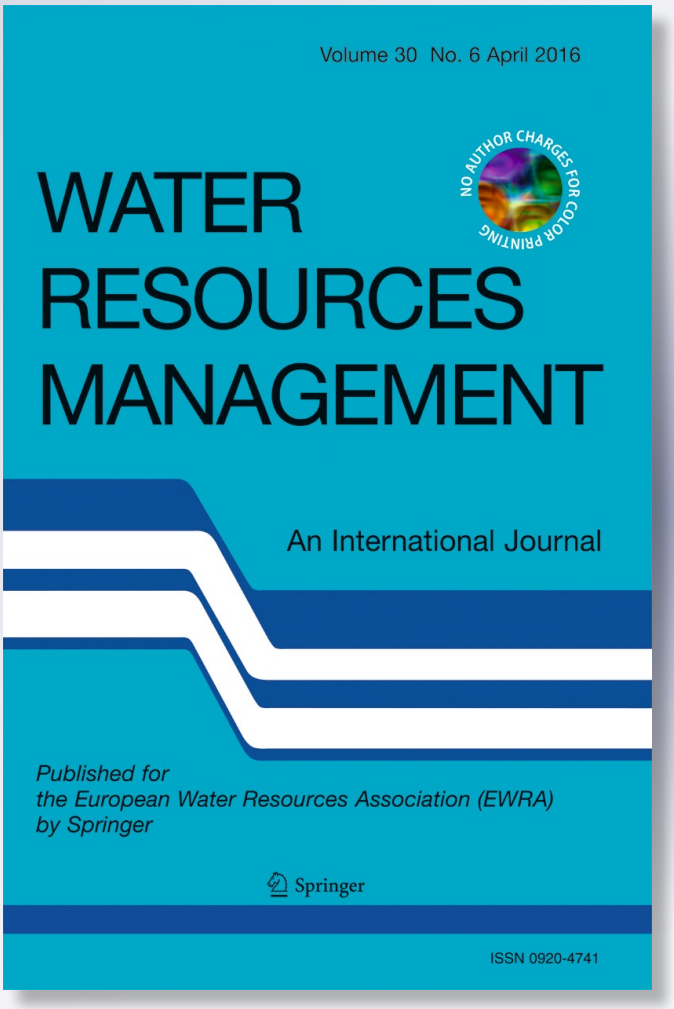
Statistical Modelling of Vertical Soil Moisture Profile: Coupling of Memory and Forcing

Manali Pal, Rajib Maity & Sayan Dey

Water Resources Management
An International Journal - Published for the European Water Resources Association (EWRA)

ISSN 0920-4741
Volume 30
Number 6

Water Resour Manage (2016)
30:1973-1986
DOI 10.1007/s11269-016-1263-4



Water Resour Manage (2016) 30:1973–1986
DOI 10.1007/s11269-016-1263-4

Statistical Modelling of Vertical Soil Moisture Profile: Coupling of Memory and Forcing

Manali Pal[1] · **Rajib Maity**[2] · **Sayan Dey**[2]

Received: 4 August 2015 / Accepted: 4 February 2016 /
Published online: 15 February 2016

Abstract Information of Soil Moisture Content (SMC) at different depths i.e. vertical Soil Moisture (SM) profile is important as it influences several hydrological processes. In the era of microwave remote sensing, spatial distribution of soil moisture information can be retrieved from satellite data for large basins. However, satellite data can provide only the surface (~0–10 *cm*) soil moisture information. In this study, a methodological framework is proposed to estimate the vertical SM profile knowing the information of SMC at surface layer. The approach is developed by coupling the memory component of SMC within a layer and the forcing component from soil layer lying above by an Auto-Regressive model with an exogenous input (ARX) where forcing component is the exogenous input. The study highlights the mutual reliance between SMC at different depths at a given location assuming the ground water table is much below the study domain. The methodology is demonstrated for three depths: 25, 50 and 80 *cm* using SMC values of 10 *cm* depth. Model performance is promising for all three depths. It is further observed that forcing is predominant than memory for near surface layers than deeper layers. With increase in depth, contribution of SM memory increases and forcing dissipates. Potential of the proposed methodology shows some promise to integrate satellite estimated surface soil moisture maps to prepare a fine resolution, 3-dimensional soil moisture profile for large areas, which is kept as future scope of this study.

Keywords Soil moisture (SM) · Vertical Soil Moisture Profile · Memory · Forcing · Auto-Regressive Model with Exogenous Input (ARX)

Electronic supplementary material The online version of this article (doi:10.1007/s11269-016-1263-4) contains supplementary material, which is available to authorized users.

✉ Rajib Maity
rajib@civil.iitkgp.ernet.in; rajibmaity@gmail.com

[1] School of Water Resources, Indian Institute of Technology Kharagpur, Kharagpur, India 721302

[2] Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India 721302

1 Introduction

Soil Moisture Content (SMC) is an Essential Climatological Variable (ECV) which plays a significant role in land and atmosphere interaction (Pielke 2001). Understanding the space-time variability of SMC is important to understand the role of SMC in various hydroclimatological processes in both local and global scales and its role in parameterization of climate and hydrological model (Argyrokastritis et al. 2009). Many hydrological, meteorological and agricultural applications need information of vertical SM profile for a better understanding of environmental processes, such as flooding, soil erosion and transportation of dissolved materials etc. and for irrigation scheduling and crop yield forecasting.

Microwave remote sensing, the most prevalent method to retrieve SMC recently, is only able to provide soil moisture information upto an approximate depth of 5–10 *cm* (Kerr et al. 2010) from the surface. So there is a potential benefit of assimilating the microwave-based surface soil moisture into soil water balance models and stochastic models (Hoeben and Troch 2000).

