Citation: Mahmood, T., Z. H. Xie, B. H. Jia, et al., 2019: A soil moisture data assimilation system for Pakistan using PODEn4DVar and the community land model version 4.5. *J. Meteor. Res.*, **33** (6), XXX－XXX, doi: 10.1007/s13351-019-9020-2. (in press)

A Soil Moisture Data Assimilation System for Pakistan using PODEn4DVar and the Community Land Model Version 4.5

Tariq MAHMOOD[1,2,3], Zhenghui XIE[1,2]*, Binghao JIA[1], Ammara HABIB[1,2], and Rashid MAHMOOD[4]

1 *State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*

2 *University of Chinese Academy of Sciences, Beijing* 100049, *China*

3 *Pakistan Meteorological Department, Islamabad, Pakistan*

4 *Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101, *China*

(Received March 18, 2019; in final form August 22, 2019)

Supported by the National Key R&D Program of China (2018YFC1506602), National Natural Science Foundation of China (41830967), and Key Research Program of Frontier Sciences, Chinese Academy of Sciences (QYZDY-SSW-DQC012).

* Corresponding author: zxie@lasg.iap.ac.cn.

ABSTRACT

Soil moisture is an important state variable for land-atmosphere interactions. It is a vital land surface variable for research on hydrology, agriculture, climate, and drought monitoring. In current study, a soil moisture data assimilation framework has been developed using the Community Land Model version 4.5 and the PODEn4DVar algorithm. Assimilation experiments were conducted at four agricultural sites in Pakistan by assimilating *in-situ* soil moisture observations. The results showed that it was a reliable system. To quantify further the feasibility of the data assimilation (DA) system, soil moisture observations from the top four soil-depths (0–5 cm, 5–10 cm, 10–20 cm, and 20–30 cm) were assimilated. The evaluation results indicated that the DA system improved soil moisture estimation. In addition, updating the soil moisture in the upper soil layers of CLM4.5 could improve soil moisture estimation in deeper soil layers (layer 7, L7~62.0 cm and layer 8, L8~103.8 cm). To further evaluate the DA system, observing system simulation experiments (OSSEs) were designed for Pakistan by assimilating daily observations. These idealized experiments produced statistical results that had higher correlation coefficients, reduced root mean square errors, and lower biases for assimilation, which showed that the DA system is reliable and improved soil moisture estimation in Pakistan.

Key words: PODEn4DVar, CLM4.5, data assimilation, soil moisture, Pakistan

1. Introduction

Soil moisture is an important land surface variable for climatological, hydrological, ecological, and biological studies and has a central role in land-atmosphere interactions. Koster et al.(2004) investigated that the soil moisture anomalies represent significant impacts on regional precipitation after undertaking elaborately designed numerical experiments. The land receives about 65 percent of the precipitation derived from evaporation over land, which is strongly linked to soil moisture (Chahine, 1992).

Accurate and precise information of soil moisture at both the spatial and temporal scales is vitally important when attempting to improve weather forecasts, climatic studies and for drought monitoring (Dai et al., 2004). However, the low number of soil moisture field measurements over land is a big barrier in acquiring the soil moisture knowledge on broad scales (Robock et al., 2000; Robinson et al., 2008; Crow et al., 2012; Zreda et al., 2012).To improve the soil moisture information, several efforts have been taken place without assimilating soil moisture observational data e.g. the North America Land Data Assimilation System (NLDAS) (Mitchell et al., 2004), the Global Land Data Assimilation System (GLDAS) (http://ldas.gsfc.nasa.gov), the Global Soil Wetness Project (http://grads.iges.org/gswp/) (Dirmeyer et al., 1999) and others (Qian et al., 2006; Sheffield and Wood, 2008).

Currently, regular and field observations, satellite observations, and hydrological modelling are the main sources used to acquire soil moisture information. Soil moisture information collected through field observations, which are of low temporal frequency and have few spatial points. As this information is on point based and therefore cannot showed the soil moisture spatial variations. These field and regular based soil moisture

have a great influence on the plants development, chemical activities of fertilizers and the generation of runoff and erosion. Therefore, it has significant impacts on agricultural and the environmental systems. The hydrological modelling is the other important source and the soil moisture simulations generated by hydrological models have good temporal frequency and spatial distributions. Though, the precision and accuracy of simulations is strongly linked to input data and model structure. Land data assimilation can provides the reasonable solution to all these issues. It is a technical method that incorporates the physical process data produced by the land surface models (Houser et al., 1998).

