model

SONG Xiaodong, ZHANG Ganlin[*], LIU Feng, LI Decheng, ZHAO Yuguo, YANG Jinling

State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

Abstract: Soil moisture content (SMC) is a key hydrological parameter in agriculture, meteorology and climate change, and understanding of spatio-temporal distributions of SMC in farmlands is important to address the precise irrigation scheduling. However, the hybrid interaction of static and dynamic environmental parameters makes it particularly difficult to accurately and reliably model the distribution of SMC. At present, deep learning wins numerous contests in machine learning and hence deep belief network (DBN), a breakthrough in deep learning is trained to extract the transition functions for the simulation of the cell state changes. In this study, we used a novel macroscopic cellular automata (MCA) model by combining DBN to predict the SMC over an irrigated corn field (an area of 22 km^2) in the Zhangye oasis, Northwest China. Static and dynamic environmental variables were prepared with regard to the complex hydrological processes. The widely used neural network, multi-layer perceptron (MLP), was utilized for comparison to DBN. The hybrid models (MLP-MCA and DBN-MCA) were calibrated and validated on SMC data within four months, i.e. June to September 2012, which were automatically observed by a wireless sensor network (WSN). Compared with MLP-MCA, the DBN-MCA model led to a decrease in root mean squared error (RMSE) by 18%. Thus, the differences of prediction errors increased due to the propagating errors of variables, difficulties of knowing soil properties and recording irrigation amount in practice. The sequential Gaussian simulation (sGs) was performed to assess the uncertainty of soil moisture estimations. Calculated with a threshold of SMC for each grid cell, the local uncertainty of simulated results in the post processing suggested that the probability of SMC less than 25% will be difference in different areas at different time periods. The current results showed that the DBN-MCA model performs better than the MLP-MCA model, and the DBN-MCA model provides a powerful tool for predicting SMC in highly non-linear forms. Moreover, because modeling soil moisture by using environmental variables is gaining increasing popularity, DBN techniques could contribute a lot to enhancing the calibration of MCA-based SMC estimations and hence provide an alternative approach for SMC monitoring in irrigation systems on the basis of canals.

Keywords: soil moisture; soil moisture sensor network; macroscopic cellular automata (MCA); deep belief network (DBN); multi-layer perceptron (MLP); uncertainty assessment; hydropedology

Soil moisture content (SMC) plays an important role in meteorology, agriculture and climate change. Accurate modeling of the spatio-temporal distributions of surface soil moisture in farmlands is especially essential for precise irrigation scheduling due to the influence of soil moisture on water migration in terms of runoff and infiltration (Sadler et al., 2005). Besides, it also could be integrated with up-to-date techniques of accurate irrigation systems, such as lateral moving irrigators and centre pivot (Kim et al., 2008).

The matching of water inputs to crop requirements is required in precision agriculture, which especially becomes critical to accurately simulate the spatial variability of soil moisture (Stafford, 1996). A stable simulated technique is necessary to predict the complex dynamic of SMC, so that sustainable water management can be utilized to decrease the increasing irrigation water requirement in arid regions. However, SMC in agricultural lands presents heterogeneous change both in space and time, resulting from the superposition of soil, terrain, evapotranspiration, precipitation, irrigation, soil management and fertilization. Recently, many distributed hydrological models were developed and successfully applied at various scales (Famiglietti and Wood, 1994). However, an accurate assessment of SMC might be restricted by the lack of adequate soil water measurements at the scale of interest, and the retrieval of SMC from local direct measurements of distributed and accurate information is costly and time-consuming.

Till now, great details of surface-subsurface groundwater interactions have been studied by using hydrological models (Furman, 2008; Sulis et al., 2010). However, few of which focused on the simulation of the hydrological system in a more comprehensive way and the reproduction of the complex interactions of the vegetation and atmospheric boundary layer (Ivanov et al., 2008). Even if the ecohydrological model considering the feedbacks of water and energy processes might be the perfect simulation tool, the idealized analytical solutions are hard to be implemented and great deployment in monitoring instruments also will be required (Tague and Band, 2004; Mendicino et al., 2013).

Cellular automata (CA) are discrete dynamic systems, in which space is divided into several regular cells with a finite number of states (von Neumann, 1966). The global evolution of each cell is derived from the simultaneous application based on a transition function governing the interactions among each cell and its neighborhoods. Different from the classical CA, macroscopic cellular automata (MCA) use the discrete approach to model various fluid dynamics of soil moisture from the macroscopic perspective (Mendicino et al., 2006; D'Ambrosio et al., 2007; Avolio et al., 2013). Under this framework, numerous studies were performed focusing on different physical problems. D'Odorico and Rodríguez-Iturbe (2000) studied the spatio-temporal variations of soil moisture through the water balance equation at the mesoscale level. With the aim of modeling complex surface-subsurface ecohydrological dynamics, an initial MCA model was presented by adding vegetation dynamic models at the catchment scale (Mendicino et al., 2006), which was tested against both field data and numerical benchmarks (Cervarolo et al., 2010). Recently, Mendicino et al. (2013) added the schematization of flow routing and runoff generation to the aforementioned system (Mendicino et al., 2006; Cervarolo et al., 2010), aiming to develop a full-coupled modeling of surface-subsurface ecohydrological dynamics. For the overland flow model, numerical stability analysis was also performed with respect to the fully coupled ecohydrological system (Mendicino et al., 2015).

Transition function, as one of the most important parts of MCA, addresses the state change of each cell at discrete time steps. Over the past few decades, artificial neural networks (ANNs) have shown considerable advantage for solving the problems with very high degree of complexity, and have often been integrated with CA model for supporting the dynamic modeling of complicated systems, such as aerodynamic modeling (Wu and Kareem, 2011), artificial intelligence (Kim and Cho, 2006; Oku and Aihara, 2010) and land use change modeling (Wang and Li, 2011; Li et al., 2013a).

In recent years, as a new technique of machine learning research, deep learning has won numerous contests (Hinton and Salakhutdinov, 2006; Schmidhuber, 2015) and brilliant achievements in technologies, and has already devoted to making advanced applications available to the public (Ji et al., 2014; Li et al., 2014). The potential benefit of the deep learning is that

many hidden layers of feature representations can be stacked in a deep network, and hence are more capable of modeling the complex structures in the data. Deep belief networks (DBNs) are currently the dominant techniques for modeling the architectural depth of brain, which are often effectively trained in a greedy layer-wise supervised learning manner. However, to date, only few studies have attempted to use these techniques to model the SMC dynamics with respect to the environmental variables.

It is quite evident that a simulation model is required for the estimation of soil moisture content under changing irrigation conditions in the human controlled agro-ecosystems especially the artificial oases. Motivated the prominence of deep learning, this study proposed a DBN (deep belief network)-based MCA (DBN-MCA) model to predict the spatio-temporal patterns of SMC over an irrigated corn field. DBNs were employed to achieve the transition functions depending on climate, topography, soil properties and evaporation. Besides, the hybrid DBN-MCA model was tested based on the observed SMC data during the period from 1 June to 30 September, 2012. The widely used neural network, multi-layer perceptron (MLP)-based MCA (MLP-MCA) model was also carried out for comparison to proposed DBN-MCA model, and the results were evaluated at different scales.