The common practices to estimate vertical SM profile are in-situ methods and estimation through different hydrologic models. However, use of conventional in-situ methods is restricted to point measurements and are also time, labour and cost intensive (Zucco et al. 2014). Other common methods of estimating vertical SM profile are Soil-Vegetation-Atmosphere-Transfer (SVAT) modelling, land-surface modelling and physical models for unsaturated zones (Downer and Ogden 2004; Moran et al. 2004). In these models, soil moisture distribution is modelled primarily by Richard's equation (1931), which physically represents the fluxes in unsaturated zone for both local and large scale applications (Haverkamp et al. 1998). This equation is highly non-linear and its solution depends on simplifications of the initial conditions (Varado et al. 2006). Moreover, it does not always provide more accurate results than other simplified infiltration models such as Horton, Philip and Green–Ampt, which are easier to apply (Chen et al. 1994; Hsu et al. 2002). Still, applications of such simplified versions are also questionable in many modelling studies for large spatial extent (Kale and Sahoo 2011).

Retrieval of vertical SM profile from the surface soil moisture measurements is possible since surface soil moisture is coupled to root-zone soil moisture through diffusion processes (Singh 2010). Mutual relationship between soil moisture at various soil depths and the stochastic features of soil moisture dynamics can be assessed by cross-correlation method and Vector Auto Regression (VAR) method (Kim and Kim 2007; Kim 2009; Kim et al. 2011; Mahmood et al. 2012). However, relatively little work has been done to explicitly quantify the information of vertical SM profile.

The principal research question addressed in this paper is how to stochastically link the surface soil moisture to the deeper layers in order to obtain the vertical soil moisture profile. The primary objective of this study is to determine the variation of the spatial and temporal dynamics of root zone soil moisture (upto ~100 *cm*). A coupling method is implemented to assess SMC at different layers by modelling its temporal persistence (memory) and considering the input from overlying layers (forcing). The persistence of the SMC time-series is significant and it is considered as the memory of the SMC. Thus, both memory and forcing can be coupled in the proposed stochastic model in order to assess the SMC at different depths. Moreover, the proposed model is developed in such a way that it uses only the surface SMC information to estimate the vertical SM profile. Thus, it may also be able to to estimate the vertical SM profile for the locations where the SMC information at deeper layers are not available. The approach also indicates the potential to assimilate the remotely sensed surface SMC data to obtain the vertical SM profile for large extent.

2 Materials and Methods

2.1 Data Source and Study Area

The Soil Climate Analysis Network (SCAN) is a continental scale network, established by U.S. Department of Agriculture (USDA)-Natural Resources Conservation Service (NRSC)-National Water and Climate Center. Hourly soil moisture data are available in the International Soil Moisture Network (ISMN) website (Dorigo et al. 2011) (http://www.wcc.nrcs.usda.gov/scan/). The ISMN was initiated by Vienna University of Technology to aid a central data hosting facility where worldwide accessible in situ soil moisture measurements from different effective networks and validation operations are collected, synchronized, and made accessible to users.

The present study is implemented for a point location. The data from one USDA SCAN network installed within the campus of Indian Institute of Technology Kanpur (IITK)-AIR-STRIP located at 26.5114°N latitude and 80.2349°E longitude, is used. The complete set of obtained data, covers a period of June 16, 2011 to October 15, 2014. The network uses the Water Scout sensor, installed at 10, 25, 50 and 80 cm depths, to configure the vertical SM profile in volumetric terms, i.e. m^3 of moisture per m^3 of soil. The temporal resolution of the acquired data is 15 minutes. Soil moisture at daily scale is calculated by taking the average over a given day. In case of the presence of missing data within a span of day, the missing time periods are ignored and the average is taken over the time span for which the data are available. If the data for all time-steps in a day are missing, daily soil moisture for that day is computed by linear interpolation. The soil type of study area may be categorized as silty according to the unified soil classification system, up to 1.5 m depth the area consists of 13 % sand, 61 % silt and 26 % clay (Jishnu et al. 2013).

2.2 Methodology

The basic modelling approach involves exploring the mutual relationship between daily variations of SMC of the two adjacent layers by coupling two important characteristics influencing the change in SMC of the deeper layer. These factors are: i) *memory* or persistence of SMC series of underlying layer and, ii) a *forcing* due to change in the SMC of the overlying layer. The soil moisture memory is the tendency of soil to retain its initial moisture content value. Thus, the soil moisture value at a specific depth can be assumed to depend upon the soil moisture values of the previous time step(s). The forcing is considered to be the effect of SMC of the overlying layer.

The variation in soil moisture dynamics due to the effect of other factors, such as precipitation and maximum air temperature prevails mainly in the near surface layers (Mahmood et al. 2012). Similarly other factors may include relative humidity, wind speed that may affect surface soil moisture variation by controlling the evaporation and evapotranspiration. Depending on the soil characteristics the variation in near surface soil moisture may influence the SMC of deeper layers with a couple of days lag (Mahmood et al. 2012). In this study, being considered at daily scale, only the effects of memory and forcing from the overlying layer are considered. The effect of the memory is represented by the auto-regressive component of the model and the forcing is represented by the SMC of overlying layer and considered as the exogenous input. The methodology of this study is explained in following sections.

2.2.1 Pre-Processing of Raw Data

Observed daily soil moisture data is available from June 16, 2011 to October 15, 2014 at 15 minutes time interval for 10, 25, 50 and 80 *cm* depths. However, there are some missing data periods. Longer periods of missing data cannot be filled up and thus discarded. In some time steps of the complete data the surface (10 *cm* depth) SMC values are missing although the SMC values of the deeper layers are available. Such time steps are also discarded from the study since the study attempts to obtain the complete vertical SM profile using only surface soil moisture information. Observed soil moisture data for the remaining period (with shorter missing periods) are shown as a time series plot in Fig. 1 (July 01, 2011 to April 28, 2013). Such shorter periods of missing values are replaced by simple linear interpolation from its preceding and successive soil moisture values, provided there is no rainfall during that period. Condition of 'interpolation during no rainfall occurrence' is violated only once during August 12, 2011—August 23, 2011. However, it is mentioned later that during the model development period, the 1st hundred data are discarded to avoid the effect of initial value assumption (discussed in detail later). Therefore, this set of missing data, which is replaced by linear interpolation, does not affect the assessment of the model performance.