Recently, there has been progress on assimilation techniques, algorithm, and their applications for many fields like as land, marine, and atmospheric studies (Tian et al., 2011; Zhang et al., 2012). Tian et al. (2011) proposed a hybrid assimilation technique known as "Proper orthogonal decomposition (POD)-based ensemble four-dimensional variational assimilation method (PODEn4DVar)". This assimilation algorithm contains the benefits of both variational and ensemble techniques and performed better than both 4DVar and the EnKF methods under perfect and imperfect model cases. The computational cost is less when compared to the EnKF and therefore it can be reliably integrated into land data assimilation studies.

Land models play a fundamental role in land data assimilation systems. The Community Land Model (CLM) (Oleson et al., 2004; Olsen et al., 2010),which is the land module of the Community Earth System Model (CESM) (Hurrell et al., 2013). Even with the scientific improvements in CLM, some studies have shown that when simulating the hydrological state variables, CLM4.0 is biased towards estimating soil moisture at the global and regional scales (Long et al., 2013; Cai et al., 2014). In another study, CLM4.5

is used to assimilate AMSR_E soil moisture data and overestimation has been observed in soil moisture simulation at most part of the study area (Liu and Mishra, 2017). The earlier versions of CLM have been used in land DA studies for the improvement of soil moisture estimation. For example, CLM2.0 has been used as forecast operator in many land DA studies to improve estimation of soil moisture by assimilating *in-situ* soil moisture data (De Lannoy et al., 2007; Tian et al., 2008a; Zhang et al., 2012), and (Kumar et al., 2009) used synthetic observations. Shi et al. (2011) incorporated CLM3.0 as a forecast model in the DA framework and assimilated satellite data for the simulation of soil moisture. In another study, Sun et al. (2015) employed CLM3.5 to assimilate the GRACE data using the PODEn4DVar assimilation technique.

The aim of this study was to build an assimilation system using CLM4.5 with the PODEn4DVar algorithm to generate the improved and more accurate soil moisture estimation for Pakistan region as a case. Pakistan is now ranked among the top few in the list of environmentally vulnerable countries, and faces considerable human challenges because soil moisture changes have implications for health, agriculture, ecology, and water resources under climate change. In such a crucial scenario, reliable and more accurate information on atmospheric and hydrological parameters is needed so that more comprehensive research on weather and climate prediction, and hydrological and agricultural studies for the region can be undertaken. In this study, a new DA system was used to obtain preliminary analysis and evaluation results for farmlands across Pakistan through the assimilation of in-situ soil moisture observations. The evaluation experiments were conducted at four agricultural sites, which were representative of various agro-climatic zones in Pakistan. This meant that the DA system has been verified under

different hydrological conditions. The second goal of this study was to see the effects on deeper soil moisture prediction when soil moisture was assimilated into the upper soil layers, were also investigated.

2. Land data assimilation system for Pakistan

Land data assimilation system consists of forecast model, assimilation algorithm and observation operator. In current study, PODEn4DVar was selected as assimilation algorithm whereas CLM4.5 was used as forecasting model.

2.1 Land Surface Model CLM4.5

The CLM4.5, a global land surface model developed by the National Center for Atmospheric Research United states of America (NCAR).It is attached with the Community Earth System Model version 1.2 (CESM1.2) as a land module. It contains several modifications over previous versions such as improved parameterizations to reduce biases in soil carbon, revised photosynthesis, and canopy radiation schemes (Oleson et al., 2013).

In CLM4.5, land surface follows the subgrid hierarchy in which each grid cell consist of land units, columns, and plant functional types (PFTs). Grid cells may contain different numbers of land units, e.g. lake, glacier, vegetated and urban. The vegetated land units contain several columns, and each column has 15 layers for soil and five layers for snow, depending upon the snow depth. The soil moisture is calculated within top 10 hydrologically activated layers.

The volumetric soil moisture content is calculated by the following equation

$$\frac{\partial\theta}{\partial t} = -\frac{\partial q}{\partial z} - E - R_{fm} \tag{1}$$

Where θ volumetric soil moisture content in soil column, E is evaporation rate, q is vertical soil water flux, R_{fm} is melting or freezing point and z is vertical distance from surface.

