

Research papers

Development of a spatially-varying Statistical Soil Moisture Profile model by coupling memory and forcing using hydrologic soil groups

Manali Pal, Rajib Maity*

Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

ARTICLE INFO

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Zhiming Lu, Associate Editor

Keywords:

Soil Moisture Content (SMC)
Statistical model
Spatial transferability
Soil moisture profile
Hydrological Soil Group (HSG)

ABSTRACT

Information on vertical Soil Moisture Content (SMC) profile is important for several hydro-meteorological processes. This study borrows the idea of coupling the memory and forcing from a previous study and develops a spatially-varying Statistical Soil Moisture Profile (SSMP) model to estimate the vertical SMC profile. It uses only surface soil moisture (0–5 cm) values and Hydrological Soil Groups (HSGs) information of the location. The focus of the study is incorporation of the HSG information to ensure the spatial transferability of the proposed model by capturing the spatial variations of soil moisture profile with the change in soil hydraulic properties. The wide range of soil moisture data for model development as well as for spatial validation is obtained from 171 stations from different networks of International Soil Moisture Network (ISMN) at five different depths, i.e., 5, 10, 20, 51 and 102 cm. The HSG information at the locations are extracted from the Web Soil Survey (WSS) database. The potential of spatial transferability of the SSMP model is assessed by applying it to the new stations within the corresponding HSG. Model performances are promising for all four depth pairs (5–10, 10–20, 20–51 and 51–102 cm) of all four HSGs during both model development and spatial validation given the model complexity. Hence, the spatially-varying SSMP model is suitable at the ungauged locations by incorporating the HSG information. The potential application of the proposed model shows the future scope to assimilate the satellite based surface SMC data into the model to develop a vertical soil moisture profile map over a large area.

1. Introduction

Soil Moisture Content (SMC) of the unsaturated zone i.e. the vertical soil moisture profile plays a significant role in determining the water and energy fluxes between soil and atmosphere (Famiglietti et al., 1998) as well as vegetation growth (Yang et al., 2012). Recently, the retrieval of surface SMC from remote sensing data is in the research interest due to its large scale and fine resolution estimation (Bertoldi et al., 2014). However, the remote sensing is capable of retrieving the soil moisture information only for the top few centimeters (5–10 cm) of surface layer (Kerr et al., 2010).

The surface SMC is associated with the root-zone SMC and it is possible to obtain soil moisture profile assessment using the surface soil moisture information (Calvet and Noilhan, 2000; Albergel et al., 2008; Singh, 2010) since it is coupled to root-zone SMC through diffusion processes (Singh, 2010). Utilizing this concept many studies have attempted to estimate the root zone soil moisture by extrapolating the surface soil moisture (Wagner et al., 1999; Manfreda et al., 2014; Manfreda et al., 2014; Renzullo et al., 2014; Dumedah et al., 2015). The data assimilation techniques and the exponential filter proposed by

Wagner et al. (1999) are the most extensively used methods among these. The exponential filter needs the wilting level, field capacity, and porosity information and can be applied to the regions with same climatic and crop conditions. These pre-requisites limit the application of exponential filter as the information may not be available for the other ungauged locations. Its application is based on the assumption of a constant hydraulic conductivity of soil, whereas, in practical scenario it can vary by several degrees of magnitude. Soil moisture profile estimation from remote sensing data has focused on data assimilation into Land Surface Models (LSMs) based on the association of near-surface soil moisture and the root-zone soil moisture through diffusion processes. LSMs use the soil hydraulic property information derived from the pedotransfer function by Cosby et al. (1984) and a set of default or spatially uniform model parameters (Li et al., 2011). These simplified and empirically derived default soil hydraulic parameters are inadequate to describe the soil moisture variability in spatially heterogeneous landscapes. Thus, the uncertainties due to the inaccurate physical description of the water and energy balance hinder the application of such techniques (Sabater et al., 2007).

It is established that the mutual association of SMC values decreases

* Corresponding author.

E-mail address: rajib@civil.iitkgp.ernet.in (R. Maity).

with the increase in the gap between two soil layers (Mahmood et al., 2012). However, the mutual association between SMC at different soil depths and the stochastic features of soil moisture dynamics can be evaluated through statistical methods such as cross-correlation method and Vector Auto Regression (VAR) method (Kim and Kim 2007; Kim 2009; Kim et al. 2011; Mahmood et al. 2012; Pal et al., 2016). This study borrows the idea of coupling the memory (temporal persistence) and forcing (input from overlying layers) from Pal et al. (2016). The memory of SMC is a measure of the time length when a moisture anomaly caused by wet or dry conditions is identifiable and impacts the atmosphere. The importance of soil moisture memory has been investigated in many studies with autocorrelation-based approaches, Markov chain, chaos theory along with studying from observations, integrations with LSMs and Atmospheric General Circulation Models (AGCMs) (Sridhar et al., 2002; Seneviratne et al., 2006; Yan et al., 2015; Sivakumar, 2017). The autocorrelation-based approaches used in AGCMs represent the variability of soil moisture memory component (Seneviratne et al., 2006). Using this soil moisture memory concept, many studies have used the coupling in a particular region or season with meteorological forcings such as evapotranspiration, precipitation or net radiation (Ghannam et al., 2016; Sörensson and Menéndez, 2011; Seneviratne et al., 2006). Pal et al. (2016), used the concept of coupling the soil moisture memory with the soil moisture information of the overlying layer which is considered the forcing to assess the root-zone soil moisture by autocorrelation-based approach. However, the spatial transferability was not investigated which is important so as to apply the developed model at ungauged locations. The spatial variation of soil moisture is controlled by its association with soil texture, vegetation, topography, precipitation and other hydroclimatic variables. Each of these interdependent controlling factors impacts the spatial distribution of SMC depending on the characteristics of heterogeneity present in the area and varies with time and scale. However, the water holding capacity in the unsaturated zone and the variability of soil moisture distribution is directly influenced by the soil hydraulic properties (Price et al., 2010; Kim and Barros, 2002) and soil texture (Jawson and Niemann, 2007). The hydraulic properties of soil determine the hydraulic conductivity, matric potentials affecting the flow of water through soil, moisture available to plants (Gaur and Mohanty, 2013) and quantity of precipitation inflowing and retained in subsurface storage (Farres, 1987; Rawls et al., 1993; Cerda, 1996). Hence, the incorporation of soil hydraulic properties in estimation of vertical soil moisture profile is beneficial. The Hydrological Soil Groups (HSGs) are categorized based on their infiltration characteristics which in turn mainly depend on the soil hydraulic properties especially the hydraulic conductivity. The classification of various soil types based on these soil-hydrologic factors is defined in the US National Engineering Handbook as four major HSGs, viz., A, B, C, and D. Originally, the soils were assigned to the four HSGs based on measured precipitation, runoff and infiltration data (Musgrave 1955). Simply, the water transmitting soil layer with the lowest saturated hydraulic conductivity, depth to any layers that is more or less water impermeable and depth to a water table (if present) determines the HSG. The rate of infiltration decreases from HSG-A to HSG-D (USDA, 2009). A brief description on different HSGs is provided in Table 1. The Web Soil Survey (WSS) operated by the USDA Natural Resources Conservation Service (NRCS) suggests that the soils are allotted to one of the four HSGs according to the infiltration rate when the soils are not protected by vegetation, thoroughly wet, and get rainfall from long-duration storms. However, the proposed approach utilizes the surface SMC that already bears the signature of the above mentioned factors. Summarizing the aforementioned discussion, the consideration of the HSGs implies the integration of the effects of the physical controls into the SMC profile estimation and may provide a potential information to develop spatially varying statistical model to estimate vertical SMC profile.

In brief, based on the research gaps, the objective of the study is to develop a spatially-varying, statistical approach linking the surface soil

moisture to the deeper layers to estimate vertical SMC profile borrowing the already established concept of coupling memory (temporal persistence) and forcing (input from overlying layers). The developed model is named as Statistical Soil Moisture Profile (SSMP) model. The spatial transferability of the proposed statistical approach is explored through incorporating the HSG information since it demonstrates the effect of the physical controls viz. soil texture, hydraulic conductivity, runoff and infiltration.

2. Data and study area

The soil moisture time series data is obtained from the International Soil Moisture Network (ISMN) website (Dorigo et al. 2011) (<http://www.wcc.nrcs.usda.gov/scan/>) initiated by Vienna University of Technology, Austria. The worldwide in situ soil moisture measurements from different networks and validation operations are collected, synchronized, and made accessible to users through the ISMN. This study utilizes the soil moisture data from three different networks to develop the SSMP model as well as for the spatial validation which are described in the previous sections. The daily time series of SMC data is used from Soil Climate Analysis Network (SCAN), U.S. Climate Reference Network (USCRN) and SNOwpack TELemetry (SNOTEL) networks.

The SNOTEL system measures the soil moisture with the hydra-probe sensor. The temperature range of the probes is from -10°C to $+65^{\circ}\text{C}$ and these are able to measure water only in the liquid state. Therefore, the studies using the soil moisture at different depths use the soil moisture values corresponding to soil temperatures $> 0^{\circ}\text{C}$. The hourly SMC data is collected from total 171 monitoring stations from these three different networks at five different depths of 5, 10, 20, 51 and 102 cm and converted to daily SMC data. The observed SMC data from at least 20 stations with good quality data from each HSG (HSG A-21, HSG B-28, HSG C-33 and HSG D-21) is used for the model development to incorporate the properties of vast range of SMC variability into the proposed models. Henceforth these stations are called as *model development stations*. It may be noted that the model testing is also carried out with these stations but for different time period. Remaining stations are used for spatial validation, henceforth called *spatial validation stations*. Numbers of such stations from each HSG are as follows: HSG A-17; HSG B-22; HSG C-17; and HSG D-17. The HSG of each monitoring stations are determined from the Web Soil Survey (WSS) (<https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>). It provides soil data and information produced by the National Cooperative Soil Survey and operated by the USDA Natural Resources Conservation Service (NRCS).

3. Methodology

3.1. Data preprocessing

3.1.1. Missing value treatment

Missing data periods may be present in observed daily soil moisture data. Extensive periods (> 20 days) of missing data are discarded since it cannot be filled up with reasonable accuracy. Moreover, some time steps consist of the complete time series for deeper layers except the surface layer information. Such time steps are also discarded since the study attempts to obtain the complete vertical SMC profile using only surface soil moisture information. However, shorter periods of missing values (20 days or less) are substituted by simple linear interpolation from its preceding and successive soil moisture values (Pal et al., 2016) (Fig. 1).