After this pre-processing, the total length of the dataset available is 668, i.e., July 01, 2011 to April 28, 2013. First two third of this dataset is used as the model development period (July 01, 2011 to September 14, 2012) and the rest is used as the testing period (September 15, 2012 to April 28, 2013).

2.2.2 Development of ARX Model

While developing the Auto-Regressive model with an exogenous input (ARX) model, as mentioned before, observed data from 10 *cm* is used as *forcing* and previous observation of SMC at 25 *cm* depth as *memory* to estimate the SMC at 25 *cm* depth. Similarly, to estimate the SMC at 50 *cm* depth, observed data from 25 *cm* (as forcing) and 50 *cm* (as memory) are used.

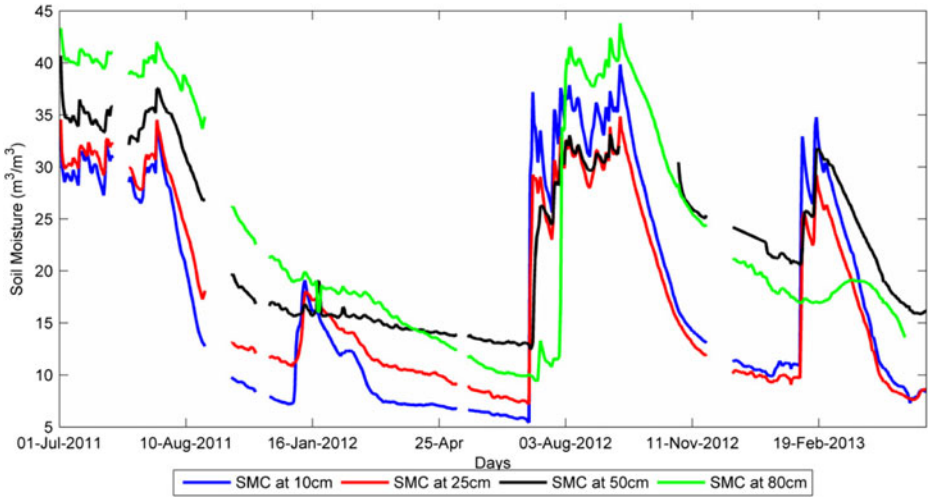


Fig. 1 Time series plot of observed SM data at 10, 25, 50 and 80 *cm* depth. The missing values replaced with linear interpolation are not shown here

The ARX model adopted in this study is expressed by equation (1),

$$SM_k(t) = \sum_{i=1}^{p} a_i SM_k(t-i) + \sum_{j=d}^{q+d-1} b_j SM_{k-1}(t-j) + e(t) \tag{1}$$

where, the $SM_k(t)$ is the soil moisture at the target depth k at time step t. The $SM_k(t-i)$ is the soil moisture at the target depth at previous time steps where $i=1,2,..,p$; $SM_{k-1}(t-j)$ is the soil moisture of the upper layer at previous time steps where $j=1,2,..,q$. The, a_i and b_j are the weighting function coefficients (aka ARX model parameters) for $SM_k(t-i)$ and $SM_{k-1}(t-j)$; p and q are the orders of the autoregressive and exogenous components respectively; and $e(t)$ is the white noise. The relative delay between the input soil moisture time series $SM_k(t-i)$ and the output soil moisture time series $SM_k(t)$, is considered by the delay parameter d, which is considered to be zero in our case. Values of $d>0$ implies delay of the output with respect to the input. Thus, the value of d is assumed to be zero in the model as it estimates soil moisture at deeper layer at the same time step as the upper layer which is reasonable for daily scale analysis of SM variation at different depths of the soil.

2.2.3 Estimating the Model Orders (p and q)

The selection of specific values of the model order is essential to have an adequate number of coefficients to take into account the required information about the association between the SMC values of surface layer and deeper layers, while maximizing such information with least number of parameters. Though theoretically p and q can take any positive integer value including zero, the optimal values of the orders of each soil moisture time series were investigated over the significant range of p and q, to obtain the desired criteria of the model order selection.

The identification of model order is mainly based on three criteria which are as follows,

a) The prediction focus or the model fit (MF).
b) The Mean Square Error (MSE) function.
c) The Akaike's Final Prediction Error (FPE).

The MF and MSE are expressed by equations (2) and (3) as follows –

$$MF = 100\left(1 - \frac{\sqrt{\sum_{i=1}^{n}\left(SM_i^{sim} - SM_i^{obs}\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(SM_i^{sim} - \overline{SM}\right)^2}}\right) \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(SM_i^{sim} - SM_i^{obs}\right)^2 \tag{3}$$

where SM^{sim} is the estimated variables in, SM^{obs} is the observed values, $\overline{SM}$ is the mean of the observed soil moisture data and n is the number of samples in the

dataset. Higher values of the MF are more favourable. The MSE is the second moment of the error that combines both the variance of the observed value and its bias. Since the MSE minimizes the variance, the lower value of MSE indicates the better model performance. The range of MF is 0 to 100 % whereas MSE may vary from 0 to infinity.