2.2 POD-based ensemble four-dimensional variational assimilation method

Tian et al. (2008b) suggested a hybrid assimilation method using ensemble and Proper Orthogonal Decomposition (POD) techniques in which the adjoint model is not needed. Tian et al. (2011) used this technique to develop the PODEn4DVar method, which combined the benefits of the both ensemble and variational approaches. In this method, the analysis field can be obtained by minimizing the following cost function:

$$\mathrm{J}(\mathrm{x}') = \frac{1}{2}(\mathrm{x}')^{\mathrm{T}}\mathrm{B}^{-1}(\mathrm{x}') + \frac{1}{2}[\mathrm{y}'(\mathrm{x}') - \mathrm{y}'_{\mathrm{obs}}]^{\mathrm{T}}\mathrm{R}^{-1}[\mathrm{y}'(\mathrm{x}') - \mathrm{y}'_{\mathrm{obs}}]. \tag{2}$$

Where $\boldsymbol{B}$ and $\boldsymbol{R}$ represent the background and observation error covariance matrices, the superscript T indicates the transpose of matrix, and $x' = x - x_b$ shows the perturbation of the background vector $\mathrm{x_b}$ at t_0 .

$$\mathrm{y}'_{\mathrm{obs}} = \begin{bmatrix} \mathrm{y}'_{\mathrm{obs},1} \\ \mathrm{y}'_{\mathrm{obs},2} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{y}'_{\mathrm{obs},s} \end{bmatrix}. \tag{3}$$

and

$$y' = y'(x') = \begin{bmatrix} y'_1(x') \\ y'_2(x') \\ \cdot \\ \cdot \\ \cdot \\ y'_s(x') \end{bmatrix} = \begin{bmatrix} (y_1)' \\ (y_2)' \\ \cdot \\ \cdot \\ \cdot \\ (y_s)' \end{bmatrix} \tag{4}$$

where y'_{obs} indicates the observation increment and y' represents the simulation of the observation increments by the forecasting model ***M*** and observation operator ***H***.

$$(y_k)' = y_k(x_b + x') - y_k(x_b). \tag{5}$$

$$y'_{obs,k} = y_{obs,k} - y_k(x_b). \tag{6}$$

$$y_k = H_k\left[M_{t_0 \to t_k}(x)\right]. \tag{7}$$

The model perturbation (***MP***) matrix is then defined as $X' = (x'_1, x'_2, \ldots, x'_N)$ and the observation perturbation (***OP***) matrix is $Y' = \left(y'_1, y'_2, \ldots, y'_N\right)$. The POD transformation is applied to the ***OP*** matrix, and which ensure the orthogonality of the transformed ***OP*** samples ϕ_y. Orthogonal ***MP*** samples ϕ_x are also obtained by applying the same POD transformation to ***MP*** matrix. The optimal solution x' is calculated by using weighted mean of ***MP*** samples.

$$x'_a = \phi_{x,r}\beta. \tag{8}$$

Where $\beta = (\beta_1, \beta_2, \ldots, \beta_r)^T$. Its corresponding optimal OPs are determined by

$$y'_a = L(x'_a) = L\left(\phi_{x,r}\beta\right) = L\left(\phi_{x,r}\right)\beta \approx L_{x_b}\left(\phi_{x,r}\right)\beta = \phi_{y,r}\beta \tag{9}$$

The control variable of cost function transferred to β after substituting x'_a and y'_a into the cost function.

The background error covariance matrix ***B*** is obtained as in the ensemble Kalman filter (EnKF) (Evensen, 2004):

$$B = \frac{\phi_{x,r}\phi_{x,r}^T}{r-1} \tag{10}$$

Equations (8) and (10) then be substituted into equation (2). By solving the optimal problem, the incremental analysis can be attained.