3.1.2. Data transformation

The transformation of data to a common probability distribution form is another important step since the soil moisture data at any two locations mostly do not follow the same range and distribution. Keeping spatial transferability as a focus of the developed model, the data

Table 1
The description of four HSGs (Das and Maity, 2015; USDA, 2009).

| HSG | Description | Infiltration Rate (mm/h) |
|-----|--|--------------------------|
| A | Soils in this group have high infiltration rates even when thoroughly wetted and have a high rate of water transmission. Group A soils typically consists of less than 10% clay and well to excessively drained sand (more than 90%) or gravel and have gravel or sand textures. Some soils having loamy sand, sandy loam, loam or silt loam textures may be considered in this group if they are well aggregated, having low bulk density, or consists of greater than 35 percent rock fragments. | > 25 |
| B | Soils in this group have moderate infiltration rates when thoroughly wetted. Water transmission through the soil is unhindered and thus it has a moderate rate of water transmission. Group B soils typically have between 10% and 20% clay and 50% to 90% sand and have loamy sand or sandy loam textures. It consists of mainly moderately deep to deep, moderately well to well-drained soils. Some soils having loam, silt loam, silt, or sandy clay loam textures may be considered in this group if they are well aggregated, have low bulk density, or contain greater than 35 percent rock fragments. | 12.5–25 |
| C | Soils in this group have moderately slow infiltration rates when thoroughly wetted. Water transmission through the soil is somewhat restricted and it has a moderate rate of water transmission. Group C soils typically consists moderately well to well-drained soils having between 20% and 40% clay and less than 50% sand and have loam, silt loam, sandy clay loam, clay loam, and silty clay loam textures. Some soils having clay, silty clay, or sandy clay textures may be considered in this group if they are well aggregated, have low bulk density, or contain greater than 35 percent rock fragments. | 2.5–12.5 |
| D | Soils in this group have very slow infiltration rates when thoroughly wetted. Water movement through the soil is restricted. Group D soils typically consists of greater than 40% clay with a high swelling potential, less than 50% sand, and have clayey textures. All soils with a depth to a water impermeable layer less than 50 cm and all soils with a water table within 60 cm of the surface are in this group. It consists of soils with a permanent high water table, soils with a clay pan or clay layer at or near the surface, and shallow soils over nearly impervious material. | < 2.5 |

transformation to a common distribution is necessary. Secondly, the proposed SSMP model is motivated by concept of the Box-Jenkins approach to couple memory and forcing as established in an earlier study (Pal et al., 2016), however, with an additional consideration of spatial transferability. As a requirement of any Box-Jenkins approach, the data should follow normal distribution, which is the requirement of the proposed SSMP model as well (Box et al., 2015). However, the SMC data may deviate (sometime significantly) from normal distribution (Choi and Jacobs, 2007). Thus, the data is transformed to follow an approximate normal distribution. Kernel density approach is used to transform the observed SMC data to follow a normal distribution since any parametric distribution could not describe the observed SMC data for any layers to develop the SSMP model. The steps to transform the data (x) are as follows. These steps can be read along with the Fig. 2 for an easy understanding.

Step-1: Non-parametric kernel density approach is applied to obtain a non-parametric Probability Density Function (PDF), which is converted to Cumulative Distribution Function (CDF).

Step-2: Non-exceedence probability of each data point is obtained using this CDF. These values are also known as reduced variate (shown

as ‘u’ in Fig. 2).

Step-3: The reduced variates are transformed through an inverse to standard normal distribution. These values are shown as ‘U’ in Fig. 2.

Step-4: The transformed standard normal variates are used for model development (details are discussed in the subsequent sections).

Step-5: The model estimated values are back transformed through a reverse process described in Step 3. These values are shown as \hat{u} in Fig. 2.

Step-6: The \hat{u} values are back transformed through the inverse CDF developed in Step 1 to obtain the corresponding estimated SMC values (\hat{x}) as shown in Fig. 2.

It may be noted that the non-parametric distribution using kernel density can be estimated as (Zambom and Dias, 2012),

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where n is the sample size, $K(\bullet)$ is the kernel smoothing function which defines the curve used to generate the PDF and h is the bandwidth which is a function of number of data-points and vary accordingly for each station of each HSG. Normal kernel function is used in the present

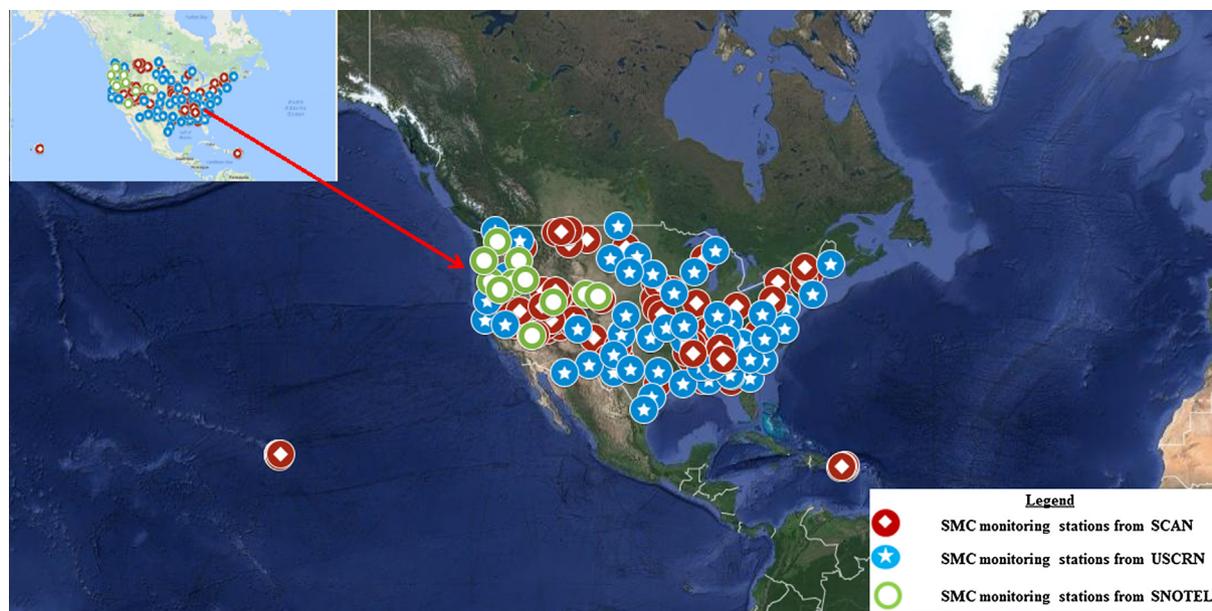


Fig. 1. The study area showing the SMC monitoring stations from SCAN, USCRN and SNOTEL.

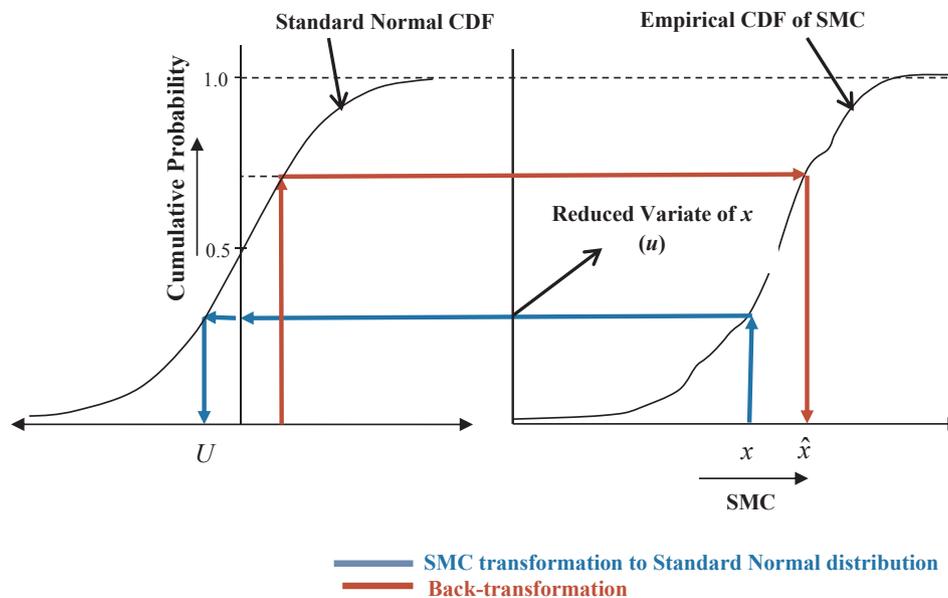


Fig. 2. The illustrative figure of data transformation to standard normal distribution using kernel density approach.

study to estimate the density of the observed SMC data for all the depths. The Normal kernel function is represented by the following equation,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2)$$

After the data transformation, the coupling approach is applied to the transformed data obtained from the observed data of each *model development* station of all four HSGs. The following sub-section describes the coupling equation to estimate the soil moisture at deeper layers which is based on the approach proposed by Pal et al. (2016).

3.2. SSMP model

Firstly, all the available stations (both for *model development stations* and *spatial validation stations*) are categorized into four groups according to the HSGs where the stations belong to. At a particular *model development* station, the estimation of soil moisture at deeper layers is based on the concept of coupling the ‘memory’ and ‘forcing’ as proposed by Pal et al. (2016). In the context of SSMP, the memory may be defined as the previous values of soil moisture in the same layer (underlying layer) and the forcing may be defined as the current and previous values of the soil moisture of the overlying layer.

3.2.1. Model at a location

Adapted from Pal et al. (2016), the basic formulation of SSMP model at a location is represented by the following form of equation which is applied on the transformed data,

$$SM_k(t) = \sum_{i=1}^p a_i SM_k(t-i) + \sum_{j=d}^{q+d-1} b_j SM_{k-1}(t-j) + e(t) \quad (3)$$

where $SM_k(t)$ is the transformed standard normal variate of soil moisture at the target depth k at time step t . The $SM_k(t-i)$ are the transformed standard normal variates of soil moisture at the target depth at preceding time steps where $i = 1, 2, \dots, p$. The $SM_{k-1}(t-j)$ are the soil moisture of the overlying layer at preceding time steps where, $j = 0, 1, \dots, q-1$. The weighting function coefficients for $SM_k(t-i)$ and $SM_{k-1}(t-j)$ are represented as a_i and b_j ; the orders of the memory and forcing components are represented as p and q respectively; and $e(t)$ represents 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 shown as d . The output is delayed with respect to the input if the values of $d > 0$. However, at daily scale this delay factor can be considered to be zero.

Parameters of this model are the orders (p and q) and the corresponding coefficients (a_1, a_2, \dots, a_p and b_0, b_1, \dots, b_{q-1}) of the memory and forcing components. The orders of memory and forcing components show the number of days (for a daily time series) of the contribution of memory and forcing respectively. The estimation of model order is based on a few statistical measures, such as Model Fit (MF), Mean Square Error (MSE) and Akaike’s Final Prediction Error (FPE), and the coefficients are estimated using the least square method which minimizes the summation of the square of the residuals as also used in Pal et al. (2016). The details of the parameter estimation criteria and methods are described in Appendix A.

Using this approach, individual models at each station are developed for four depth pairs (the adjoining layers) i.e. 5–10 cm, 10–20 cm, 20–51 cm and 51–102 cm. The next step is to define the specific model order and coefficient values for each HSG from the individual models developed to impart spatial transferability, which is discussed in the next section.

3.2.2. Model with spatial transferability

Individual models, developed at *model development stations*, are grouped according to the HSGs. The median value of the model orders corresponding to each depth pairs across all the stations within a particular HSG is computed. These values are used as the final model orders for that particular depth pair and HSG. This is repeated for all the depth pairs and HSGs. Subsequently, the model coefficients are re-computed with the final model orders at each station within a particular HSG used for model development. Next, the mean values of the coefficients are computed. These values are used as the final coefficients for that particular depth pair and HSG subject to the validity of stationarity assumption. Generally, averaging the coefficients of different Box-Jenkins models may lead to non-stationarity of the resulting model. Hence, the stationarity is checked before finalizing the set of SSMP model.