1 Study area and data collection

1.1 Study area

The study area (38°52′15″N, 100°21′47″E; Fig. 1), covering an irrigated corn field with an area of about 22 km^2, is located in the Zhangye oasis, Gansu province, Northwest China. This area is characterized by a typical temperate arid climate with relatively low precipitation and high evaporation. The average annual precipitation varies from 100 to 250 mm, whereas annual potential evaporation ranges from 1,200 to 1,800 mm, which is 10 times higher than annual precipitation (Li et al., 2013b). The major crop of the observation matrix is seed corn (accounting for 68.6% of the total cropland). Other land cover types, including wheat and vegetable fields, villages and irrigation canals, are randomly interspersed in the crop fields.
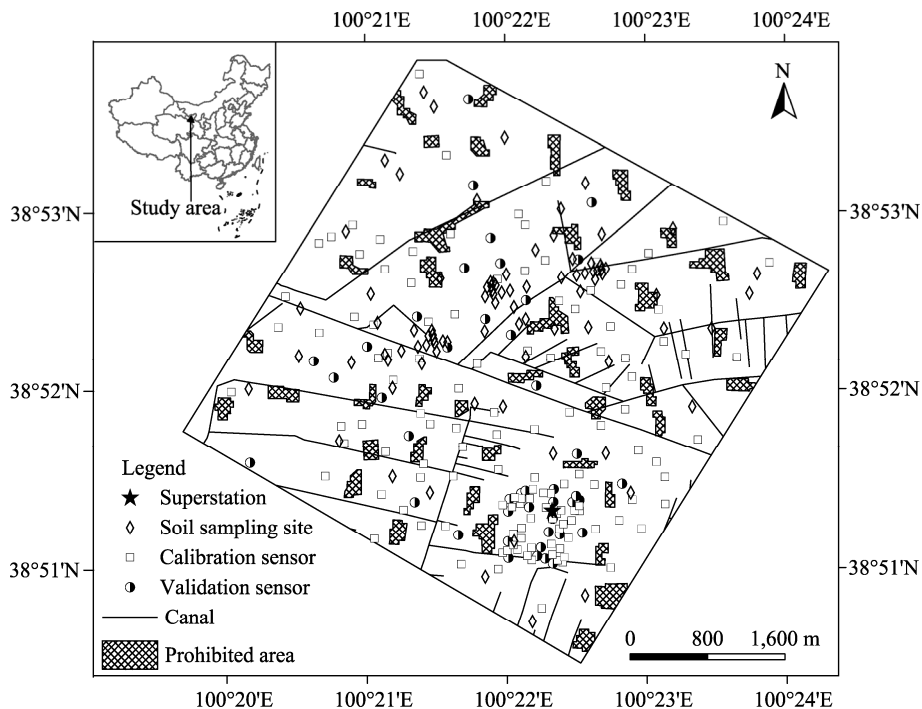

Fig. 1 Geographical locations of soil sampling sites and soil moisture sensors in the study area. The prohibited area is covered by villages.

Irrigation water in the study area is mainly supplied from the middle reaches of Heihe River. With many canals, the well-developed irrigation systems were formed. The traditional water supply systems based on canals are managed by many smallholder farmers and the agriculture is mainly depended on flooding irrigation, which often leads to serious water waste. During the growing season, most of the crop field is irrigated four times (Zhu et al., 2015), while only a small unit with areas of about 600–2,000 m^2 is irrigated once. The soil properties (bulk density, texture and organic content) vary in the study area, and soil samples are considered as silt loam with sand (9%–36%), silt (56%–81%) and clay (5%–19%). Based on regular and nested sampling, we conducted a soil field survey in July 2012. Eighty-nine soil samples in the topsoil layer of 0–20 cm were taken (Fig. 1). Soil samples were analyzed for soil organic matter content, sand content, silt content, bulk density and soil porosity.

1.2 Data

Environmental variables can be classified into static or physical and dynamic variables. The dynamic variables vary spatially and temporally, while the static variables do not change temporally.

1.2.1 Soil moisture data

Jin et al. (2014) designed an ecohydrological wireless sensor network (WSN) to model the variations of soil moisture both in space and time. Without any spatial autocorrelation assumption, a hybrid sampling method was proposed to optimize the deployment of WSN by Kang et al. (2014). In this study, three WSNs, namely WATERNET, SoilNet and Beijing Normal University Sensor Network (BNUNET) were implemented. These sensor networks are an important part of the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) (Li et al., 2013b; Gao et al., 2014). In this study, a total of 172 sensor nodes were installed primarily for observing soil temperature and moisture to the depth of 4 cm, and these nodes were randomly divided into calibration (n=129; from 1 June to 31 July, 2012) and validation (n=43; from 1 August to 30 September, 2012) datasets (Fig. 1). Detailed design of sensor network and the data communication implementation were referred to Jin et al. (2014) and Kang et al. (2014).

Soil moisture data were expressed as volumetric soil water content (%; v/v), and the meteorological data were collected from June to September in 2012. The magnitude of instrument errors for each type of WSN ranged from 1% to 2% measured by laboratory calibration using the gravimetric method (Wang et al., 2014). Soil moisture observations were conducted at 10-min intervals, but it was noted that the prediction procedure was applied to the daily mean SMC data. The calibration experiments indicated that the consistency between sensors is greater than 95% and the accuracy of the surface radiative temperature is 0.15°C.

1.2.2 Dynamic variables

The dynamic environmental variables included evapotranspiration, leaf area index (LAI), meteorological information (precipitation, global radiation, air pressure and humidity), land surface temperature (LST), neighborhood states and irrigation time, which were collected from 1 June to 30 September, 2012. Daily meteorological data were recorded by a superstation located within the corn field (Li et al., 2013b; Fig. 1). Processed meteorological measurements were employed in all variables because of the flat terrain and the close proximity of wireless sensors towards the superstation.

The collection of daily image data and uncertainty of the LST retrievals prohibited the use of LST data. Since the temporal dynamics of LST can be captured by the WSN (Jin et al., 2014), the daily spatial distribution of LST was interpolated by the ordinary kriging approach (from 1 June to 30 September, 2012). The monthly LAI data were retrieved from the Huanjing satellite charge coupled device (HJ-CCD) images at a 30-m spatial resolution (Liao et al., 2013). In this study, the HJ-CCD data from the following four days were employed: 19 June, 8 July, 14 August and 6 September, 2012. The daily actual evapotranspiration data estimated by ETWatch algorithm (Wu et al., 2012) with a 30-m spatial resolution from May to September 2012 were obtained from the Heihe Plan Science Data Center (http://www.heihedata.org). In addition, the missing data were

infilled by the ETMonitor dataset with a 250-m spatial resolution (Hu and Jia, 2015).