$$\Phi_{y,r} = \left[(r-1)I_{r\times r} + \phi_{y,r}^{T} R^{-1} \phi_{y,r}\right]^{-1} \phi_{y,r}^{T} R^{-1}. \quad (11)$$

$$x_a' = \phi_{x,r} \Phi_{y,r} y_{obs}' \quad (12)$$

The final analysis x_a is expressed as follow

$$x_a = x_b + x_a' = x' + \phi_{x,r} \Phi_{y,r} y_{obs}' \quad (13)$$

2.3 Soil moisture data assimilation system for Pakistan

The soil moisture data assimilation system for Pakistan consists of the land surface model CLM4.5, the assimilation algorithm PODEn4DVar, and the observation operator. The observation operator (***H***) is needed to create a relationship between observations and the forecast model CLM4.5 simulated state variables. . In this study, the observation operator is simply a real matrix, which is used to link simulated soil moisture to observed soil moisture. The observation operator is expressed as

$$H = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \quad (14)$$

Where n indicates the dimension of model state vector, w_i is (the weight calculated from the distance between two points (x, x_i), y_i is the function value at point x_i.

This data assimilation system consists of two steps: (1) forecasting and (2) updating of state variable soil moisture. First, the daily simulated hydrogeological variables are obtained by running the CLM4.5 in the current assimilation window and then the updating procedure for the state variables according to PODEn4DVar

assimilation method. The updating process for state variables includes the following steps (Fig. 1):

(a) Read the CLM4.5 daily simulation outputs and historical simulation results to obtain sample matrix and then construct the background field vector.

(b) Construct the model perturbations (***MP***) and observation perturbations (***OP***) matrices.

(c) Generate ***OP*** samples ϕ_y and ***MP*** samples ϕ_x by applying the POD transformation to the ***OP*** matrix and ***MP*** matrix respectively.

(d) Calculate the optimal assimilation increment x_a' and the analysis field x_a as described in the assimilation method.

(e) Update the initialization file of CLM4.5 using the analysis field x_a and use this updated initialization file to run CLM4.5 to get a forecast for the next assimilation window and repeat the same steps for updating the state variables.

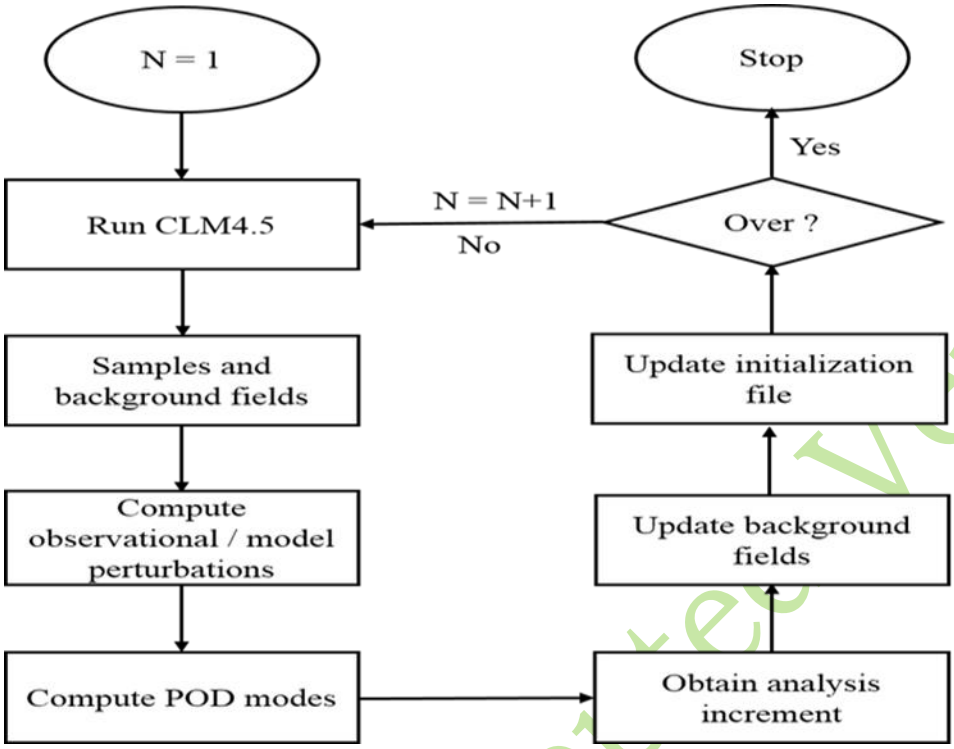

Fig. 1. Flow chart of data assimilation system.