3.3. Spatial validation

Spatial validation of the developed SSMP model should be carried out at new locations having daily time series of only surface SMC values

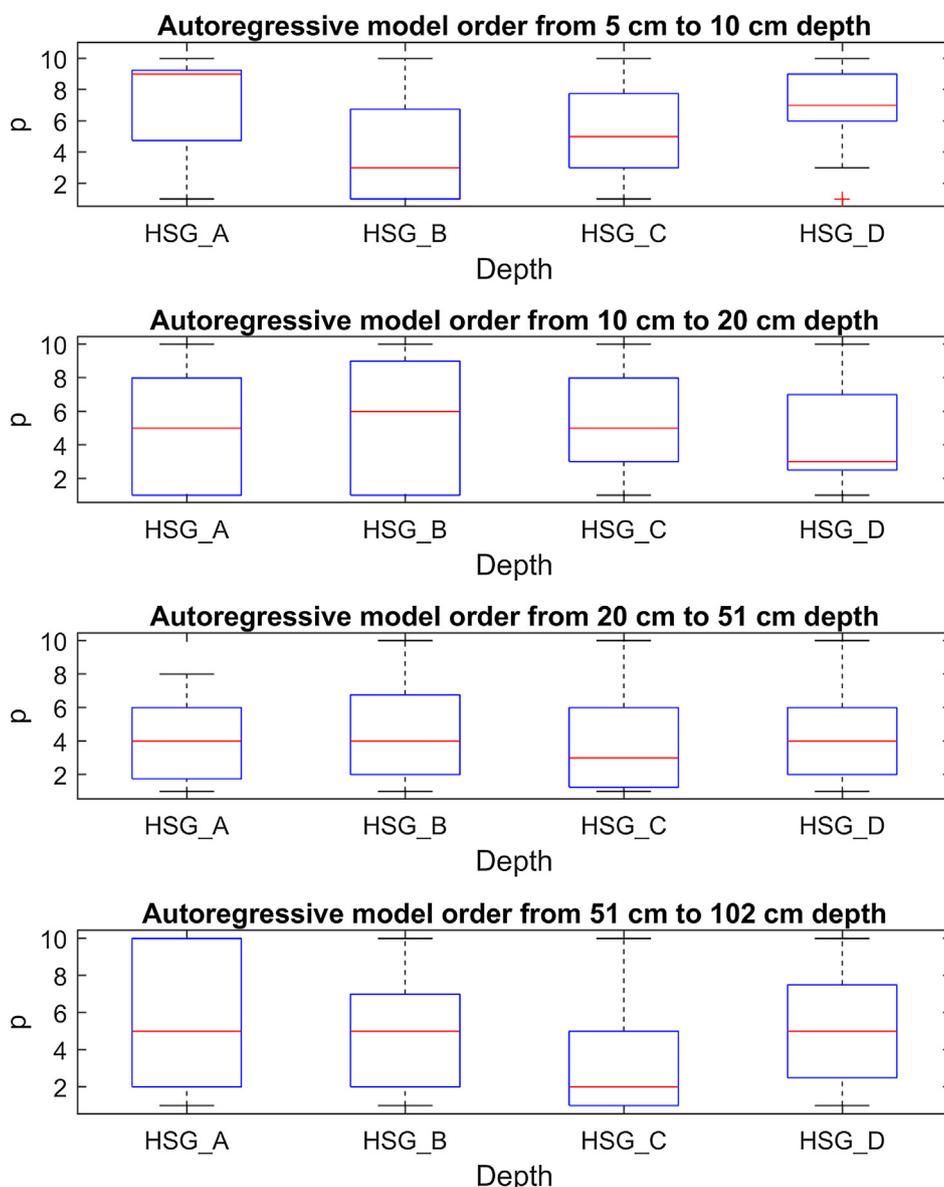


Fig. 3a. The boxplot of the memory orders for each HSG and each depth pair obtained from the observed SMC data of all the stations.

and known HSG. For this purpose, the *spatial validation stations* are used. The step-wise description of the spatial validation is described as follows,

Step-1: For a new location, firstly the HSG information is noted and the observed surface SMC values are transformed to standard normal variates using non-parametric kernel density approach (as explained in data transformation section, Steps 1–3).

Step-2: A Cumulative Distribution Function (CDF) of SMC values at each depth from all the *model development stations* within a particular HSG is generated using non-parametric kernel density approach (as explained in data transformation section, Steps 1–2). This is denoted as ‘reference CDF’. Since at ungauged stations, the SMC data at deeper layers is not available, the reference CDF is assumed to be valid for the same depth within the same HSG.

Step-3: According to the HSG of the new station, the SSMP model is applied and the estimates values are obtained. These are back transformed using standard normal distribution (Step 5 in data transformation section) and then using the reference CDF (for corresponding depth and HSG), which is similar to the method explained in Step 6 in data transformation section (Fig. 2).

Step-4: The back transformed values for all the depths are finally

corrected by adding the Deviation in Mean (DM) between the soil moisture regimes of the new locations and the *model development stations*. The DM is computed by taking the difference between the mean of the surface SMC values of the *model development stations* (for the particular HSG) and the target station. This is due the fact that the soil moisture regimes of the new station may differ from the soil moisture regime of the *model development stations*. The estimated SMC values, before and after DM correction, are compared to the observed SMC values.

4. Results and discussions

4.1. Model orders and coefficients

To determine the model orders, the SSMP models for individual stations with the corresponding observed SMC data from each HSG are developed. During model development at each station, the ranges of p and q are varied between 1 and 10 from which a parsimonious choice of the model order is identified using three criteria i.e., MF, MSE and FPE. Figs. 3a and 3b shows the box plots of orders of memory and forcing of all the four depth pairs of these individual models. The median values

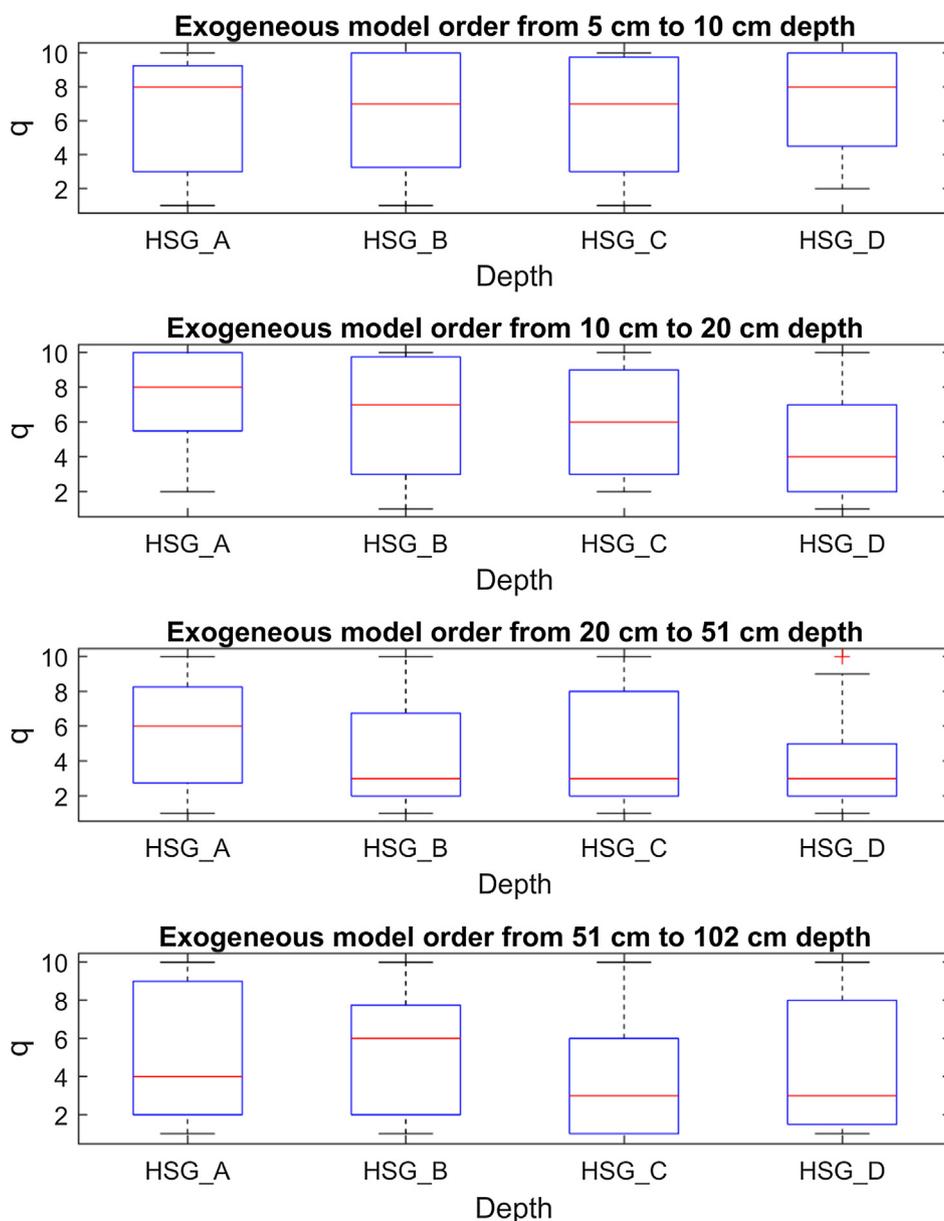


Fig. 3b. The boxplot of the forcing orders for each HSG and each depth pair obtained from the observed SMC data of all the stations.

Table 2

Memory and forcing orders of the models for HSG A, B, C and D for each depth pair.

| D_1 | D_2 | HSG A | | HSG B | | HSG C | | HSG D | |
|-------|-------|-------|---|-------|---|-------|---|-------|---|
| | | p | q | p | q | p | q | p | q |
| 5 | 10 | 9 | 8 | 3 | 7 | 5 | 7 | 7 | 8 |
| 10 | 20 | 6 | 8 | 6 | 7 | 5 | 6 | 3 | 4 |
| 20 | 51 | 4 | 6 | 4 | 3 | 3 | 3 | 4 | 3 |
| 51 | 102 | 7 | 4 | 5 | 6 | 2 | 3 | 5 | 3 |

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

of the model orders observed, in this step from each station corresponding to each HSG and each depth pair are finally assigned as the memory and forcing orders for that particular HSG. The selected values of the model orders for all the depth pairs for each HSG are shown in Table 2. With these selected model orders, data from the model development stations are used to determine the model coefficients. The obtained coefficients are averaged across all the stations within a HSG.

The mean values obtained for each depth pairs for a HSG are specified as the set of final coefficients of the SSMP model. These values of the coefficients are shown in Table 3.