Irrigation data play irreplaceable roles in estimating the surface soil moisture in irrigated fields (Bertoldi et al., 2014). In the study area, the irrigated cropland mainly depends on the flooding irrigation. Hence, the soil moisture will abruptly increase to saturation and might last for several hours. The exact irrigation data (including irrigation time and water amount) were unavailable due to the fragmental ownership of crop fields. Thus, an alternative retrieval of irrigation data from real-time observed SMC was carried out based on two principals: (1) there was no precipitation within few hours; and (2) the SMC increased sharply by at least 5%. The new irrigation data were a Boolean variable, in which value 1 indicates an irrigation event and value 0 denotes no irrigation. Furthermore, the daily distribution of irrigated values will be achieved, as one point can represent the status of an independent irrigated field.

The states of the neighboring cells were difficultly measured in situ. We used the von Neumann two-dimensional neighborhood to identify the neighborhood states of each cell, aiming to account for the subsurface lateral flow and the surface runoff (Brocca et al., 2007). The daily SMC maps were interpolated by using the ordinary kriging approach. These predicted SMC maps were applied to derive the neighborhood states, which were based on the observed soil moisture data during the MCA model training. Nevertheless, at the model validation stage, this variable will be calculated by using the simulated SMC maps based on MCA model at a previous time. The amount of neighboring cell which was larger than that of central cell was recorded and saved at the corresponding location of a new raster. A daily variable was generated after the moving three-by-three window across the SMC grid. The auxiliary layer (statistics of neighborhoods), even though with certain uncertainty concerning the measured values, could offer additional information of lateral water movement for the SMC modeling.

1.2.3 Static variables

The land cover map produced from orthophotos at a 10-m spatial resolution was taken into account in the predicting process of SMC, which was acquired from the Heihe Plan Science Data Center. Digital elevation model (DEM) data were derived from airborne light detection and ranging (LiDAR, pixel size of 2 m) and resampled to a 10-m spatial resolution. Topographic properties, elevation, slope aspect, slope gradient and SAGA TWI (System of Automated Geoscientific Analyses Topographic Wetness Index), were derived in SAGA GIS (http://saga-gis.org/). Euclidean distances in the entire area from the canals were calculated and saved in separated TIFF files. Multivariate geostatistical approach, i.e. kriging with an external drift, was utilized for estimating the thematic soil maps at a 10-m spatial resolution, following by a principal component analysis.

All the aforementioned variables were converted to a TIFF format, and resampled to a 10-m resolution. Variables were selected by stepwise regression in both directions (forward and backward), using Akaike information criterion (Akaike, 1974). Then the selected variables were incorporated in MLP-MCA and DBN-MCA models.

2 Methods

2.1 MCA framework and procedure

There are few states controlled by simplified transition function in the classical CA, which cannot deal with macroscopic phenomena. Using complicated transition rules, the MCA is developed so as to represent plenty of states. A formal definition of MCA can be found in Mendicino et al. (2015), and the basic structure can be characterized as follows:

$$\text{MCA}=(E^d, X, P, Q, \sigma, \gamma). \tag{1}$$

Where, E^d denotes cells set, X the neighborhood set influencing the cell states, P the global constant parameters for each cell, Q the finite set of cell states given by the Cartesian product with several sub-states ($Q=Q_1 \times Q_2 \times \ldots \times Q_n$), σ the deterministic transition function of the MCA, and γ the finite set of functions defining the external covariates. The values of sub-states will be changed in time either based on the interactions among sub-states inside the cells or local

interactions among neighboring cells.

There are three phases in the structure of MCA model. The first phase is relative to the data preparation of input data including calibration data, validation data and standardized variables to ensure that all types of inputs are at the same scale. The second phase involves the transition functions of cell states based on the calibration data. Various data mining techniques available were proposed to extract the transition functions. In this study, DBN and MLP were mobilized to serve as the weight optimization tool for the learning capacity of the model. In the last phase, the generalization capacity of the model will be verified using the validation data.

2.2 MLP network

Artificial neural networks (ANNs) can simulate human learning processes through a flexible mathematical structure. Among numerous ANN algorithms, MLP network with the back-propagation training algorithm is perhaps the most popular model. It was selected for simulating the dynamics of SMC in this study. There are three layers in the MLP network, including input, hidden and output layers, each with a set of interconnected neurons working in parallel to transform the input data into output values (Fig. 2). In the input layer, the neurons have equal amount to the variables; whereas neurons in the hidden layer, which carry weights representing the linkages between the variables and SMC, are determined based on both the calibration and validation data. The output layer only comprises a single neuron that represents the output SMC values, which can be expressed as:

$$y_i = \sum_{j=1}^{n} w_{ij} x_j + b_i, \quad (2); \qquad Y=F(y_i). \quad (3)$$

Where, y_i is the sum of weighted nodes, w_{ij} the weight of connection between nodes i and j, x_j the value of node j, b_i the bias, Y the node output and F the transfer function.
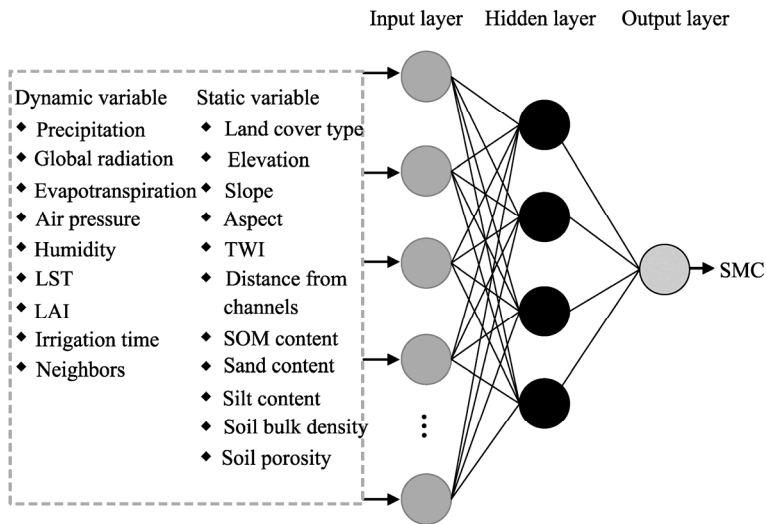

Fig. 2 Architecture of the MLP (multi-layer perceptron) neural network. LST, land surface temperature; LAI, leaf area index; TWI, topographic wetness index; SOM, soil organic matter; SMC, soil moisture content.

2.3 Deep belief network

Deep learning (also known as deep belief networks (DBNs)), proposed by Hinton (Hinton et al., 2006), is a new branch of the machine learning. DBN comprises a generative probabilistic architecture including several restricted Boltzmann machine (RBM) layers (Hinton and Salakhutdinov, 2006).