3. Evaluation Experiments

3.1 Data description

In this study, we used atmospheric forcing data to run land model CLM4.5 and in-situ soil moisture information for the preliminary analysis and evaluation of the DA system for Pakistan. CRUNCEP version 4, with spatial resolution of 0.5°×0.5°, is a 110-year (1900–2010) dataset, which is the standard atmospheric forcing data provided with CLM4.5 and is used to derive model in offline mode. This dataset is generated by combining two datasets: (1) the 6-hourly NCEP reanalysis data with resolution of 2.5° (1948-2010) and (2) the monthly CRU TS3.2 data with 0.5°resolution (1901-2002) (Mitchell and Jones, 2005) (More details on the CRUNCEP dataset are accessible at

http://www.cesm.ucar.edu/models/cesm1.2/clm/clm_forcingdata_esg.html). This dataset has been used widely to derive CLM in studies on plant and vegetation development, and evapotranspiration (Mao et al., 2012; Mao et al., 2013; Shi et al., 2013), and in the TRENDY project (Piao et al., 2012).

Pakistan Meteorological Department (PMD) provided the soil moisture observational data for this study. The available soil moisture data from PMD was relative soil moisture and collected three times in a month i.e. 7th, 17th, and 27th, from the meteorological stations situated in agricultural fields across Pakistan. The collected relative soil moisture contents were then changed to volumetric water contents (multiply relative soil moisture contents to soil bulk density and divide it by water density) and used for assimilation and DA system evaluation.

The four selected agro-meteorological data sites were considered to be representative of different agro-climatic zones in Pakistan. They ranged from arid to humid (Chaudhry and Rasul, 2004).The localities of these data sites are presented in Fig. 2. Rawalpindi (RWP) agro-meteorological station is situated at the northern side of the Potohar Plateau. It represents rain fed plains with a sub-humid agro-climate. The major crops grown in this region are wheat, groundnut, and fodder. The Faisalabad (FSD) site represents the irrigated plains of central and southern Punjab and is in the dry semi-arid agro-climatic zone. Due to well managed canal system, it is a highly productive zone where wheat, rice, sugarcane, and cotton are the major crops. Quetta (QTA) is a high elevation agricultural rain fed site and has arid climatic characteristics. Wheat is the major crop in this zone. Aridity and low rainfall are the major causes of crop failure in this climatic-zone. Tandojam (TND) represents irrigated arid agro-climatic plains. It has

a well-organized irrigation system, and wheat, cotton, and rice are the major crops in this region.

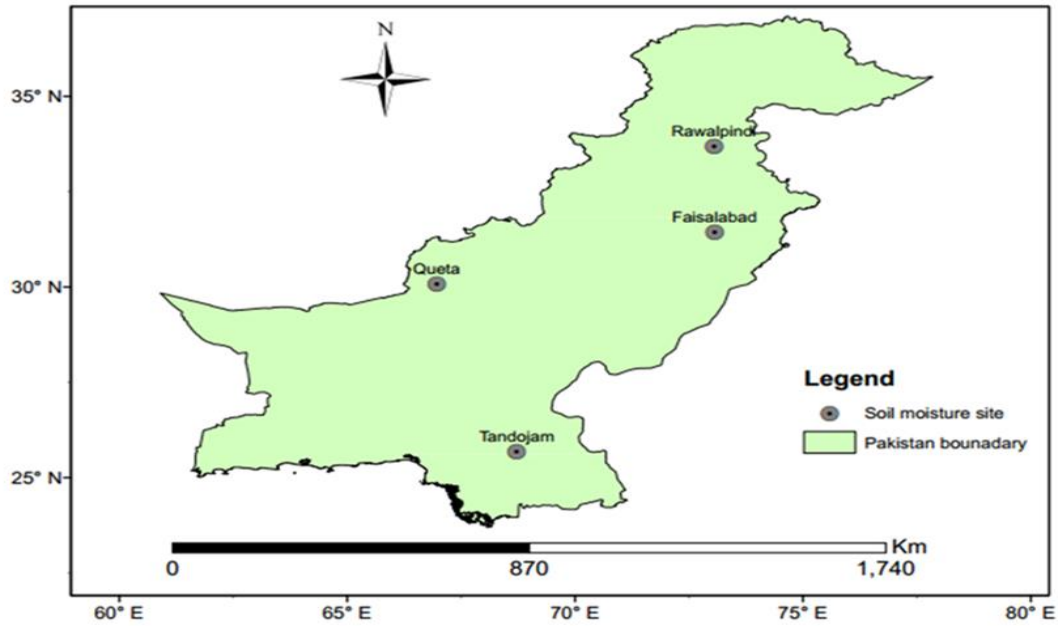

Fig. 2. Location map of the study sites in Pakistan.