A close inspection of Tables 2 and 3 reveals a few points. The total number of model orders decreases from HSG A to HSG D, i.e. with the decreasing infiltration capacity of the HSG. It indicates that the total number of model orders decreases with the decrease in infiltration capacity of the HSGs. A few more detailed observations especially the values of the model coefficients indicate some physical interpretation. Firstly, it can be noted that the memory coefficients tend to be low to high from HSG A to HSG D in general, indicating the dominating role of memory at deeper layers for the soils having low infiltration rate (e.g. HSG D). On contrast, the forcing coefficients are distinctly noted to decrease from HSG A to HSG D for all the four depth pairs except a few aberrations. It indicates that the forcing varies directly with the infiltration rate of the soil layer as the infiltration rate decreases from HSG A to HSG D. Hence, it can be expected that the values of forcing coefficients for a particular depth accurately capture the infiltration trend across the four HSGs. It is also noted that first few coefficients are much higher than the rest. However, these are not discarded because

Table 3

The memory ($a_1 - a_{10}$) and forcing ($b_1 - b_{10}$) model coefficients developed corresponding to their assigned model orders for HSG A, B, C and D for each depth pair.

| Coefficients | 5_10 | | | | 10_20 | | | | 20_51 | | | | 51_102 | | | |
|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | A | B | C | D | A | B | C | D | A | B | C | D | A | B | C | D |
| a_1 | 0.887 | 0.908 | 0.965 | 0.924 | 0.877 | 0.946 | 1.014 | 0.993 | 0.939 | 0.972 | 1.064 | 1.094 | 0.995 | 0.953 | 1.067 | 1.049 |
| a_2 | -0.090 | -0.089 | -0.152 | -0.134 | -0.065 | -0.127 | -0.177 | -0.162 | -0.114 | -0.078 | -0.152 | -0.219 | -0.111 | -0.123 | -0.121 | -0.143 |
| a_3 | 0.042 | 0.092 | 0.072 | 0.067 | 0.026 | 0.071 | 0.056 | 0.090 | 0.044 | 0.035 | 0.043 | 0.082 | 0.059 | 0.077 | - | 0.058 |
| a_4 | 0.015 | - | 0.010 | -0.011 | 0.016 | -0.010 | 0.014 | - | 0.046 | -0.001 | - | -0.014 | -0.013 | 0.013 | - | -0.022 |
| a_5 | 0.029 | - | 0.039 | 0.030 | 0.021 | 0.027 | 0.033 | - | - | - | - | - | 0.009 | 0.021 | - | 0.023 |
| a_6 | 0.010 | - | - | 0.011 | 0.039 | 0.022 | - | - | - | - | - | - | 0.006 | - | - | - |
| a_7 | 0.023 | - | - | 0.033 | - | - | - | - | - | - | - | - | -0.003 | - | - | - |
| a_8 | -0.008 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| a_9 | 0.011 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| a_{10} | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| b_0 | 0.736 | 0.689 | 0.611 | 0.698 | 0.663 | 0.614 | 0.589 | 0.580 | 0.525 | 0.431 | 0.401 | 0.368 | 0.269 | 0.397 | 0.291 | 0.262 |
| b_1 | -0.533 | -0.468 | -0.450 | -0.513 | -0.415 | -0.424 | -0.463 | -0.433 | -0.283 | -0.267 | -0.269 | -0.245 | -0.093 | -0.283 | -0.170 | -0.178 |
| b_2 | -0.028 | -0.054 | -0.010 | -0.010 | -0.072 | -0.038 | 0.004 | -0.014 | -0.090 | -0.091 | -0.087 | -0.066 | -0.083 | 0.001 | -0.071 | -0.050 |
| b_3 | -0.003 | -0.058 | -0.026 | -0.019 | 0.006 | -0.022 | -0.011 | -0.054 | -0.006 | - | - | - | -0.036 | -0.018 | - | - |
| b_4 | -0.029 | -0.011 | -0.022 | -0.010 | -0.024 | -0.017 | -0.023 | - | -0.039 | - | - | - | - | -0.011 | - | - |
| b_5 | -0.023 | -0.003 | -0.027 | -0.020 | -0.029 | -0.009 | -0.036 | - | -0.017 | - | - | - | - | -0.029 | - | - |
| b_6 | -0.009 | -0.004 | -0.010 | -0.018 | -0.018 | -0.030 | - | - | - | - | - | - | - | - | - | - |
| b_7 | -0.030 | - | - | -0.030 | -0.014 | - | - | - | - | - | - | - | - | - | - | - |
| b_8 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| b_9 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |

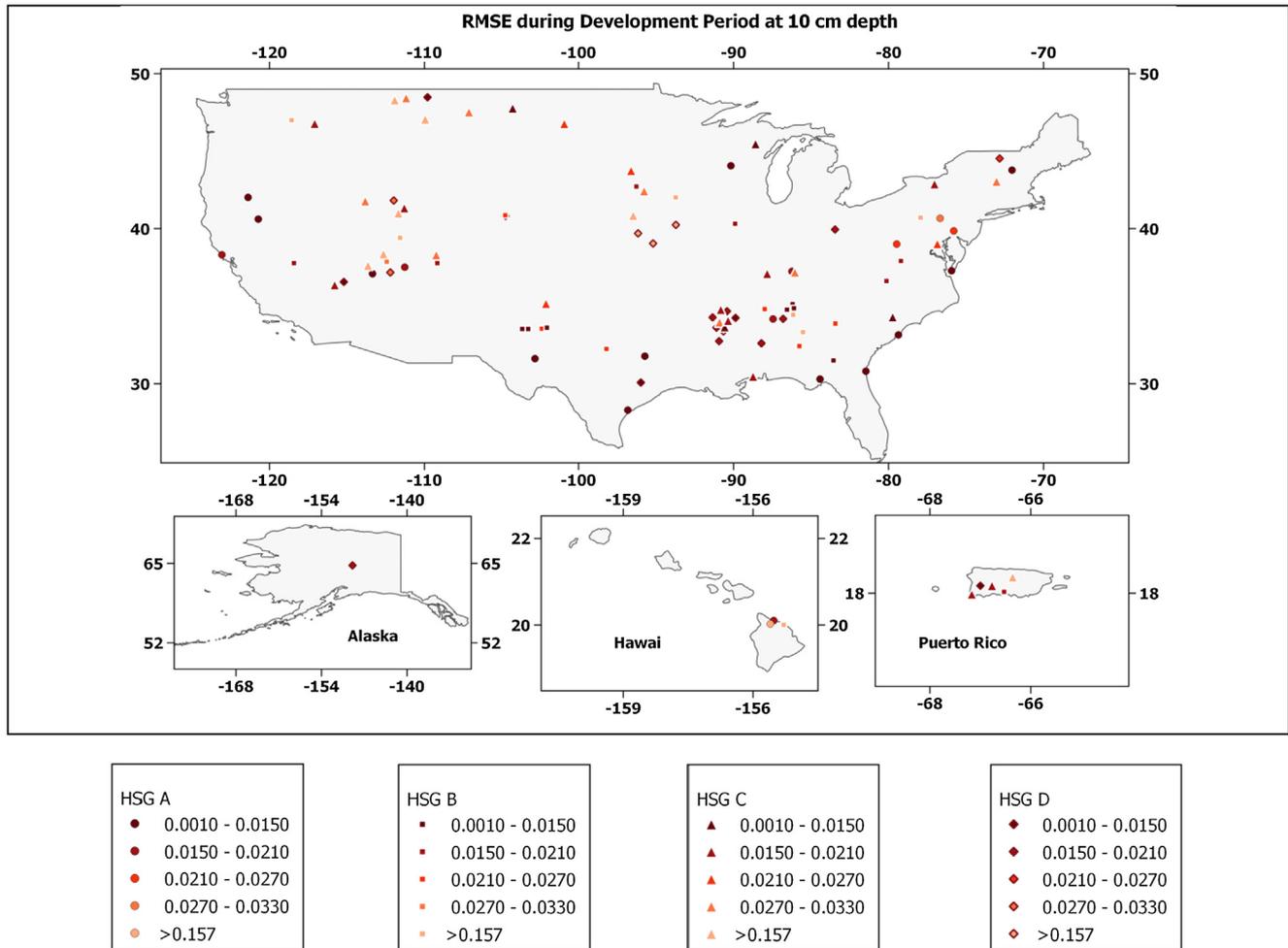


Fig. 4a. Spatial variability of RMSE across the stations of HSG A, B, C and D at 10 cm depth during development period.

the model fitting criteria were already fixed and, whatever small, these coefficients may have some contribution to the model performance.

Moreover, it is observed that the memory coefficient of immediate previous time step (a_1) increases, in general, with the increase in depth

for all the four HSGs. Inversely, the forcing coefficients of same and immediate previous time steps (and b_1) decrease with the increase in depth for all the HSGs. It is established in the literature that the memory dominates over the forcing component at the deeper layers whereas the

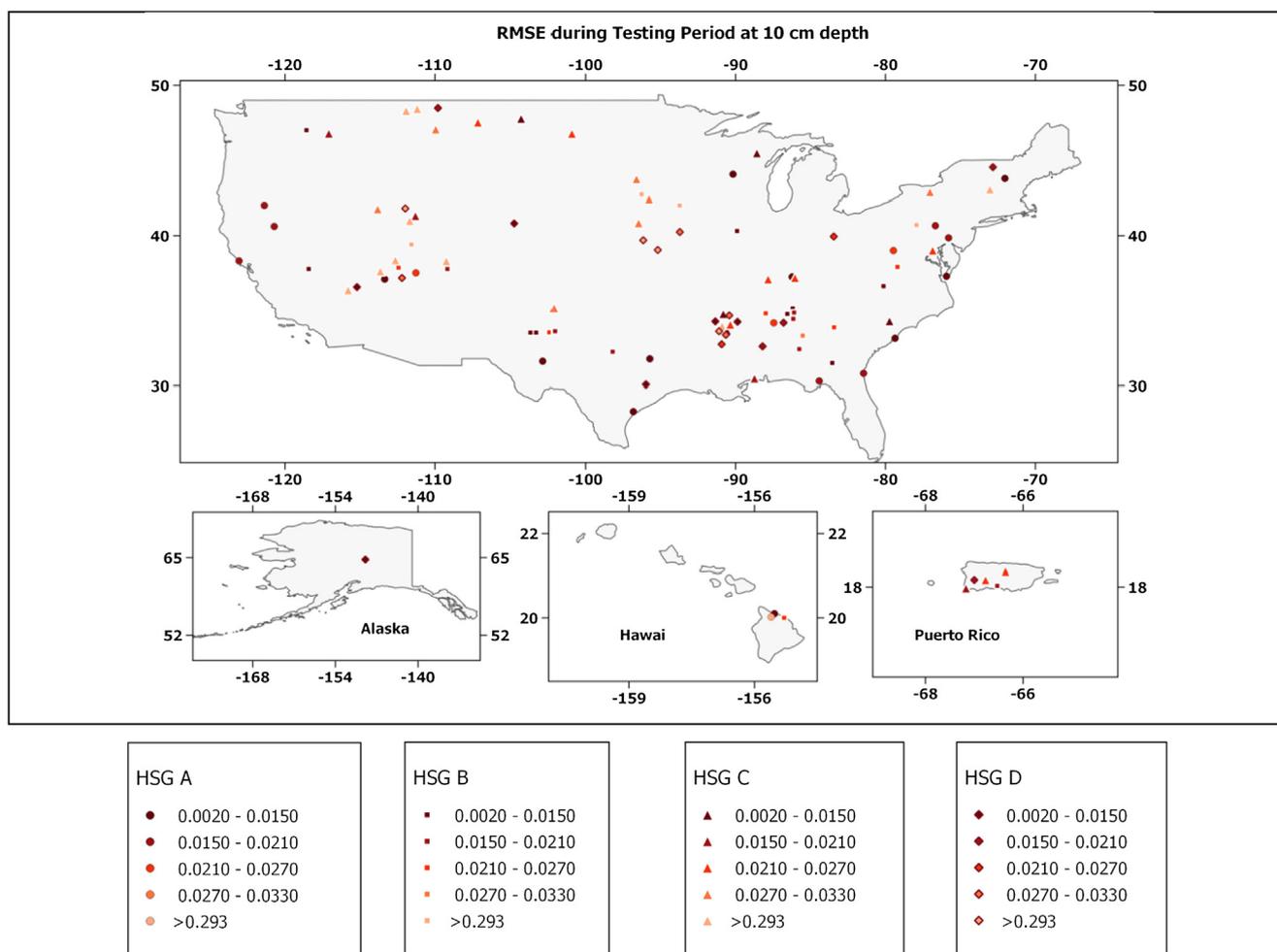


Fig. 4b. Spatial variability of RMSE across the stations of HSG A, B, C and D at 10 cm depth during testing period.

forcing dominates over memory components at near-surface layers (Pal et al., 2016; Mahmood et al., 2012). This hypothesis is reinforced by the estimated coefficients of the proposed spatially-varying SSMP model.