2.3.1 RBM

As a special type of Markov random field, RBM is a bipartite undirected graphical model and consists of a layer of visible units (random variables) $\mathbf{v} \in \{0, 1\}^V$, and a layer of binary hidden

units (random variables) $\mathbf{h} \in \{0, 1\}^H$. The visible units represent the observed data and the hidden units learn to capture the higher-order features based on the input. Visible and hidden layers are connected by a symmetrical weight matrix $W \in R^{V \times H}$, whereas there are no connections within a layer (Fig. 3). The joint distribution over the visible and hidden units is as follows:

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h})), \tag{4}$$

$$Z = \sum_v \sum_h \exp(-E(\mathbf{v}, \mathbf{h})). \tag{5}$$

Where, Z is a normalization constant, and $E(\mathbf{v}, \mathbf{h})$ the energy of the state $(\mathbf{v}, \mathbf{h})$. The energy of an observed state $(\mathbf{v}, \mathbf{h})$ can be obtained as Eq. 6.

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{V} \sum_{j=1}^{H} W_{ij} v_i h_j - \sum_{j=1}^{H} b_j h_j - \sum_{i=1}^{V} c_i v_i. \tag{6}$$

Where, v_i and h_j are binary states of the visible unit i and the hidden unit j, b_j and c_i are hidden and visible unit biases, respectively. As shown in Eq. 6, units of v and h are independent of each other as there are no direct connections between units in the same layer (Fig. 3). That is, there are no visible-visible or hidden-hidden connections. Conditioned on the visible layer, the hidden units should be independent Bernoulli random variables if the values of hidden units are binary. In addition, the visible units are conditioned on the hidden layer and are independent Gaussian random variables if the visible units own real values.
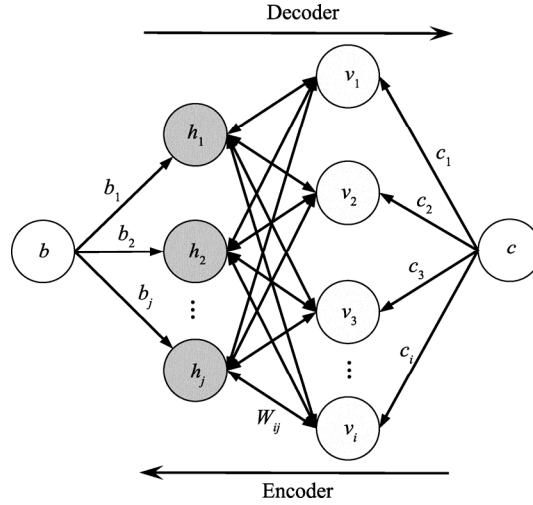

Fig. 3 Architecture of a restricted Boltzmann machine (RBM). v_i and h_j are binary states of the visible unit i and the hidden unit j, respectively; b_j and c_i are hidden and visible unit biases, respectively; w_{ij} is the weight of connection between nodes i and j.

2.3.2 DBN architecture and training

Feature activations can be implemented through RBMs and hence adopted to effectively learn within DBN architecture (Fig. 4). The hidden layer will be visible to the next RBM if the previous RBM was trained. The extent of the data's negative log probability will be enhanced by adding more layers and perceiving pattern representations (Hinton et al., 2006). If the stochastic gradient ascent is carried out for the log-likelihood manifold, a new learning function of a weight w_{ij} can be achieved as follows:

$$\Delta W_{ij} = \varepsilon \frac{\partial \log p(\mathbf{v})}{\partial W_{ij}} = \varepsilon \left(\langle v_i h_j \rangle p^0 - \langle v_i h_j \rangle p_\theta^\infty \right). \tag{7}$$

Where, ε is a learning rate, p^0 the distribution of the data and p_θ^∞ the equilibrium distribution defined by RBM. $\theta = \{W_{ij}, b_j, c_i\}$ includes all parameters needed to learn. The modeling accuracy

of DBN with back-propagation layer totally depends on the output feature of the last DBN layer. A DBN with 256-512-512-256-15 structure was trained in this study by using the training set to exhibit the feature extracting ability.
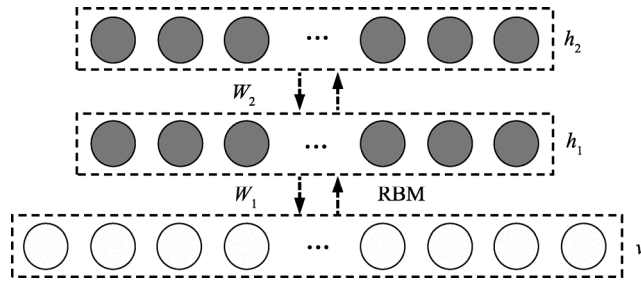

Fig. 4 Deep belief network (DBN) with two hidden layers (h_1 and h_2), and the RBM that is being trained for layer v. If the layer v is trained, the RBM will be trained for h_1 and h_2 in turn. W is the weight vector between visible unit and hidden unit.

2.4 Evaluation and uncertainty assessment of models

The estimation accuracy of models was assessed by the root mean square error (RMSE) according to Eq. 8:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^* - x_i\right)^2} \ . \tag{8}$$

Where, x_i^* is the predicted SMC expressed as volumetric soil water content (%) at location i, x_i the observed SMC at the same location and n the number of instances presented to the model.

The uncertainty of predicted SMC was evaluated by the sequential Gaussian simulation (sGs) (Goovaerts, 1997; Delbari et al., 2009). Local uncertainty of each cell is defined as the probability of predicted SMC value less than a specific SMC threshold. In this study, the specific SMC threshold was set to 25%. The spatial uncertainty is defined as the joint uncertainty concerning multi-locations in the prediction, so that locations with similar uncertainty can be represented through a specific probability (Zhao et al., 2005). If a specific probability (p) is defined, the multi-locations (or an area) can be described as Prob[SMC<25%]≥p. Hence, the joint uncertainty can be calculated by n/N, where N denotes the total number of realizations generated and n the number of realizations, of which all the predicted SMC values at the same multi-locations are with values less than 25% in each realization.

Data preparations, analyses and visualization were carried out using ArcGIS 10.0 (http://www.esri.com/) and software R version 3.1.1 (http://cran.r-project.org/) with its add-in packages: "h2o", "RSNNS", "gstat" and "raster".