Akaike's Final Prediction Error (FPE) criterion provides a comparative measure of model performance. According to Akaike's theory, the most accurate model has the smallest FPE when the model is evaluated on a different testing data set. The FPE can be estimated by the equation (4),

$$FPE = V\left(\frac{1+\frac{m}{n}}{1-\frac{m}{n}}\right) \tag{4}$$

where, m is the number of estimated parameters which is equal to $(p+q)$, n is the number of values in the estimation dataset and V is a loss function. Mathematically, the loss function is the determinant of the obtained noise covariance matrix which can be computed by equation (5),

$$V = \det\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\varepsilon_i^T\right) \tag{5}$$

where, ε_i is the error at i th time step. In the simulation when $m<<n$, the FPE is approximated with equation (6),

$$FPE = V\left(1+\frac{2m}{n}\right) \tag{6}$$

The range of FPE is 0 to infinity. After determining the model orders based on prediction focus, MSE and FPE, the next step is to estimate the ARX model coefficients. In the following section the method to estimate the model coefficients, is described.

2.2.4 Determining the ARX Model Parameters (a_i and b_j)

During the determination of coefficients of the ARX model, the Least Square (LS) method minimizes the summation of the square of the residuals. As mentioned before, considering $d=0$, equation (1) can be expanded as equation (7),

$$SM_k(t) = a_1SM_k(t-1) + ... + a_pSM_k(t-p) + b_0SM_{k-1}(t) + ... + b_{q-1}SM_{k-1}(t-q+1) + e(t) \tag{7}$$

In equation (7), t varies from λ to n, where $\lambda=\max(p+1, q)$. The summation of the square of the error terms, i.e. $e(t)$, $t=\lambda,\cdots,n$, can be written as equation (8),

$$S = \sum_{t=\lambda}^{n}\left[\left\{SM_k(t)-a_1SM_k(t-1)-....-a_pSM_k(t-p)-b_0SM_{k-1}(t)-....-b_{q-1}SM_{k-1}(t-q+1)\right\}^2\right] \tag{8}$$

Since the LS method minimizes the summation of the square of the errors, the parameter values are computed by differentiating S with respect to each parameter, and assigning the differentiated value equal to zero. Mathematically, this can be represented as equation (9),

$$\left.\begin{aligned} \frac{\partial S}{\partial a_i} &= 0 \quad \forall i = 1, \cdots, p \\ \frac{\partial S}{\partial b_j} &= 0 \quad \forall j = 0, \cdots, (q-1) \end{aligned}\right\} \tag{9}$$

The matrix notation of equations (9) can also be expressed as equation (10),

$$B = \left(X^T X\right)^{-1} X^T Y \tag{10}$$

where, B is the column matrix consisting of the coefficients (a_i and b_j); X is the matrix formed by input variables, i.e. soil moisture values of target depth and forcing depths at previous time steps and Y is the output variable matrix i.e. the soil moisture at target depth. The details of the derivation for computation of ARX parameters by LS method is provided in the supplementary document.

The performance of developed model is evaluated using i) Root Mean Square Error (RMSE); ii) Unbiased RMSE (uRMSE); iii) Degree of agreement (Dr) and iv) Pearson Correlation Coefficient (CC). The details of these performance metrics are provided in the supplementary document.

While using the developed model, two issues need to be noted. The developed model uses the daily observed soil moisture data of 10 *cm* depth to estimate the soil moisture at 25 *cm*. However, the observed data may not be available for deeper layers (at new locations) and thus, the estimated (rather than observed) SMC at overlying layer (say 25 *cm* depth) is used as 'forcing' for underlying layer (i.e., 50 *cm*). This approach helps to evaluate the model efficacy to estimate the vertical SM profile using only the observed SM data from the surface layer. Secondly, since the SMC at a specific depth depends on the memory as well as on forcing, the initial values of the target depth are selected at time $(t-p)$ where p is equal to the number of the order of the AR parameters in the developed ARX model. The effect of such assumed initial values on the model output is also investigated. This is explained as follows.

Since the model aims to be kept independent of observed SMC values of underlying depths, observed SMC values of those layers are evaded to be used to initialize the model. Four alternatives are used as the initial value: i) Minimum SMC value of the forcing layer, ii) Mean SMC value of forcing layer, iii) Same SMC value as the forcing layer at the initial time-steps and iv) Maximum SMC value of the forcing layer. Whereas, these alternatives may not be theoretically or physically justified, it is better to treat them just as reasonable arbitrary values to explore the time horizon over which effect of initial assumption dies down. It is indeed possible to assume many alternatives as initial values. Models are developed and validated using these initial values and their performances are tested to assess the time horizon over which the effect of initial value assumption exists and can be ignored while evaluating the model performance.

3 Results

3.1 Sensitivity of Initial Value Assumption

Sensitivity of the estimated soil moisture to the assumed initial SMC values of the target depth is demonstrated for SMC estimate at 25 *cm* depth. The time series of observed value of soil

moisture at 25 *cm* depth is plotted with the estimated soil moisture time series with different initial values in Fig. 2. It is observed that the estimated SMC eventually converges to the observed SMC for all the four cases and behaves similarly thereafter. It is also seen that the estimated SMC values of 25 *cm* depth converge to the observed SMC within the first 100 time steps of the study period. Hence, for evaluating the model performance, the estimated SMC values of first 100 time steps are discarded during the model application to avoid the effect of initial value assumption. The performance metrics of both model development and testing periods for all four cases are shown in Table 1. Henceforth, the average of the SMC time series from 10 *cm* depth is used as the initial value to estimate the SMC of successive layers. This is to retain the fact that the approach utilizes only the surface layer's SMC information.

3.2 Model Development and Performance

To select the order of the model, the values of p and q are varied over one to five each. Different combinations of p and q and corresponding values of fitness criteria are described in supplementary document (Table S1). Based on maximum MF, minimum MSE and FPE, and the model parsimony criterion i.e. optimum number of model parameters should be minimum possible, the final model order is selected. The coefficients of the models are estimated by LS method as described in section 2.2.4. Table 2 summarizes the values of the selected model orders, the coefficients, model fit, FPE and MSE.