4.2. Model performance

Four different performance metrics are used in this study, namely, Correlation of Coefficient (CC), Refined Degree of Agreement (Dr), Root Mean Square Error (RMSE) and unbiased Root Mean Square Error (uRMSE). The performance of the SSMP model is evaluated at each station (SSMP performance). This performance is compared with the performance of the model developed individually at each station (individual performance). The SSMP performance is expected to be bit inferior as compared to individual performance. The model performances are carried out in terms of the average values of the performance metrics across the stations for each depth pair of each HSG. The complete range of performance metrics i.e. the minimum, maximum and average values of the performance metrics for all four depth pairs of all HSGs for the two cases mentioned are represented in Tables S1 to S4 in the supplementary document. It has been observed for all the depth pairs of all four HSGs that the model performances are almost comparable between two cases, SSMP performance being a bit inferior for model development and testing periods. Further, it is noticed that for both the cases and for all four HSGs the model performances decrease with increase in depth. At some depths the model performance is better with the SSMP model for that particular HSG than the model developed with the observed data. This implies that the HSG driven spatial transferability of the SSMP model is beneficial as reflected

through the comparable SSMP and individual performance. For rest of the manuscript, model performance refers to SSMP performance.

The model performance at each station for all the HSGs at all four depths during development and testing periods is assessed through RMSE (Figs. 4a and 4b at 10 cm depth, for other depths refer to Figs. S1–S4 in the Supplementary document) and other performance metrics (not shown). In these figures, different tones are used to depict the variation of the model performance; the stations with deeper tones indicate even better performance (lower RMSE) as compared to lighter tones. Comparing the performance for all the depths, it is noticed that for each HSG more number of stations are showing better performance towards the near-surface layers compared to deeper layers.

A typical time-series plots of observed and estimated SMC using the SSMP model at one station from each HSG at all four depths are shown in Fig. 5 (for HSG A) and Figs. S5–S7 in the supplementary document (for HSG B to HSG D). From the figures, it is noticed that the model is able to capture almost all the peaks of SMC variation for the upper layers i.e. 10 and 20 cm depths for all four HSGs. However, for the deeper layers (i.e. 51 and 102 cm) higher deviation of estimated SMC from the observed SMC is noted.

The performance of SSMP model is also checked in terms of the variation of mean and standard deviation values of the estimated SMC values for each depth pair and each HSG which are shown in Fig. 6 (for HSG A) and Figs. S8–S10 in the supplementary document (for HSG B to HSG D). In these figures, the average value of these mean and the standard deviation values of observed and estimated SMC are shown by the white diamonds in the boxes. It has been investigated that these average values for observed and estimated SMC vary insignificantly

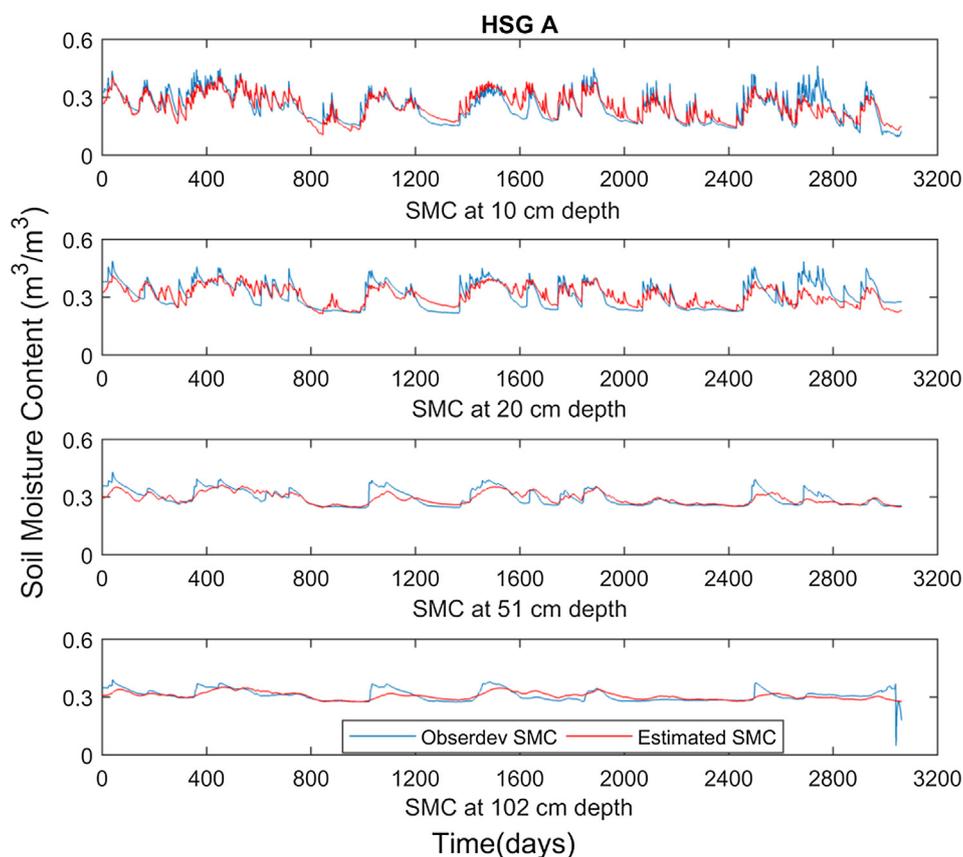


Fig. 5. Comparison of observed and estimated SMC for all four depths of one station from HSG A.

within the order of ~ 0.001 both during model development and testing periods indicating a good model performance.

4.3. Spatial validation

4.3.1. Reference Cumulative Distribution Function (CDF)

For spatial validation of the developed SSMP model, the reference CDFs for all the depths except the surface layer of all HSGs are prepared as discussed in methodology. The depth wise CDFs obtained from the SMC time series of all the *model development* stations are shown in Fig. 7 for HSG A and Figs. S11–S13 in the Supplementary document for HSG B to HSG D respectively. It is noted that the SMC values mostly vary from 0 to 0.65 for almost all the depths of all HSGs. This general tendency is violated only for two cases i.e. the SMC varies from 0 to 0.98 at 10 cm depth for HSG B and at 20 cm depth for HSG D. We assume that the mean soil moisture at a real location may change but the distribution shape will remain same.

4.3.2. Model performance

During the spatial validation, firstly, the developed SSMP model for each HSG is applied with the observed surface SMC (5 cm) at new locations to estimate the SMC at 10, 20, 51 and 102 cm depths. However, the systematic bias i.e. the difference between the mean of observed SMC and estimated SMC at deeper layers are noted to be high for each HSG. It can be attributed to the diverse regime of SMC values of the *model development* stations than the SMC values of *spatial validation* stations. Therefore, as discussed in the methodology of spatial validation, the regimes of observed surface SMC values for the *model development* and the target station are matched by computing the Deviation in Mean (DM) for the surface layer. The mean values of the surface SMC time series of all the model developing stations are found to be 0.141, 0.171, 0.198 and 0.239 for HSG A, B, C and D respectively. The DM for each station for a particular HSG is obtained by subtracting the mean

SMC of surface layer of the target station from the mean value of the surface layer from the *model development* stations for that particular HSG. The DM is added to the estimated SMC values for the deeper layers of the target stations to obtain the final DM-corrected SMC of the deeper layers. This helps to correct the model bias to some extent and, maintain the soil moisture regime for the target station. In this process, possible occurrences of negative SMC values, if any, are replaced by zero.

A typical example of the complete process of obtaining the DM-corrected SMC values at deeper layers for HSG A (similar for all four HSGs) can be described as following. The SSMP model for HSG A is obtained from Tables 2 and 3 respectively. For implementing the SSMP model for spatial validation, a new station belonging to HSG A is selected of which surface SMC information would be utilized. However, the SMC distribution information of the deeper layers of the newly selected station is unavailable which is required to back-transform the reduced variates to attain the corresponding SMC. The unavailability demands the use of the reference CDFs obtained from the model developing stations as shown in Fig. 7 and Figs. S13–S15 in the Supplementary document for each depth. After estimating the SMC at each layer using the SSMP model the DM correction is performed to attain the final estimates of the SMC values as discussed. The mean of surface SMC values of all *model development* stations from HSG A is 0.141 and the mean value of surface SMC of the target station (PauA-kala) is computed as 0.337. Subsequently, the DM i.e. the difference between the mean of surface layer SMC of *model development* stations and the target station is found to be (-0.1968) . Lastly, the DM-corrected final SMC values are obtained after subtracting the DM from the values estimated from SSMP model. Table S5 in the supplementary document shows the estimated SMC values of first 10 steps after removing the initial values (Pal et al., 2016).

To check whether the DM can be beneficially used in spatial transferability the following analysis is carried out. Systematic bias i.e.