3 Results

3.1 Spatio-temporal variations of observed soil moisture

Figure 5 summarizes the main statistics of observed SMC on each day during the study period (from 1 June to 30 September, 2012), suggesting considerable spatial heterogeneity in soil moisture dynamics. There were several irrigation events which can be illustrated by the peak of mean SMC. Variation of observed soil moisture was shown by the scatter plots in terms of the standard deviation versus mean soil moisture for each day (Fig. 6). Along with the increase of mean soil moisture, there was also an increase in the standard deviation. Even if all the observations in this study were conducted under dry conditions, the present result concurred with the previous studies (Ryu and Famiglietti, 2005; Sur et al., 2013). In addition, this study suggested that more soil samples need to be required for arid region in the prediction of mean soil moisture, which was also high-lighted in other related studies (Brocca et al., 2007, 2010; Wang et al., 2008).
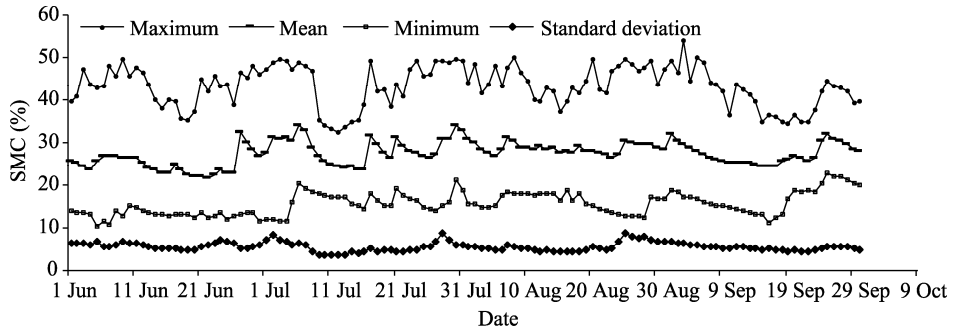
Fig. 5 Statistics of soil moisture content (SMC) at the depth of 4 cm from 1 June to 30 September in 2012 observed by a WSN (wireless sensor network). SMC is expressed in terms of volumetric soil water content.
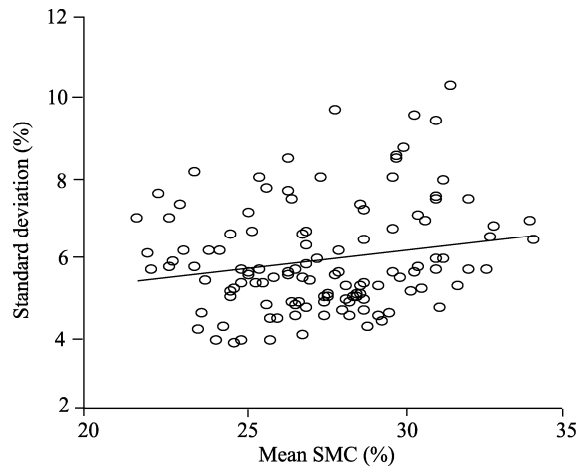

Fig. 6 Relationship between mean SMC and standard deviation from 1 June to 30 September in 2012

3.2 Evaluation of variable importance

The increases in RMSEs as the dynamic and static variables excluded one by one are shown in Fig. 7. It was noted that the RMSEs were generated by calibration data from 1 June to 31 July. The importance of variables was judged by the increase of RMSEs after excluding each variable from the model, as both MLP and DBN algorithms cannot reveal the functional correlations between SMC and variables. Besides, the magnitude of increased RMSE suggested that the soil bulk density plays the most important role in explaining the spatial variations of SMC in MLP-MCA model. Broadly speaking, the contributions of different variables to the DBN-MCA model were more or less the same. That was, their exclusion marginally decreased the prediction error in terms of the RMSEs. This finding differed from previous study of SMC mapping based on soil-landscape concept, in which soil porosity might be significantly correlated with hydraulic conductivity under wet conditions (Famiglietti et al., 1998). Most of the previous studies on neural network showed their low interpretability due to the black-box implementation depending on the input and output data (Wang and Li, 2011). As the spatial variation of SMC cannot be simulated by the standard geostatistical methods (Western et al., 2001), the learning model provides an alternative technique for predicting SMC.

3.3 Performance of the spatial models

Generally speaking, the DBN-MCA model yielded the minimum prediction errors and generated the reasonable results compared to the MLP-MCA model. The RMSE for the DBN-MCA model (5.54%) was lower than that for the MLP-MCA model (6.77%; Fig. 8), indicating a modest relative improvement of prediction accuracy by the DBN-MCA. Furthermore, the standard deviations of estimated SMC from all models (3.15% and 6.76% for DBN-MCA and MLP-MCA

models, respectively; Table 1) were lower than that of observed SMC (6.18%; Fig. 5), which suggested that the MCA model generated better predictions than what was expected basing on the observed values alone.

Table 1   Descriptive statistics of the estimated SMC basing on MLP-MCA and DBN-MCA models

Date	Estimated SMC from MLP-MCA model (%)				Estimated SMC from DBN-MCA model (%)			
	Maximum	Mean	Minimum	SD	Maximum	Mean	Minimum	SD
1 Aug	39.04	26.13	8.36	5.78	37.57	23.37	15.67	3.64
15 Aug	34.94	26.89	13.66	5.02	36.96	28.20	18.32	3.27
1 Sep	39.96	30.24	18.62	6.72	39.20	28.02	19.98	4.25
15 Sep	38.86	24.90	14.60	5.61	29.84	23.86	15.58	2.53
Study period	51.27	25.88	8.36	6.76	49.52	24.01	11.23	3.15

Note: SMC, soil moisture content; MLP, multi-layer perceptron; MCA, macroscopic cellular automata; DBN, deep belief network; SD, standard deviation.
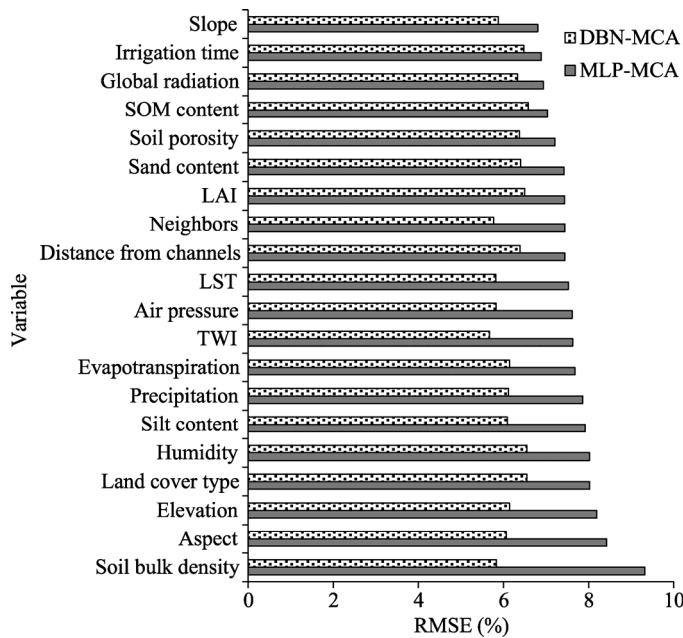

Fig. 7   Variable importance shown by increase in the RMSEs (root mean square errors) of the MLP (multi-layer perceptron)-MCA (macroscopic cellular automata) and DBN (deep belief network)-MCA models in which one variable was excluded once
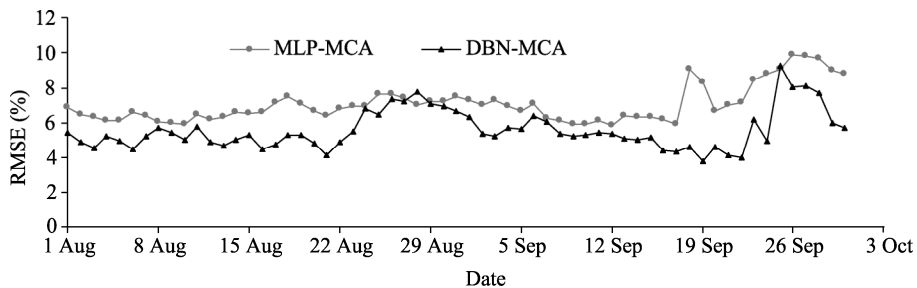

Fig. 8   Daily prediction errors of the SMC derived from the MLP-MCA and DBN-MCA models from 1 August to 30 September, 2012. RMSE, root mean square error.