The time-series plots of estimated and observed SMC for target depths of 25, 50 and 80 *cm* are shown in Figs. 3, 4 and 5 respectively. It is noticed that the estimated SMC follows the

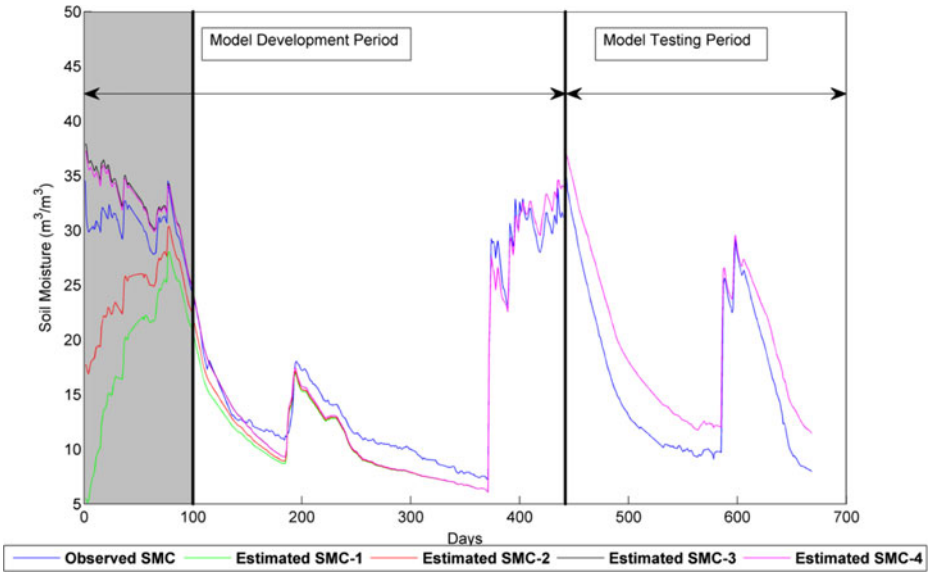


Fig. 2 Sensitivity analysis of the Initial values. Estimated SMC-1: The estimated SMC at target depth with initial value equals to the minimum value of SMC time series of the forcing layer; Estimated SMC-2: The estimated SMC at target depth with initial value equals to the average value of SMC time series of the forcing layer; Estimated SMC-3: The estimated SMC at target depth with initial value equals to the value of SMC of the forcing layer at same time step. Estimated SMC-4: The estimated SMC at target depth with initial value equals to the maximum value of SMC time series of the forcing layer. The first 100 data points are ignored from the performance assessment, which is shown in the shaded part

Table 1 Performance metrics for sensitivity analysis of the initial values of the target depths

Initial values	Performance metrics							
	Development period				Testing period			
	CC	D_r	RMSE	uRMSE	CC	D_r	RMSE	uRMSE
1	0.9926	0.8662	1.8004	1.2747	0.9488	0.8081	2.7148	2.6451
2	0.9939	0.8807	1.6125	1.1969	0.9525	0.8169	2.6429	2.5411
3	0.9930	0.8965	1.4621	1.2454	0.9566	0.8265	2.5898	2.4224
4	0.9931	0.8966	1.4632	1.2403	0.9565	0.8266	2.5902	2.4252

Note: Initial Value1: Minimum value of the SMC time series of the Forcing layer; Initial Value 2: Mean of the SMC time series of the Forcing layer; Initial Value3: Identical as the Forcing layer at same time step; Initial Value 4: Maximum value of the SMC time series of the Forcing layer

observed SMC for all three target depths reasonably well. The model performance is best at 25 *cm* and reasonably good at 80 *cm* except for significant overestimation for the last peak of SMC. At 50 *cm* target depth, there is visible increase in the error especially in the testing period.

The performance metrics for the development and testing periods of the final models are summarized in Table 3. The superior model performance, indicated by the high values of CC and Dr, and low RMSE and uRMSE during both model development and testing periods indicates that the models have reasonably good performance for all three layers.

3.3 Comparative Study of Memory and Forcing

To study the individual effects of memory and forcing, order of only one component (memory or forcing) is kept as zero and the other component (forcing or memory) is varied from 1 to 5 at a time. Table 4 represents the model performance criteria described in section 2.2.3, when the effect of *only memory* and *only forcing* on the estimated SMC is considered at different depths. After developing the ARX model with only forcing component, the new models were applied to the estimated data with assumed initial values, to assess the model performance at different depths. Table 5 describes the performance metrics of the ARX model with only forcing component.

4 Discussion

The sensitivity test of initial value assumption suggests that effect of initial value dies down gradually after initial 100 time steps. Since average of the SMC time series from 10 *cm* depth is used as the initial value, suitability of the developed model is ensured at ungauged locations,

Table 2 Estimated values of the parameters during model development

D1	D2	p	q	a_1	a_2	a_3	b_1	b_2	b_3	MF(%)	FPE	MSE
25	10	1	3	0.978	NA	NA	0.576	−0.443	−0.110	96.4	0.112	0.110
50	25	2	2	1.197	−0.236	NA	0.229	−0.185	NA	94.8	0.206	0.203
80	50	3	2	1.224	−0.333	0.073	0.400	−0.359	NA	92.6	0.779	0.764