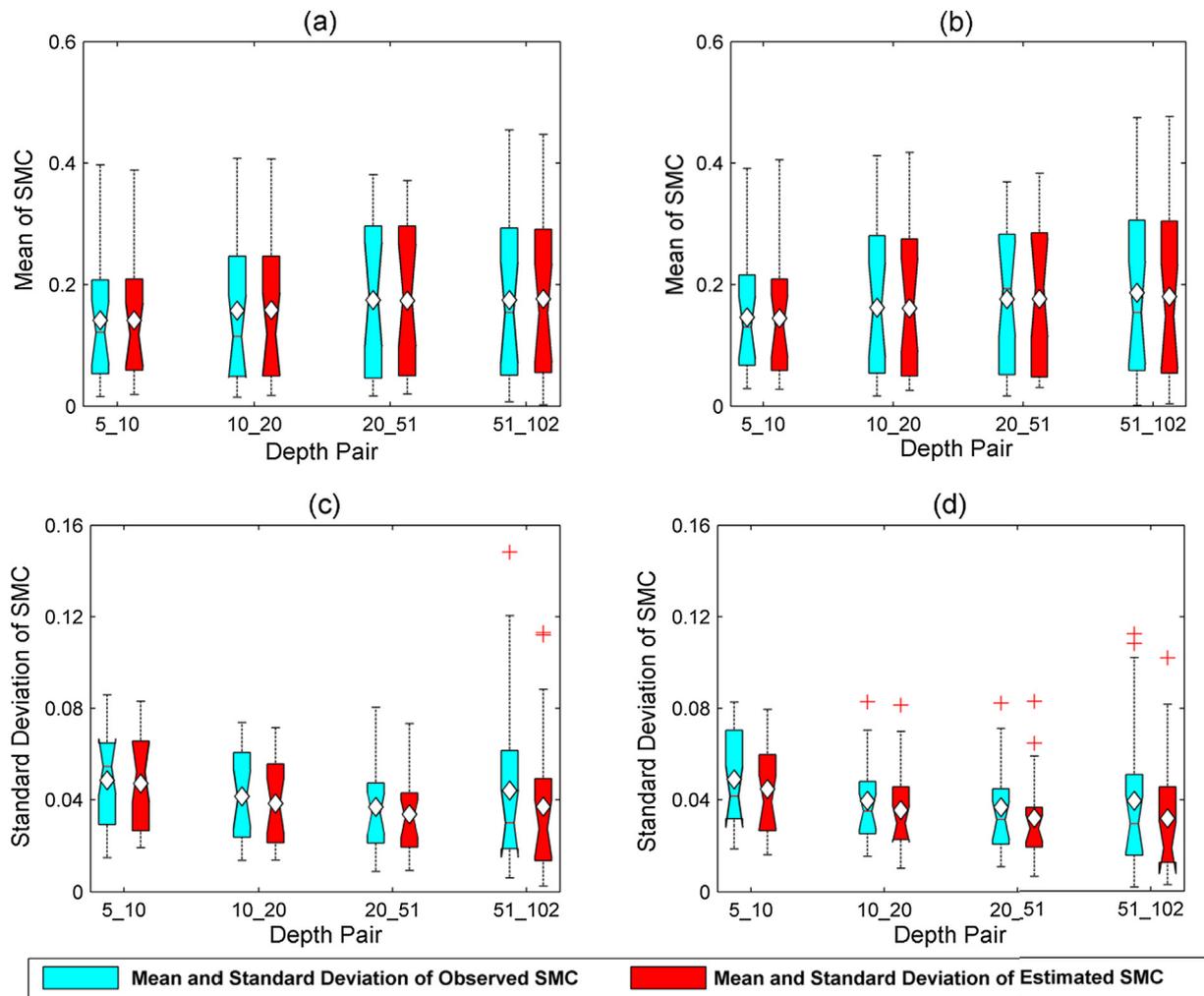


Fig. 6. The boxplots showing the comparison of observed and estimated SMC for HSG A. Comparison of mean values during (a) development period and (b) testing period. Comparison of standard deviation during (c) development period and (d) testing period. The average values of these mean and standard deviation across all stations are shown by the white diamond in the plot.

the difference in the mean of estimated SMC and the mean of observed SMC for each deeper layer for the target station is computed. It has been noticed that the DM successfully indicates the biases (both magnitude and direction) for the deeper layers for almost all the stations since in almost all the cases the biases vary within 10% of the surface DM. However, there are some stations for each depth from each HSG which do not follow the trend and number of such stations is shown in Table 4. Exception of this fact for some cases as seen in Table 4 can be ignored keeping the possibility of uncertainty in the data in mind.

The comparison of the model performances during spatial validation for the two cases i.e. i) Case 1: the model performance without the DM-corrected estimated SMC for deeper layers; and ii) Case 2: the model performance with the DM-corrected values, are presented to show the requirement of DM correction in Tables S6–S9 in the supplementary document. It clearly shows that the superiority of the model performance during the second case i.e. after DM correction. The performance metrics values i.e. CC, Dr, RMSE and uRMSE shows that the differences of CC of all the four depths from all four HSGs, between the two cases range from 0.006 to 0.012 which is quite low. However, the values of Dr, RMSE and uRMSE of all the four depths from all four HSGs are observed to be significantly improved in the second case where DM-corrected values are used. Fig. 8 represents a time series plot of observed, estimated SMC values with and without DM correction for all four depths of a station from HSG A where the improvement of SMC estimation after DM correction at each depth is evident. Other such

plots from remaining HSGs are also prepared and similar outcome is observed. Hence, in the manuscript only one such example is shown to avoid repetition. Therefore, the comparative analyses discussed above suggests the DM correction of the SSMP model estimated values is essential for ungauged locations and hence is the final step of spatial validation.

The performances of the SSMP model for each HSG during spatial validation are also shown by the comparison of the mean and the standard deviation values of each station for all HSGs of the observed and estimated SMC in Figs. 9a and 9b respectively for each depth pair. The difference in average values of the mean and standard deviation of observed SMC and DM-corrected estimated SMC for all four HSGs vary from 0.02 to 0.077, and 0.022–0.085 respectively for all four depth pairs. Hence, it can be concluded that the proposed statistical approach can almost correctly characterize the mean of SMC for all depths of all four HSGs although the characterization of standard deviation is comparatively poor. However, the values of performance metrics during spatial validation show a reasonably acceptable model performance given the possibility of uncertainty present due to various unaccounted physical controls of the ungauged location. Finally, the theory of using only the surface SMC information to estimate the vertical SMC profile for different HSGs in proposed spatially varying SSMP model shows the efficient applicability to the locations where the SMC information for deeper layers is not available.

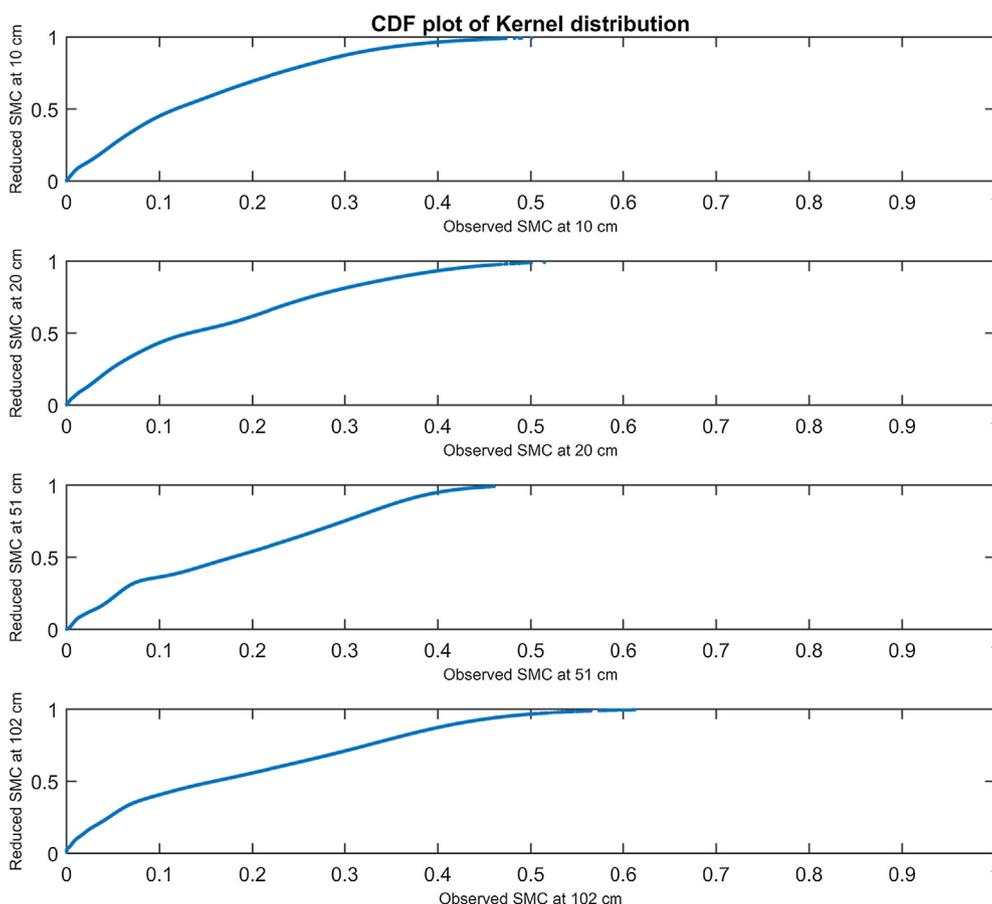


Fig. 7. Reference CDF plots of SMC data using kernel distribution at 10, 20, 51 and 102 cm depths from the *model development* stations for spatial validation of HSG A.

Table 4
Number of station showing greater than 10% variation from the surface DM for each HSG.

| HSG | No. of stations having > 10% bias than DM in surface layer | Depth Pair | | | |
|-----|--|-----------------------|-------|-------|--------|
| | | Total no. of stations | | | |
| | | 5_10 | 10_20 | 20_51 | 51_102 |
| A | 17 | 1 | 2 | 4 | 4 |
| B | 22 | 0 | 0 | 3 | 7 |
| C | 17 | 2 | 1 | 2 | 6 |

5. Summary and conclusions

In this study, a spatially-varying Statistical Soil Moisture Profile (SSMP) model is developed. The key features of the SSMP model are – 1) estimating the vertical SMC profile using only the surface SMC; and 2) imparting the spatial transferability by incorporating the HSG information. The SMC data are obtained from 171 monitoring stations from SCAN, USCRN and SNOTEL network scattered over entire USA from ISMN database at 5, 10, 20, 51 and 102 cm depths and the HSG information of each monitoring stations are determined from the WSS.

During the model development the forcing components show the trend of decreasing in the direction of HSG A to HSG D i.e. the forcing coefficients are higher for high infiltration (HSG A) and low for low infiltration (HSG D) of the soil. This specific feature of the forcing components for different HSGs having different infiltration characteristics including the effects of the of soil hydraulic properties on the SMC dynamics, justifies the applicability of the spatially varying SSMP model for each HSG group to new locations.

The potential of spatially-varying SSMP model developed are investigated by comparison of the model performances of the station-wise observed-data-specific models and the proposed spatially varying models. The efficacy of the developed spatially varying SSMP model in terms of spatial transferability is also investigated by applying the models to new monitoring stations of each HSG.

The difference in SMC regimes of the *model development* stations and the target stations drives the study to compute the Deviation in Mean (DM) of the model developing stations and stations selected for spatial validation, for surface layers. The DM-corrected estimated SMC values of deeper layers obtained during the spatial validation gives better model performances for all four HSGs. For, both the cases (model development and spatial validation), the model performances illustrated, indicate that the developed spatially varying SSMP model are able to characterize the SMC at deeper layers from only the surface SMC information for all HSG and all depths. However, it has also been studied that the model performance consistently decreases with increase in depths for all the four HSGs but still acceptable given the complexity of the model.

Considering the key features of the developed model, future scope lies in the integration of remotely sensed surface soil moisture content (0–5 cm) in the estimation of large scale fine resolution, vertical soil moisture profile (up to root zone). It is expected to be useful information in several fields of applications.

Declaration of interests

None.