The predicted SMC values of this study might be influenced by many factors like measurement, laboratory, statistical and random errors. For example, the soil properties and LST used as variables were interpolated by kriging; the LAI data were retrieved with a mean error of 0.46 (Liao et al., 2013); and the daily actual evapotranspiration data produced the best performance

with determination of 0.87 (Hu and Jia, 2015). Thus, those associated interpolation errors were propagated to the subsequent SMC estimations. The more variables we utilized from different sources, the more uncertainty the predicted model might be involved.

3.4 Spatial prediction and mapping of SMC

The spatial distribution of SMC on 1 August, 15 August, 1 September and 15 September simulated by MLP-MCA and DBN-MCA models with a 10-m spatial resolution is shown in Fig. 9, in which soil moisture of each point was assumed to be homogeneous. The maps showed similar macro-patterns on different days from a qualitative perspective, where low SMC values appeared in the northeastern part and high values in the southwest. Besides, the SMC at the depth of 4 cm, estimated by DBN-MCA model, ranged from 11.23% to 49.52% with a mean value of 24.01%. Compared with DBN-MCA, the MLP-MCA model estimated a higher average SMC (25.88%; Table 1). All SMC maps derived from DBN-MCA model showed more details than those from MLP-MCA model. The detailed maps derived from DBN-MCA model suggested that this approach provided reasonable and good results compared to the MLP-MCA model. It was interesting to find that the SMC maps on 15 August were obviously different between the two models, and the wetness in SMC map based on MLP-MCA model appeared to be more homogeneous. The converse results suggested that there was certain uncertainty within the predicting model, and the uncertainty should be discussed further. Furthermore, it was easy to observe an obvious overlap between irrigated canals (Fig. 1) and spatial variation of soil moisture (Fig. 9). It was also suggested that the lateral redistribution of soil water led to a significant influence on SMC, which was implied by the connected architecture (Western et al., 1998).
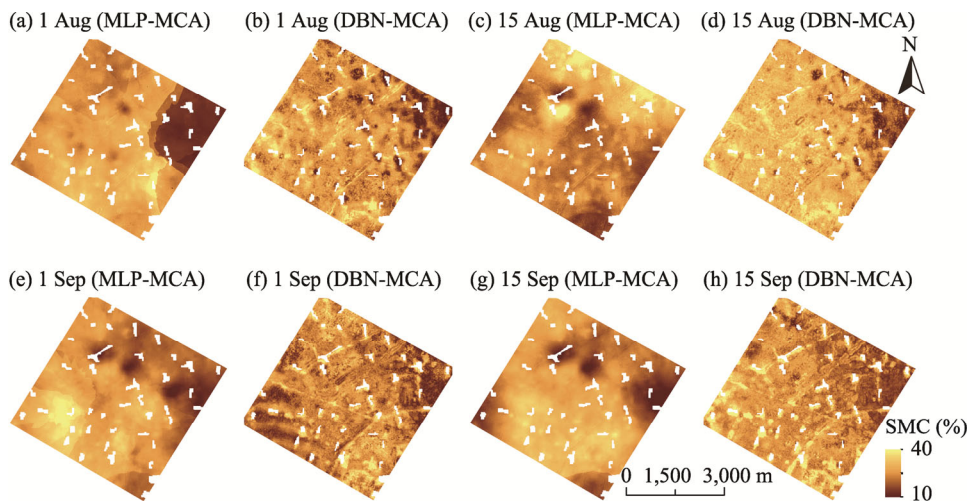

Fig. 9 Spatial distribution maps of SMC simulated by MLP-MCA (a, c, e, g) and DBN-MCA (b, d, f, h) models on 1 August (a, b), 15 August (c, d), 1 September (e, f) and 15 September (g, h). Note that masked areas are colored white in each map.

3.5 Uncertainty evaluation of SMC

The simulation based on the sGs was executed for 100 times with the same probability of SMC prediction at a 10-m spatial resolution, and hence one hundred SMC maps were produced. Those SMC maps were post processed for the calculation of probability map and the joint probability. The distribution of probability of SMC less than 25% on certain days is shown in Fig. 10. Broadly speaking, there will be a high degree of uncertainty for the cell with a moderate range of probability, because the simulated SMC values significantly changed (larger or less than 25%) during the multiple realizations (e.g. probabilities between 40%–60% in Fig. 10). The SMC less than 25% was mainly found in the northern part of the study area on all days except 15 August. Probabilities of SMC on 15 August were less than 50% for 97% of the pixels, illustrating a low

uncertainty evaluation of SMC.

By linking Figs. 9 and 10 together, it was clearly found that there were similar distributions of both the predicted SMC and probabilities on 1 and 15 September. It may also imply that the prediction of SMC based on MLP-MCA and DBN-MCA models can be validated partially in the upper part of the study area with a relatively lower uncertainty (Fig. 10). We utilized the spatial uncertainty (joint probability) of SMC to evaluate the stability of multi-location uncertainty measures, where SMC threshold was 25% and two given critical probabilities were 60% and 80%. All the joint probabilities of those four days were zero, which also illustrates the high spatial uncertainty.
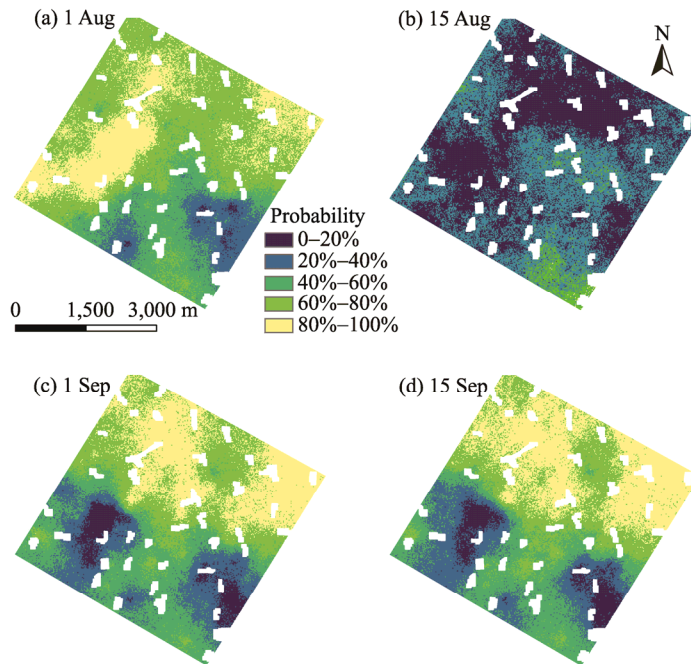