Note: D_1: Target depth of estimation; D_2: Depth of forcing layer

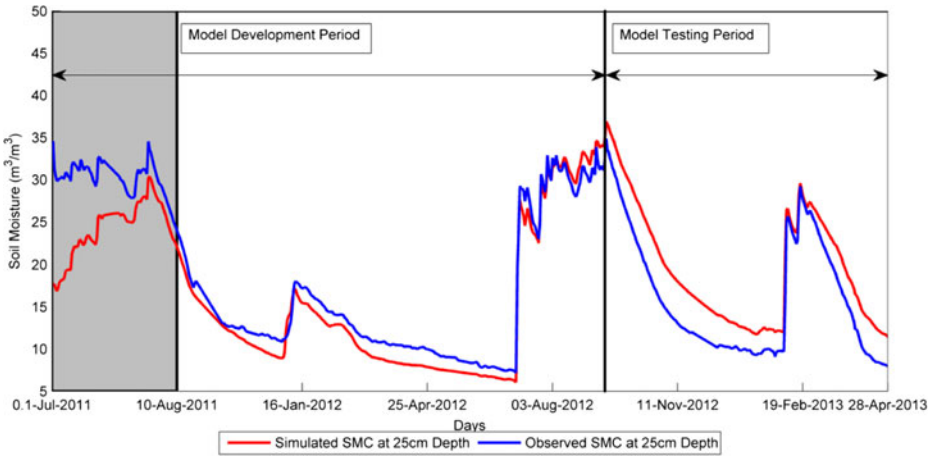


Fig. 3 Time series plot of training and testing dataset of estimated output and observed soil moisture at 25 *cm* depth

where information on SMC at deeper layers are not available. This may be treated as a valuable potential of the proposed approach to integrate satellite estimated 2-dimensional surface soil moisture map to prepare a 3-dimensional soil moisture profile.

The observation from Table 2 indicates the increasing order of memory and decreasing order of forcing with increase in soil depth. Thus, it can be concluded that the effect of forcing reduces with increase in depth while memory component dominates at deeper layers. The higher order of the forcing component indicates faster its propagation caused by low water retention capacity and high hydraulic conductivity indicating the soil type of the study area. The value of model fit decreases while MSE and FPE increase with increase in target depth in model development showing that model performance is expected to reduce with increase in target depth. This trend is also noticed from Table 3 that the model performs best for target depth of 25 *cm* and the performance gradually decreases for subsequent depths. This may be attributed to the increasing distance between the surface layer and the target depths. Also, the

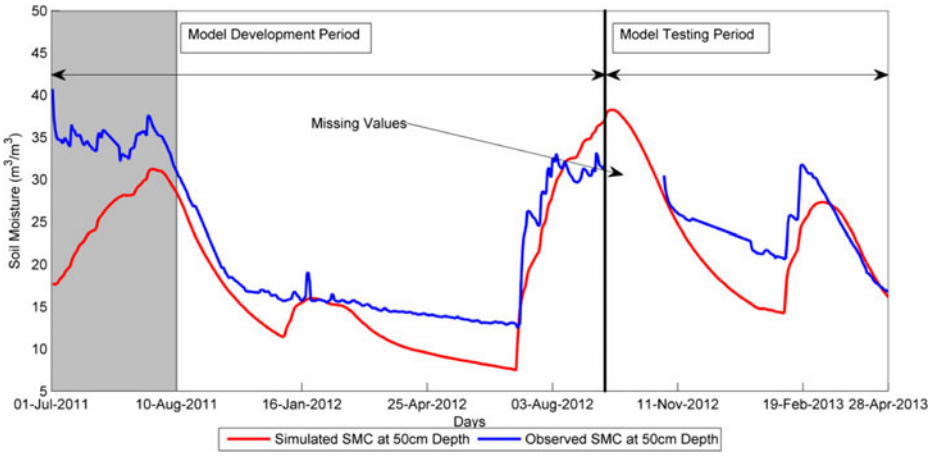


Fig. 4 Time series plot of training and testing dataset of estimated output and observed soil moisture at 50 *cm* depth

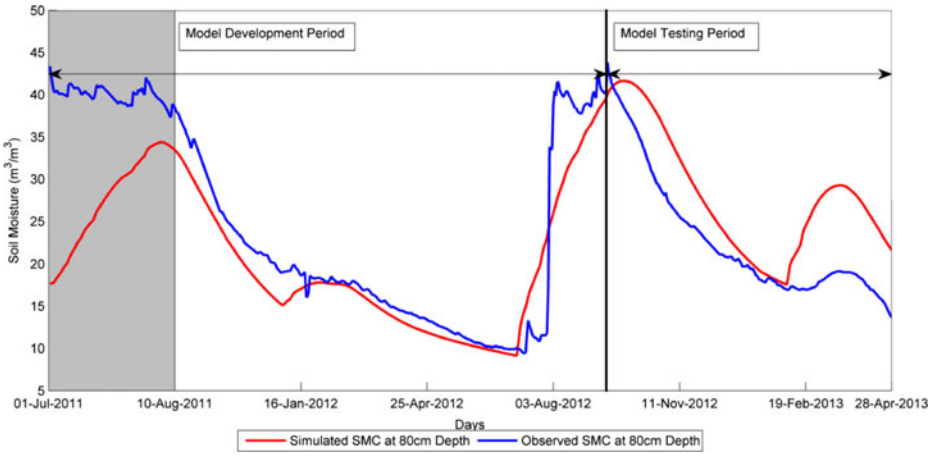


Fig. 5 Time series plot of training and testing dataset of estimated output and observed soil moisture at 80 *cm* depth

error in estimation of SMC at 25 *cm* depth propagates to the estimation of SMC at 50 *cm* and so forth. This explains the heterogeneous and complex nature of moisture propagation towards deeper layers which is influenced not only by the surface soil moisture but also soil properties and the lateral subsurface flow.

It is noticed in case of 25 and 80 *cm* target depths, there is a difference between the RMSE and uRMSE during the testing period. This indicates that there could be some bias between the observed and estimated SMC values. The computed biases are found to be below 25 % (21.6 and 23.2 %) of the observed values for the target depths 25 and 80 *cm* respectively.