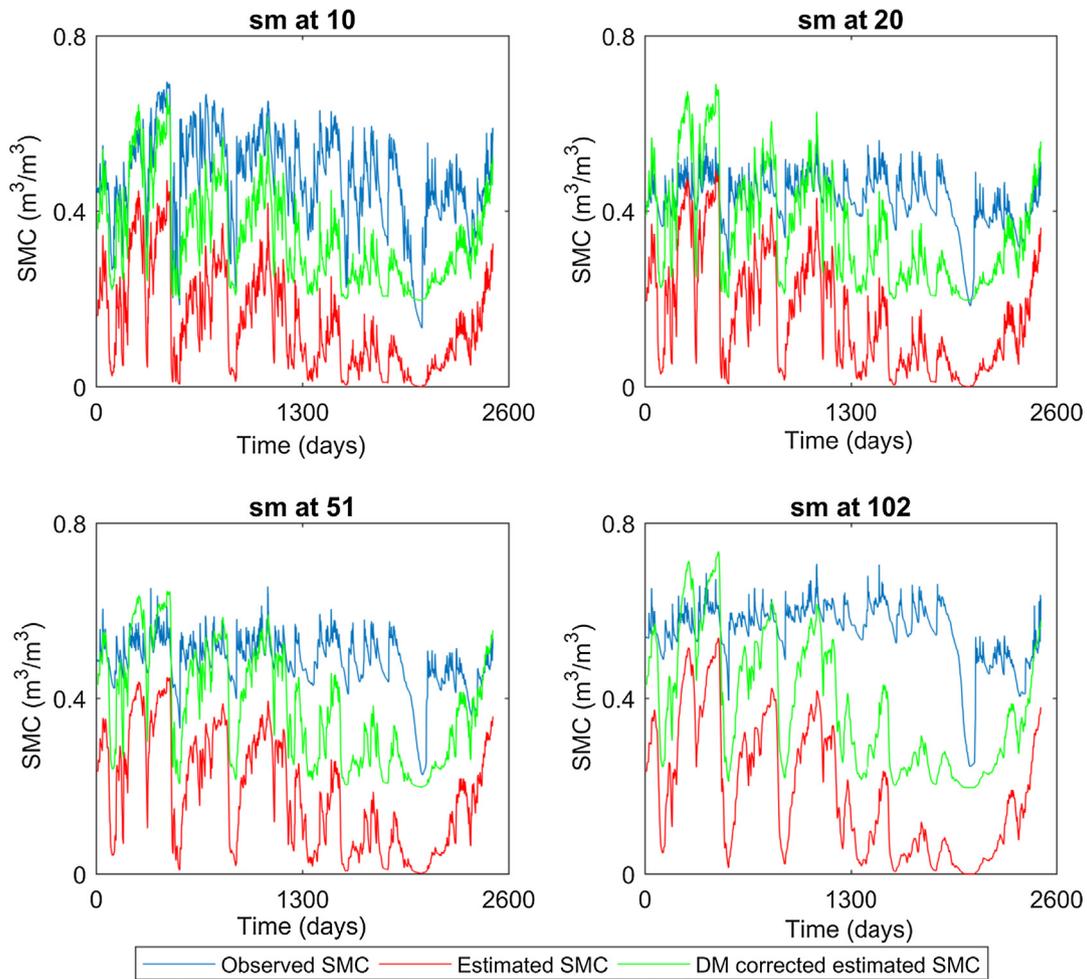


Fig. 8. Time series plot of observed, estimated and DM-corrected estimated SMC of all four depths for one station selected for spatial validation from HSG A.

Acknowledgements

This study is partially supported by a research project sponsored by the Space Application Centre (SAC), Indian Space Research

Organization (ISRO), Govt. of India (Ref No.: IIT/SRIC/CE/VIR/2016-17/88).

Appendix A. Parameter estimation of the SSMP model

A1. The orders of memory and forcing components (p and q)

The orders of the memory and forcing components are represented as *p* and *q* respectively. The optimum values of model order of each soil moisture time series are selected based on the desired model order identification criteria which are discussed in detail in this section. The model order identification criteria for SSMP model are – a) The prediction focus or the model fit (MF); b) The Mean Square Error (MSE) function; and c) The Akaike’s Final Prediction Error (FPE).

Higher values of the MF indicate better model performance. On the other hand, the lower values of MSE are preferable for better model performance since it minimizes the variance. The MF and MSE can be represented by the following equations,

$$MF = 100 \left(1 - \frac{\sqrt{\sum_{i=1}^n (SM_i^{sim} - SM_i^{obs})^2}}{\sqrt{\sum_{i=1}^n (SM_i^{sim} - \bar{SM})^2}} \right) \tag{A1}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (SM_i^{sim} - SM_i^{obs})^2 \tag{A2}$$

where SM_i^{sim} is the estimated transformed standard normal variates, SM_i^{obs} is the transformed standard normal variates from the reduced variates of observed soil moisture values using kernel distribution, \bar{SM} is the mean of the transformed standard normal variates from the reduced variates of observed soil moisture data using kernel distribution and *n* is the number of samples in the dataset.

The Akaike’s Final Prediction Error (FPE) can be estimated by the following equation,

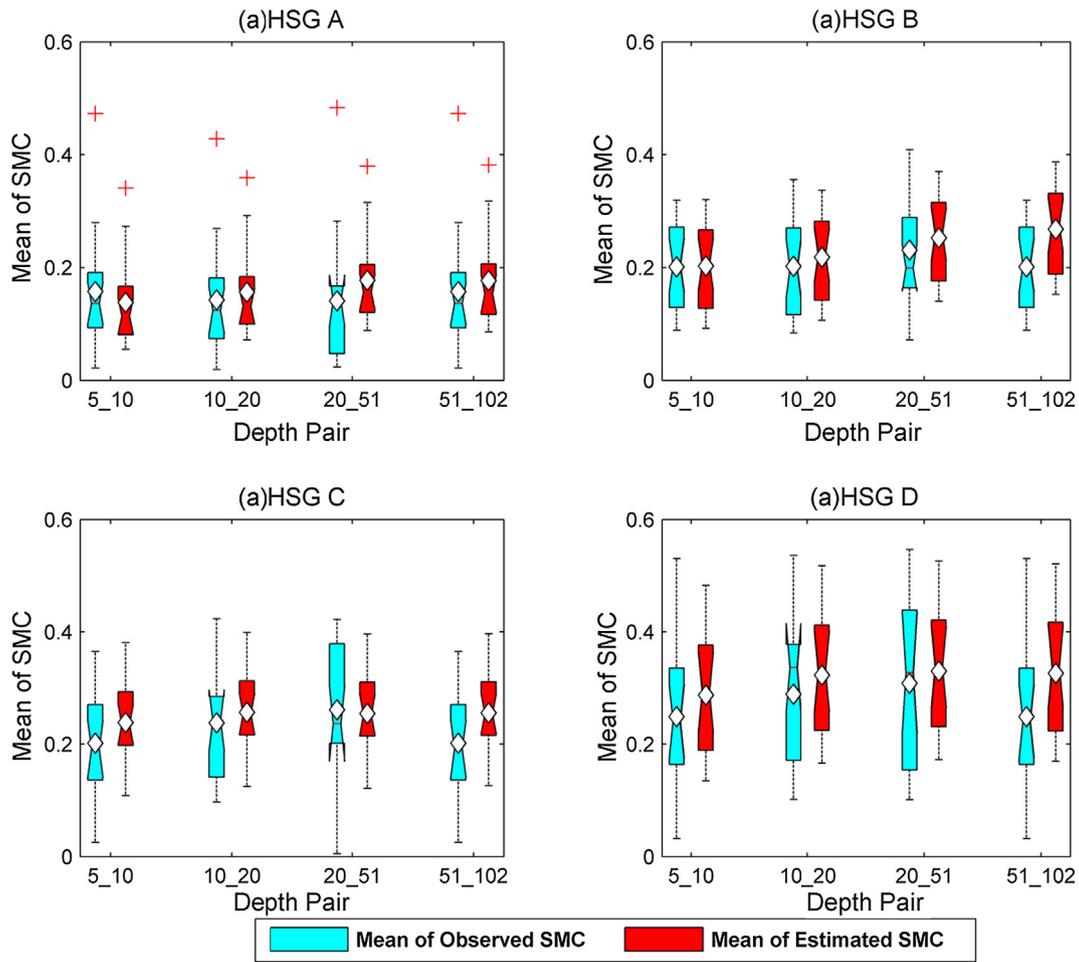


Fig. 9a. The comparison of mean observed SMC and estimated SMC values (after DM-correction) for all the depths during the spatial validation for a) HSG A; b) HSG B; c) HSG C; and d) HSG D. The average values of mean values across all stations are shown by the white diamond in the plot.

$$FPE = V \left(\frac{1 + \frac{m}{n}}{1 - \frac{m}{n}} \right) \tag{A3}$$

where m equals to $(p + q)$ representing the number of estimated parameters n is the number of values in the estimated dataset and V is loss function which is the determinant of the obtained noise covariance matrix and can be obtained from the following equation,

$$V = \det \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \varepsilon_i^T \right) \tag{A4}$$

where ε_i is the error at i th time step. In the simulation when $m \ll n$, the FPE is computed with the following equation,

$$FPE = V \left(1 + \frac{2m}{n} \right) \tag{A5}$$

The FPE criterion is a comparative measure of model performance where the smallest value FPE indicates the most precise model while assessed on different testing data set. The ranges of p and q are investigated over a range considering the model parsimony and to ensure the desired values of the aforementioned model order identification criteria. The next step of the SSMP model development is to determine the model coefficients. The methodology to estimate the model coefficients is described in the following section.

A2. Model coefficients

The coefficients of the SSMP model are estimated by the least square method which minimizes the summation of the square of the residuals. Therefore, from Eq. (3) the summation of the square of the error terms, i.e. $e(t)$, $t = \lambda, \dots, n$, can be written as,

$$S = \sum_{t=\lambda}^n \left[\left\{ SM_k(t) - a_1 SM_k(t-1) - \dots - a_p SM_k(t-p) - b_0 ks_SM_{k-1}(t) - \dots - b_{q-1} SM_{k-1}(t-q+1) \right\}^2 \right] \tag{A6}$$

In Eq. (A6), t varies from λ to n , where $\lambda = \max(p + 1, q)$ in order to avoid initial few steps to accommodate the lags. Finally, the coefficients are estimated by differentiating S with respect to each parameter and assigning the differentiated values equal to zero which is shown as,