Fig. 10   Distribution maps of probability of SMC less than 25% on 1 August (a), 15 August (b), 1 September (c) and 15 September (d)

4   Discussion

The validation results of this study demonstrated that the proposed deep learning-based MCA model can accurately estimate the spatio-temporal patterns of SMC by combining field measurements and environmental variables together. Different from relative studies about soil hydrological model (Schwärzel et al., 2009) and numerical simulations (Bertoldi et al., 2014), the presented MCA model in this study provided a new management opportunity for controlling SMC changes in the field irrigation. The knowledge between soil moisture and environmental variables enhanced the estimation quality. The validation results also showed that the proposed models were comparable to the sensors-based techniques in the range of visible and near/shortwave/thermal infrared (Sánchez-Ruiz et al., 2014) and ground penetrating radar (Ardekani, 2013) regions. As many dynamic variables were necessary for the input, the proposed models involved a fair complex pre-process and was less intuitive than those regressions obtained between variables and SMC (Paloscia et al., 2013). However, MCA models are not only insensitive to missing data, but also can be available over vegetated terrains. From a practical point of view, this model is promising for the real-time soil moisture simulation based on WSN.

Due to the fragmental ownership of fields, an alternative retrieval of irrigation data was achieved through the artificial WSN, as one point represented the status of an independent

irrigated field. The pre-processed Boolean variable of irrigation data played an important role in the prediction in terms of validation at the plot scale (data not shown), even if the variable importance was not significant (Fig. 7). Bertoldi et al. (2014) also revealed that the irrigation data could significantly affect the prediction of SMC. Through the case studies reported in this paper, it was clear that the spatial variation and magnitude of soil moisture depended on the function of irrigation amount and simulated models in question.

Furthermore, the predicted SMC maps could also be compared with other maps constructed by other learning models (e.g. multi-layer perceptron). By comparison, it was observed that the deep learning was capable of modeling the SMC variation in spite of the generally qualitative and quantitative nature of hydrological data as noisy and highly variable. The achieved findings were consistent with other studies that the deep learning is a powerful modeling technique to learn the non-linear and complex behaviors of datasets (Schmidhuber, 2015).

In this study, the MCA models generated better predictions than what were expected, because the standard deviations of the predicted SMC values from all MCA models were lower than those of the observed values. However, there were several uncertainties that should be pointed out and paid much more attention in the further application of quasi real-time soil moisture monitoring as the following reasons. (1) Soil moisture data were collected from a hybrid WSN, in which different types of sensors may be resulting in various calibration errors. (2) There were certain uncertainties within the variables, e.g. soil properties and neighborhood states predicted by geostatistics. (3) Intraday dynamic variables were substituted by historical data, since the daily remote sensing data were hard to be collected.

The observed SMC data in this study presented an almost normal frequency, and the autocorrelation analysis showed that SMC data were well spatially correlated. Whereas the predicted maps on four days did not derive a smoothing effect (Fig. 9). The 100-times realization of sGs revealed that there might be several uncertainties of SMC prediction, which was similar to the conclusion of Delbari et al. (2009). Moreover, the knowledge of spatial uncertainty might be helpful to optimize the designing of WSN with respect to the representative sampling (Kang et al., 2014). Concerns about the lack of prior information of soil moisture may also arise. Although the simulation based on the sGs approach was working in hindsight, a more representative WSN regarding hierarchical SMC variation might be achieved at several scales, which took up a little time and energy for the field sampling to avoid the frequent uncertainty within some local areas.

5 Conclusions

In this study, a novel DBN-MCA model was presented and thus applied to a prediction of SMC over an irrigated corn field in the Zhangye oasis. DBN was trained to extract the transition functions for the simulation of the cell state changes. After the incorporation of static and dynamic variables, DBN-MCA model led to a reduction in RMSE by 18% in comparison with the MLP-MCA model. Cross validation results showed that MCA model is a promising technique for modeling the temporal and spatial variations of SMC. Furthermore, it was also found that the predicted biases were probably due to the difficulties in collecting the irrigation amount. If the detailed environmental variables could be employed, the MCA model would achieve a more accurate simulation.

Acknowledgements

References

Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6): 716–723.

Ardekani M R M. 2013. Off- and on-ground GPR techniques for field-scale soil moisture mapping. Geoderma, 200–201: 55–66.

Avolio M V, Di Gregorio S, Lupiano V, et al. 2013. SCIDDICA-SS$_3$: a new version of cellular automata model for simulating fast moving landslides. Journal of Supercomputing, 65(2): 682–696.

Bertoldi G, Della Chiesa S D, Notarnicola C, et al. 2014. Estimation of soil moisture patterns in mountain grasslands by means of SAR RADARSAT2 images and hydrological modeling. Journal of Hydrology, 516: 245–257.

Brocca L, Morbidelli R, Melone F, et al. 2007. Soil moisture spatial variability in experimental areas of central Italy. Journal of Hydrology, 333(2–4): 356–373.

Brocca L, Melone F, Moramarco T, et al. 2010. Spatial-temporal variability of soil moisture and its estimation across scales. Water Resources Research, 46(2): W02516, doi: 10.1029/2009WR008016.

Cervarolo G, Mendicino G, Senatore A. 2010. A coupled ecohydrological-three-dimensional unsaturated flow model describing energy, H_2O and CO_2 fluxes. Ecohydrology, 3(2): 205–225.

D'Ambrosio D, Iovine G, Spataro W, et al. 2007. A macroscopic collisional model for debris-flows simulation. Environmental Modelling & Software, 22(10): 1417–1436.

Delbari M, Afrasiab P, Loiskandl W. 2009. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. CATENA, 79(2): 163–169.

D'Odorico P, Rodríguez-Iturbe I. 2000. Space-time self-organization of mesoscale rainfall and soil moisture. Advances in Water Resources, 23(4): 349–357.

Famiglietti J S, Wood E F. 1994. Multiscale modeling of spatially variable water and energy balance processes. Water Resources Research, 30(11): 3061–3078.

Famiglietti J S, Rudnicki J W, Rodell M. 1998. Variability in surface soil moisture content along a hillslope transect: Rattlesnake Hill, Texas. Journal of Hydrology, 210(1–4): 259–281.

Furman A. 2008. Modeling coupled surface-subsurface flow processes: A review. Vadose Zone Journal, 7(2): 741–756.