Apart from this, there is overestimation (e.g. testing period: 25 and 80 *cm*) and underestimation (e.g. testing period: 50 *cm*) during some part of the model development and testing period. Though it cannot be specifically attribute to some reason, presence of missing values (testing period: 50 cm), propagation of its effect to deeper layers and role of memory only situation (initial part of testing period: 25 and 80 *cm*) might be some of possible reasons. Longer and more accurate data might be necessary to fully exploit the potential of the proposed model.

While studying the individual effects of memory and forcing, it is seen that the use of only memory component is almost comparable with the developed model (Table 2 and Table 4). But in ungauged locations the observed soil moisture data is not available for deeper layers. Therefore, the model development with only memory component is not adopted. In contrast

Table 3 Performance metrics during model development and testing periods

D_1	D_2	Performance metrics							
		Development period				Testing period			
		CC	D_r	RMSE	uRMSE	CC	D_r	RMSE	uRMSE
25	10	0.994	0.881	1.613	1.197	0.983	0.716	3.691	1.281
50	25	0.963	0.702	3.631	2.562	0.853	0.478	4.457	4.016
80	50	0.948	0.831	3.826	3.399	0.888	0.535	6.310	3.428

Note: D_1: Target depth of estimation; D_2: Depth of forcing layer

Table 4 Performance metrics during the model development and testing period when the memory component ($p = 0$) and forcing component are absent ($q = 0$) respectively

D_1	D_2	Model order (p, q)	MF (%)	FPE	MSE	Model order (p, q)	MF (%)	FPE	MSE
25	10	0, 1	66.77	9.39	9.37	1, 0	90.00	0.85	0.85
25	10	0, 2	67.78	8.87	8.81	2, 0	90.48	0.78	0.77
25	10	0, 3	68.34	8.60	8.50	3, 0	90.51	0.78	0.76
25	10	0, 4	68.77	8.41	8.27	4, 0	90.55	0.78	0.76
25	10	0, 5	69.14	8.24	8.08	5, 0	90.56	0.78	0.76
50	25	0, 1	57.61	13.51	13.49	1, 0	93.38	0.33	0.33
50	25	0, 2	59.65	12.31	12.22	2, 0	94.00	0.27	0.27
50	25	0, 3	61.21	11.42	11.30	3, 0	94.01	0.28	0.27
50	25	0, 4	62.74	10.59	10.42	4, 0	94.03	0.28	0.27
50	25	0, 5	63.77	10.06	9.86	5, 0	94.05	0.27	0.27
80	50	0, 1	59.20	23.37	23.33	1, 0	91.87	0.93	0.93
80	50	0, 2	61.01	21.45	21.31	2, 0	92.15	0.87	0.86
80	50	0, 3	62.06	20.40	20.17	3, 0	92.18	0.87	0.86
80	50	0, 4	63.31	19.16	18.87	4, 0	92.49	0.81	0.79
80	50	0, 5	64.35	18.17	17.81	5, 0	92.49	0.82	0.79

Note: D_1: Target depth of estimation; D_2: Depth of forcing layer

Table 5 Performance metrics during model development and testing period when memory component ($p = 0$) was absent and the model coefficients were applied to the estimated SMC

D1	D2	q	Performance metrics							
			Development period				Testing period			
			CC	D_r	RMSE	uRMSE	CC	D_r	RMSE	uRMSE
25	10	1	0.984	0.751	3.372	3.126	0.995	0.833	2.511	1.469
25	10	2	0.986	0.757	3.279	3.023	0.998	0.825	2.535	3.023
25	10	3	0.987	0.761	3.227	2.969	0.997	0.819	2.604	1.402
25	10	4	0.987	0.765	3.189	2.930	0.997	0.815	2.659	1.422
25	10	5	0.987	0.768	3.156	2.894	0.996	0.811	2.714	1.438
50	25	1	0.918	0.565	5.196	4.570	0.868	0.315	5.671	4.502
50	25	2	0.929	0.577	5.029	4.359	0.878	0.335	5.584	4.500
50	25	3	0.937	0.585	4.899	4.201	0.880	0.344	5.526	4.499
50	25	4	0.944	0.595	4.773	4.051	0.879	0.348	5.487	4.519
50	25	5	0.944	0.595	4.773	4.051	0.879	0.348	5.487	4.519
80	50	1	0.887	0.761	5.010	4.543	0.828	0.614	5.850	4.469
80	50	2	0.905	0.772	4.720	4.169	0.835	0.601	5.931	4.424
80	50	3	0.914	0.778	4.575	3.984	0.838	0.592	6.011	4.421
80	50	4	0.922	0.784	4.429	3.790	0.840	0.581	6.109	4.429
80	50	5	0.929	0.790	4.310	3.634	0.841	0.572	6.207	4.446

Note: D_1: Target depth of estimation; D_2: Depth of forcing layer

that the model with only forcing component performs much worse than the developed model (Table 4 and Table 2). Physically, it can be said that the effect of forcing gradually diminishes for the deeper layers in the soil. Hence, estimation of soil moisture with only forcing component gives better performance metrics in near surface layers compared to deeper layers.

5 Conclusions

The modelling philosophy of the present study aims at modelling two important factors—memory and forcing, to estimate the vertical soil moisture profile using an ARX model. Following conclusions can be drawn from the findings of this study.