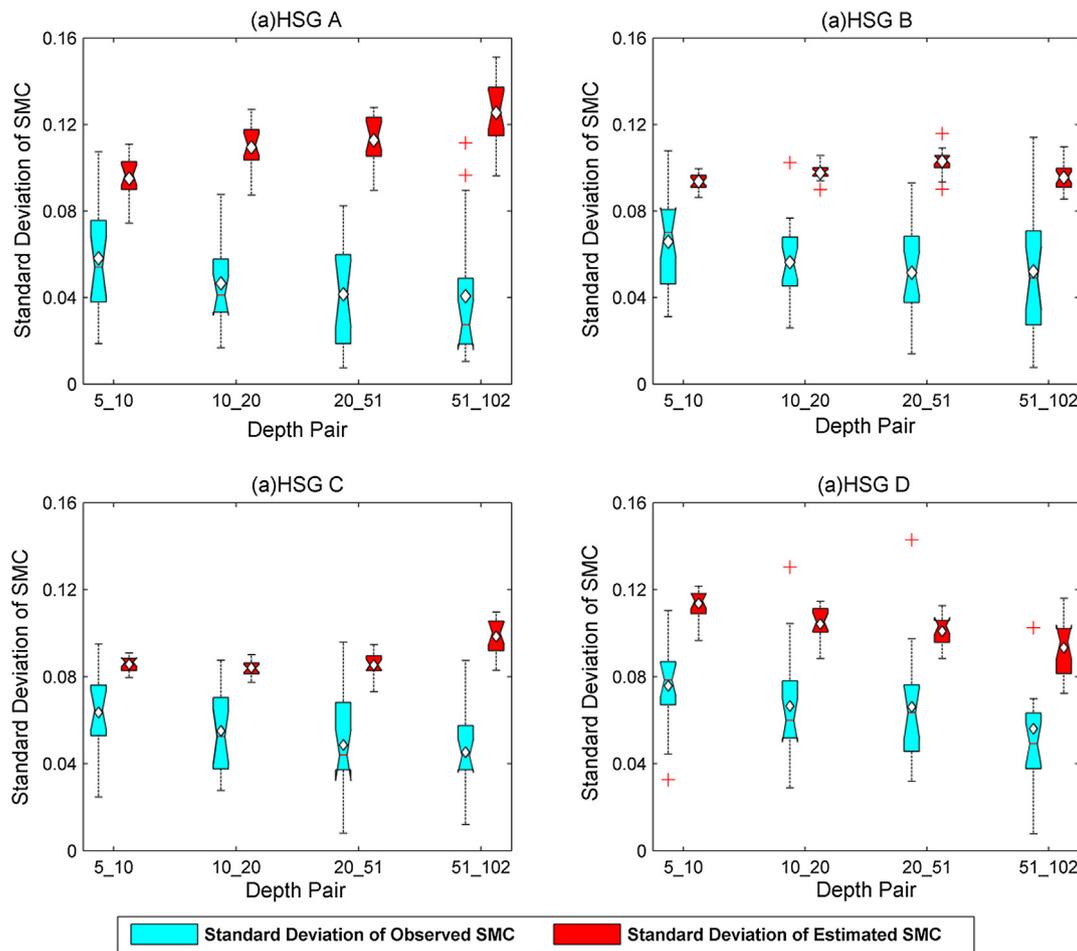


Fig. 9b. The comparison between standard deviation of observed SMC and estimated SMC (after DM-correction) values for all the depths during the spatial validation for a) HSG A; b) HSG B; c) HSG C; and d) HSG D. The average values of standard deviations across all stations are shown by the white diamond in the plot.

$$\left. \begin{aligned} \frac{\partial S}{\partial a_i} &= 0 \quad \forall i = 1, \dots, p \\ \frac{\partial S}{\partial b_j} &= 0 \quad \forall j = 0, \dots, (q - 1) \end{aligned} \right\} \tag{A7}$$

For solving the values of the coefficients (a_i and b_j) in Eq. (A7), the matrix notations are used and is represented as,

$$B = (X^T X)^{-1} X^T Y \tag{A8}$$

where B is the column matrix consisting of the coefficients (a_i and b_j); X is the matrix consisting of input variables, i.e. soil moisture values of target depth at previous time steps equal to values of memory orders and forcing depths at same and previous time steps equal to values of forcing orders and Y is the output variable matrix i.e. the soil moisture at target depth.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2018.12.042>.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Pignatelli, B., Martin, E., 2008. From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci.* 12, 1323–1337. <https://doi.org/10.5194/hess-12-1323-2008>.

Bertoldi, G., Della Chiesa, S., Notarnicola, C., Pasolli, L., Niedrist, G., Tappeiner, U., 2014. Estimation of soil moisture patterns in mountain grasslands by means of SAR RADARSAT2 images and hydrological modeling. *J. Hydrol.* 516, 245–257.

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Calvet, J.C., Noilhan, J., 2000. From near-surface to root-zone soil moisture using year-round data. *J. Hydrometeorol.* 1 (5), 393–411.

Cerda, A., 1996. Soil aggregate stability in three Mediterranean environments. *Soil Technol.* 9 (3), 133–140.

Choi, M., Jacobs, J.M., 2007. Soil moisture variability of root zone profiles within SMEX02 remote sensing footprints. *Adv. Water Resour.* 30, 883–896. <https://doi.org/10.1016/j.advwatres.2006.07.007>.

Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resour. Res.* 20, 682–690. <https://doi.org/10.1029/WR020i006p0682>.

Das, S.K., Maity, R., 2015. A hydrometeorological approach for probabilistic simulation of monthly soil moisture under bare and crop land conditions. *Water Resour. Res. Am. Geophys. Union (AGU)* 51, 2336–2355. <https://doi.org/10.1002/2014WR016043>.

Dorigo, W.A., 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. <https://doi.org/10.5194/hess-15-1675-2011>.

Dumedah, G., Walkera, J.P., Merlin, O., 2015. Root-zone soil moisture estimation from

- assimilation of downscaled Soil Moisture and Ocean Salinity data. *Adv. Water Resour.* 84, 14–22.
- Famiglietti, J.S., Rudnicki, J.W., Rodell, M., 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. *J. Hydrol.* 210, 259–281. [https://doi.org/10.1016/S0022-1694\(98\)00187-5](https://doi.org/10.1016/S0022-1694(98)00187-5).
- Farres, P.J., 1987. The dynamics of rainsplash erosion and the role of soil aggregate stability. *Catena* 14, 119–130.
- Gaur, N., Mohanty, B.P., 2013. Evolution of physical controls for soil moisture in humid and subhumid watersheds. *Water Resour. Res.* 49, 1244–1258. <https://doi.org/10.1002/wrcr.20069>.
- Ghannam, K., Nakai, T., Paschalis, A., Oishi, C.A., Kotani, A., Igarashi, Y., Kumagai, T., Katul, G.G., 2016. Persistence and memory timescales in root-zone soil moisture dynamics. *Water Resour. Res.* 52, 1427–1445. <https://doi.org/10.1002/2015WR017983>.
- Jawson, S.D., Niemann, J.D., 2007. Spatial patterns from EOF analysis of soil moisture at a large scale and their dependence on soil, landuse, and topographic properties. *Adv. Water Resour.* 30 (3), 366–381.
- Kerr, Y.H., Wadlteufel, P., Wigneron, J.P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C., Juglea, S.E., Drinkwater, M.R., Hahne, A., Martin-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, 318–328. <https://doi.org/10.1016/j.jhydrol.2009.06.025>.
- Kim, G., Barros, A.P., 2002. Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data. *Remote Sens. Environ.* 81 (2–3), 393–403.
- 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, 1–11. <https://doi.org/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, 353–363. <https://doi.org/10.1016/j.jhydrol.2011.01.012>.
- Li, H., Huang, M., Wigmosta, M.S., Ke, Y., Coleman, A.M., Leung, L.R., Wang, A., Ricciuto, D.M., 2011. Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed. *J. Geophys. Res. Atmos.* 116. <https://doi.org/10.1029/2011JD016276>.
- Mahmood, R., Littell, A., Hubbard, K.G., You, J., 2012. Observed data-based assessment of relationships among soil moisture at various depths, precipitation, and temperature. *Appl. Geogr.* 34, 255–264. <https://doi.org/10.1016/j.apgeog.2011.11.009>.
- Manfreda, S., Brocca, L., Moramarco, T., Melone, F., Sheffield, J., 2014. A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrol. Earth Syst. Sci.* 18, 1199–1212. <https://doi.org/10.5194/hess-18-1199-2014>.
- Musgrave, G.W., 1955. How much of the rain enters the soil? In: *Water: U.S. Department of Agriculture. Yearbook*. Washington, DC, pp. 151–159.
- Pal, M., Maity, Rajib, Dey, Sayan, 2016. Statistical modelling of vertical soil moisture profile: coupling of memory and forcing. *Water Resour. Manage., Springer* 30 (6), 1973–1986. <https://doi.org/10.1007/s11269-016-1263-4>.
- Price, K., Jackson, C.R., Parker, A.J., 2010. Variation of surficial soil hydraulic properties across land uses in the southern Blue Ridge Mountains, North Carolina, USA. *J. Hydrol.* 383, 256–268. <https://doi.org/10.1016/j.jhydrol.2009.12.041>.
- Rawls, W.J., Ahuja, L.R., Brakensiek, D., Shirmohammadi, A., 1993. Infiltration and soil water movement. In: Maidment, D. (Ed.), *Handbook of Hydrology*. McGraw-Hill, New York, pp. 5.1–5.51.
- Renzullo, L.J., van Dijk, A.I.J.M., Perraud, J.M., Collins, D., Henderson, B., Jin, H., Smith, A.B., McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *J. Hydrol.* 519 (2014), 2747–2762.
- Sabater, J.M., Jarlan, L., Calvet, J.-C., Bouysse, F., De Rosnay, P., 2007. From near surface to root zone soil moisture using different assimilation techniques. *J. Hydrometeorol.* 8, 94–206.
- Seneviratne, S.I., Koster, R.D., Guo, Z., Dirmeyer, P.A., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Mocko, D., Oleson, K.W., Versegny, D., 2006. Soil moisture memory in AGCM simulations: analysis of global land-atmosphere coupling experiment (GLACE) data. *J. Hydrometeorol.* 7 (5), 1090–1112.
- Singh, V.P., 2010. Entropy theory for movement of moisture in soils. *Water Resour. Res.* 46, 1–12. <https://doi.org/10.1029/2009WR008288>.
- Sivakumar, B., 2017. Applications to other hydrologic data. In: *Chaos in Hydrology*. Springer, Dordrecht, pp. 297–319. https://doi.org/10.1007/978-90-481-2552-4_11.
- Sörensson, A.A., Menéndez, C.G., 2011. Summer soil-precipitation coupling in South America. *Tellus* 63A, 56–68. <https://doi.org/10.1111/j.1600-0870.2010.00468.x>.
- Sridhar, V., Elliott, R.L., Chen, F., Brotzge, J.A., 2002. Validation of the NOAA-OSU land surface model using surface flux measurements in Oklahoma. *J. Geophys. Res.* 107 (D20), 4418. <https://doi.org/10.1029/2001JD001306>.
- USDA, N., 2009. National Engineering Handbook Chapter 7 Hydrologic Soil Groups.
- Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.* 70 (2), 191–207.
- Yan, H., DeChant, C.M., Moradkhani, H., 2015. Improving soil moisture profile prediction with the particle Filter-Markov Chain Monte Carlo Method. *IEEE Trans. Geosci. Remote Sens.* 53 (11). <https://doi.org/10.1109/TGRS.2015.2432067>.
- Yang, L., Wei, W., Chen, L., Jia, F., Mo, B., 2012. Spatial variations of shallow and deep soil moisture in the semi-arid Loess Plateau, China. *Hydrol. Earth Syst. Sci.* 16, 3199–3217. <https://doi.org/10.5194/hess-16-3199-2012>.
- Zambom, A., Dias, R., 2012. A review of kernel density estimation with applications to econometrics. *arXiv Prepr. arXiv1212.2812* 1–35.