Gao S G, Zhu Z L, Liu S M, et al. 2014. Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing. International Journal of Applied Earth Observation and Geoinformation, 32: 54–66.

Goovaerts P. 1997. Geostatistics for Natural Resources Evaluation. New York: Oxford University Press.

Hinton G E, Osindero S, Teh Y W. 2006. A fast learning algorithm for deep belief nets. Neural Computation, 18(7): 1527–1554.

Hinton G E, Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. Science, 313(5786): 504–507.

Hu G C, Jia L. 2015. Monitoring of evapotranspiration in a semi-arid inland river basin by combining microwave and optical remote sensing observations. Remote Sensing, 7(3): 3056–3087.

Ivanov V Y, Bras R L, Vivoni E R. 2008. Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks. Water Resources Research, 44(3): W03429, doi: 10.1029/2006WR005588.

Ji N N, Zhang J S, Zhang C X. 2014. A sparse-response deep belief network based on rate distortion theory. Pattern Recognition, 47(9): 3179–3191.

Jin R, Li X, Yan B P, et al. 2014. A nested ecohydrological wireless sensor network for capturing the surface heterogeneity in the midstream areas of the Heihe River Basin, China. IEEE Geoscience and Remote Sensing Letters, 11(11): 2015–2019.

Kang J, Li X, Jin R, et al. 2014. Hybrid optimal design of the eco-hydrological wireless sensor network in the middle reach of the Heihe River Basin, China. Sensors, 14(10): 19095–19114.

Kim K J, Cho S B. 2006. Evolved neural networks based on cellular automata for sensory-motor controller. Neurocomputing, 69(16–18): 2193–2207.

Kim Y, Evans R G, Iversen W M. 2008. Remote sensing and control of an irrigation system using a distributed wireless sensor network. IEEE Transactions on Instrumentation and Measurement, 57(7): 1379–1387.

Li H, Li X Y, Ramanathan M, et al. 2014. Identifying informative risk factors and predicting bone disease progression via deep belief networks. Methods, 69(3): 257–265.

Li X, Lao C H, Liu Y L, et al. 2013a. Early warning of illegal development for protected areas by integrating cellular automata with neural networks. Journal of Environmental Management, 130: 106–116.

Li X, Cheng G D, Liu S M, et al. 2013b. Heihe watershed allied telemetry experimental research (HiWATER): Scientific

objectives and experimental design. Bulletin of American Meteorological Society, 94(8): 1145–1160.

Liao Y R, Fan W J, Xu X R. 2013. Algorithm of leaf area index product for HJ-CCD over Heihe River Basin. In: Proceedings of 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Melbourne, VIC: IEEE, 169–172.

Mendicino G, Senatore A, Spezzano G, et al. 2006. Three-dimensional unsaturated flow modeling using cellular automata. Water Resources Research, 42(11): W11419, doi: 10.1029/2005WR004472.

Mendicino G, Pedace J, Senatore A. 2013. Cellular Automata based modeling for the assessment of ecohydrological dynamics at the hillslope scale: Preliminary results. Procedia Environmental Sciences, 19: 311–320.

Mendicino G, Pedace J, Senatore A. 2015. Stability of an overland flow scheme in the framework of a fully coupled eco-hydrological model based on the Macroscopic Cellular Automata approach. Communications in Nonlinear Science and Numerical Simulation, 21(1–3): 128–146.

Oku M, Aihara K. 2010. Noise-robust realization of Turing-complete cellular automata by using neural networks with pattern representation. Physics Letters A, 374(48): 4859–4863.

Paloscia S, Pettinato S, Santi E, et al. 2013. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. Remote Sensing of Environment, 134: 234–248

Ryu D, Famiglietti J S. 2005. Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment. Water Resources Research, 41(12): W12433, doi: 10.1029/2004WR003835.

Sadler E J, Evans R G, Stone K C, et al. 2005. Opportunities for conservation with precision irrigation. Journal of Soil and Water Conservation, 60(6): 371–379.

Sánchez-Ruiz S, Piles M, Sánchez N, et al. 2014. Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates. Journal of Hydrology, 516: 273–283.

Schmidhuber J. 2015. Deep learning in neural networks: An overview. Neural Networks, 61: 85–117.

Schwärzel K, Feger K H, Häntzschel J, et al. 2009. A novel approach in model-based mapping of soil water conditions at forest sites. Forest Ecology and Management, 258(10): 2163–2174.

Stafford J V. 1996. Essential technology for precision agriculture. In: Proceedings of the 3rd International Conference on Precision Agriculture. Minneapolis, Minnesota, 595–604.

Sulis M, Meyerhoff S B, Paniconi C, et al. 2010. A comparison of two physics-based numerical models for simulating surface water-groundwater interactions. Advances in Water Resources, 33(4): 456–467.

Sur C, Jung Y, Choi M. 2013. Temporal stability and variability of field scale soil moisture on mountainous hillslopes in Northeast Asia. Geoderma, 207–208: 234–243.

Tague C L, Band L E. 2004. RHESSys: Regional Hydro-Ecologic Simulation System: An object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. Earth Interactions, 8(19): 1–42.

von Neumann J. 1966. Theory of Self-Reproducing Automata. Urbana: University of Illinois Press.

Wang C M, Zuo Q, Zhang R D. 2008. Estimating the necessary sampling size of surface soil moisture at different scales using a random combination method. Journal of Hydrology, 352(3–4): 309–321.

Wang J H, Ge Y, Song Y Z, et al. 2014. A geostatistical approach to upscale soil moisture with unequal precision observations. IEEE Geoscience and Remote Sensing Letters, 11(12): 2125–2129.

Wang Y, Li S C. 2011. Simulating multiple class urban land-use/cover changes by RBFN-based CA model. Computers & Geosciences, 37(2): 111–121.

Western A W, Blöschl G, Grayson R B. 1998. How well do indicator variograms capture the spatial connectivity of soil moisture? Hydrological Processes, 12(12): 1851–1868.

Western A W, Blöschl G, Grayson R B. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. Water Resources Research, 37(1): 83–97.

Wu B F, Yan N N, Xiong J, et al. 2012. Validation of ETWatch using field measurements at diverse landscapes: A case study in Hai Basin of China. Journal of Hydrology, 436–437: 67–80.

Wu T, Kareem A. 2011. Modeling hysteretic nonlinear behavior of bridge aerodynamics via cellular automata nested neural network. Journal of Wind Engineering and Industrial Aerodynamics, 99(4): 378–388.

Zhao Y C, Shi X Z, Yu D S, et al. 2005. Uncertainty assessment of spatial patterns of soil organic carbon density using sequential indicator simulation, a case study of Hebei province, China. Chemosphere, 59(11): 1527–1535.

Zhu Z L, Tan L, Gao S G, et al. 2015. Observation on soil moisture of irrigation cropland by cosmic-ray probe. IEEE Geoscience and Remote Sensing Letters, 12(3): 472–476.