- The proposed model is found to be effective in modelling the vertical soil moisture profile using only the surface soil moisture information with better performance for the near surface layers. Still the model performance for all the depths is reasonably well given the complexity of problem.
- The contribution of memory increases with the increase in depth whereas the contribution of forcing decreases with the increase in depth.
- The developed model needs to be initialized with a few initial values assumptions. However, the effect of this initial assumption dies down gradually and it is recommended to ignore the first 100 time steps to ensure the exclusion of the effect of initial value assumption. It is also found that the mean of the surface soil moisture time series can be used as the initial value for all the depths.
- The use of only memory component is almost comparable with the developed model. But implementing the model with only the memory component is not possible since the observed vertical SM profile data are not available for ungauged locations. Thus, one of the potential of the proposed approach is that the developed model is suitable for ungauged locations, where information on SMC at deeper layers are not available.

The future scope of the study lies in linking the soil characteristics to the memory and forcing components included in the proposed approach. As mentioned before, the integration of remote sensing surface soil moisture data to the proposed approach will allow to implement the model in different ungauged locations provided the soil texture is same as the present study area. This aspect is kept as future scope of this study.

Acknowledgments Authors wish to acknowledge Dr. Shivam Tripathi, IIT Kanpur for providing the daily soil moisture data from the study area. This is a part of International Soil Moisture Network. For details: https://ismn.geo.tuwien.ac.at/.

Compliance with Ethical Standards

Funding This study is partially supported by a research project sponsored by the Ministry of Earth Science (MoES), Government of India (Sanction order# MoES/PAMC/H&C/30/2013-PC-II).

Conflict of Interest Corresponding author Dr. Rajib Maity has received research grants from the Ministry of Earth Science (MoES), Government of India (Sanction order# MoES/PAMC/H&C/30/2013-PC-II). The other authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals The authors declare that the research does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent The authors declare that the 'Informed Consent' is not applicable in the research since it does not contain any studies with human participants or animals performed by any of the authors.

References

Argyrokastritis I, Kargas G, Kerkides P (2009) Simulation of soil moisture profiles using K(h) from coupling experimental retention curves and one-step outflow data. Water Resour Manag 23(15):3255–3266. doi:10.1007/s11269-009-9432-3

Chen ZQ, Govindaraju RS, Kavvas ML (1994) Spatial averaging of unsaturated flow equations under infiltration conditions areally heterogeneous fields:Numerical simulations. Water Resour Res 30(2):535–548

Dorigo WA, Wagner W, Hohensinn R, Hahn S, Paulik C, Xaver A, Gruber A, Drusch M, Mecklenburg S, van Oevelen P, Robock A, Jackson T (2011) The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrol Earth Syst Sci 15:1675–1698. doi:10.5194/hess-15-1675-2011

Downer CW, Ogden FL (2004) Appropriate vertical discretization of Richards' equation for two-dimensional watershed-scale modelling. Hydrol Process 18(1):1–22. doi:10.1002/hyp.1306

Haverkamp R, Parlange JY, Cuenca R, Ross PJ, Steenhuis TS (1998) Scaling of the Richards' equation and its application to watershed modeling. In: Sposito G (ed) In scale invariance and scale dependence in hydrology. Cambridge University Press, New York, pp 190–223

Hoeben R, Troch PA (2000) Assimilation of active microwave observation data for soil moisture profile estimation. Water Resour Res 36:2805–2819. doi:10.1029/2000WR900100

Hsu SM, Asce M, Ni C et al. (2002) Assessment of three infiltration formulasbased on model fitting on Richards equation. J Hydrol Eng (October), 373–379

Jishnu RB, Naik SP, Patra NR, Malik JN (2013) Ground response analysis of Kanpur soil along Indo-Gangetic Plains. Soil Dyn Earthq Eng 51:47–57

Kale R, Sahoo B (2011) Green–Ampt infiltration models for varied field conditions: a revisit. Water Resour Manag 25(14):3505–3536

Kerr YH, Wadlteufel P, Wigneron JP, Delwart S, Cabot F, Boutin J, Escorihuela MJ, Font J, Reul N, Gruhier C, Juglea SE, Drinkwater MR, Hahne A, Martı'n-Neira M, Mecklenburg S (2010) The SMOS Mission: new tool for monitoring key elements of the global water cycle. Proc IEEE 98(5):666–687

Kim S (2009) Multivariate analysis of soil moisture history for a hillslope. J Hydrol 374(3-4):318–328. doi:10.1016/j.jhydrol.2009.06.025

Kim S, Kim H (2007) Stochastic analysis of soil moisture to understand spatial and temporal variations of soil wetness at a steep hillside. J Hydrol 341(1–2):1–11. doi:10.1016/j.jhydrol.2007.04.012

Kim S, Sun H, Jung S (2011) Configuration of the relationship of soil moistures for vertical soil profiles on a steep hillslope using a vector time series model. J Hydrol 399(3–4):353–363. doi:10.1016/j.jhydrol.2011.01.012

Mahmood R, Littell A, Hubbard KG, You J (2012) Observed data-based assessment of relationships among soil moisture at various depths, precipitation, and temperature. Appl Geogr 34:255–264. doi:10.1016/j.apgeog.2011.11.009

Moran MS, Peters-Lidard CD, Watts JM, McElroy S (2004) Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can J Remote Sens 30(5):805–826. doi:10.5589/m04-043

Pielke RA Sr (2001) Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. Am Geophys Union 39(2):151–177

Singh VP (2010) Entropy theory for movement of moisture in soils. Water Resour Res 46:W03516. doi:10.1029/2009WR008288

Varado N, Braud I, Ross PJ, Haverkamp R (2006) Assessment of an efficient numerical solution of the 1D Richards' equation on bare soil. J Hydrol 323(1–4):244–257. doi:10.1016/j.jhydrol.2005.07.052

Zucco G, Brocca L, Moramarco T et al. (2014) Influence of land use on soil moisture spatial–temporal variability and monitoring. J Hydrol 1–7. doi: 10.1016/j.jhydrol.2014.01.043