3.2 Experimental design

3.1.1 In-situ soil moisture assimilation

The assimilation experiments were conducted at the four available soil moisture sites in Pakistan to evaluate the performance of the DA system based on CLM4.5 and PODEn4DVar. For reasonable initial conditions, a 100-year simulation of CLM4.5 was run at every data site using CRUNCEP atmospheric forcing data. The outputs of the spin-up simulation were choose as the initial conditions for all types of assimilation experiments. The spatial resolution of the model was set to be 0.1°×0.1° for all in-situ soil moisture data assimilation experiments. In all assimilation experiments, the historical

sampling scheme (Wang et al., 2010) was used and ensemble size was fixed to be 50 members. Another 50-year simulation of CLM4.5 was run using the spin-up results from the 100-year simulations as initial conditions.

Table 1. Assimilating soil depths and the corresponding CLM4.5 layers

In-situ soil depth (cm)	CLM4.5 layers (depth cm)
5	L3 (~6.2)
10	L4 (~11.9)
20	L5 (~21.2)
30	L6 (~36.6)

In current study, the in-situ soil moisture information from four soil-depths (0–5 cm, 5–10 cm, 10–20 cm, and 20–30 cm) for year 2006 were assimilated and the corresponding layers of CLM4.5 for these soil-depths are described in Table 1. To check the performance of the DA system, alternative soil moisture observations were assimilated and non-assimilated observations were considered to assess the DA system.

The effects of assimilating soil moisture observations at these four upper soil-depths on deeper soil-depths moisture simulations (30–40 cm, 40–50, 50–70, and 70–90 cm) by CLM4.5 were also investigated. Table 2 shows the deep soil layers information. It should be noted that both the 40 cm and 50 cm soil-depths exist within a single layer of CLM4.5, but observed soil moisture data was available for these depths. Therefore, this

information was used to evaluate the 40 and 50 cm soil-depths as well as the70 and 90 cm soil-depths.

Table 2. Evaluating soil depths and the corresponding layers of CLM4.5

In-situ soil depth (cm)	CLM4.5 layers (depth cm)
40	L7 (~62.0)
50	L7 (~62.0)
70	L8 (~103.8)
90	L8 (~103.8)

3.1.2 Observing system simulation experiments (OSSEs)

Observing system simulation experiments (OSSEs) are considered one of the best options for the assessment and evaluation of a DA system because it produces both the "observations" and "true" states. In this study, OSSEs were conducted for Pakistan. The 100-year spin up simulation of CLM4.5 with 1 degree horizontal resolution was run using CRUNCEP data to acquire the suitable initial conditions for the DA experiments. Daily simulations of CLM4.5 for year 2004 using CRUNCEP atmospheric forcing data were treated as the "true" fields in this experiment. The daily averaged soil moisture values calculated by adding errors to the "true" fields were used as the "observations" for the assimilation. Both the simulation (without DA) and assimilation experiments were driven by QIAN atmospheric forcing data for year 2004. In OSSEs, the ensemble size and

sampling strategy were kept the same as those used in the in-situ soil moisture assimilation experiments. In these experiments, assimilation was carried out for all ten layers of the land model whereas for in-situ soil moisture assimilation experiments, only four CLM4.5 layers were used for assimilation.

3.3 Results and discussion

3.3.1 In-situ soil moisture assimilation results

3.3.1.1 Assimilation and evaluation results for the top layers

The preliminary results of the DA system for the top layers assimilation are described in Fig. 3. The black dots in Fig. 3 indicate the observations used for the evaluation, whereas the green dots are the assimilated observations. Figure 3 (a, b, c, d) shows the assimilation results for the 0–5 cm soil layer whereas Fig. 3 (e, f, g, h) for the 20–30 cm soil layer at the experimental sites. The assimilation results for soil-depths 5–

10 and 10-20 cm are not shown because they produced similar results.

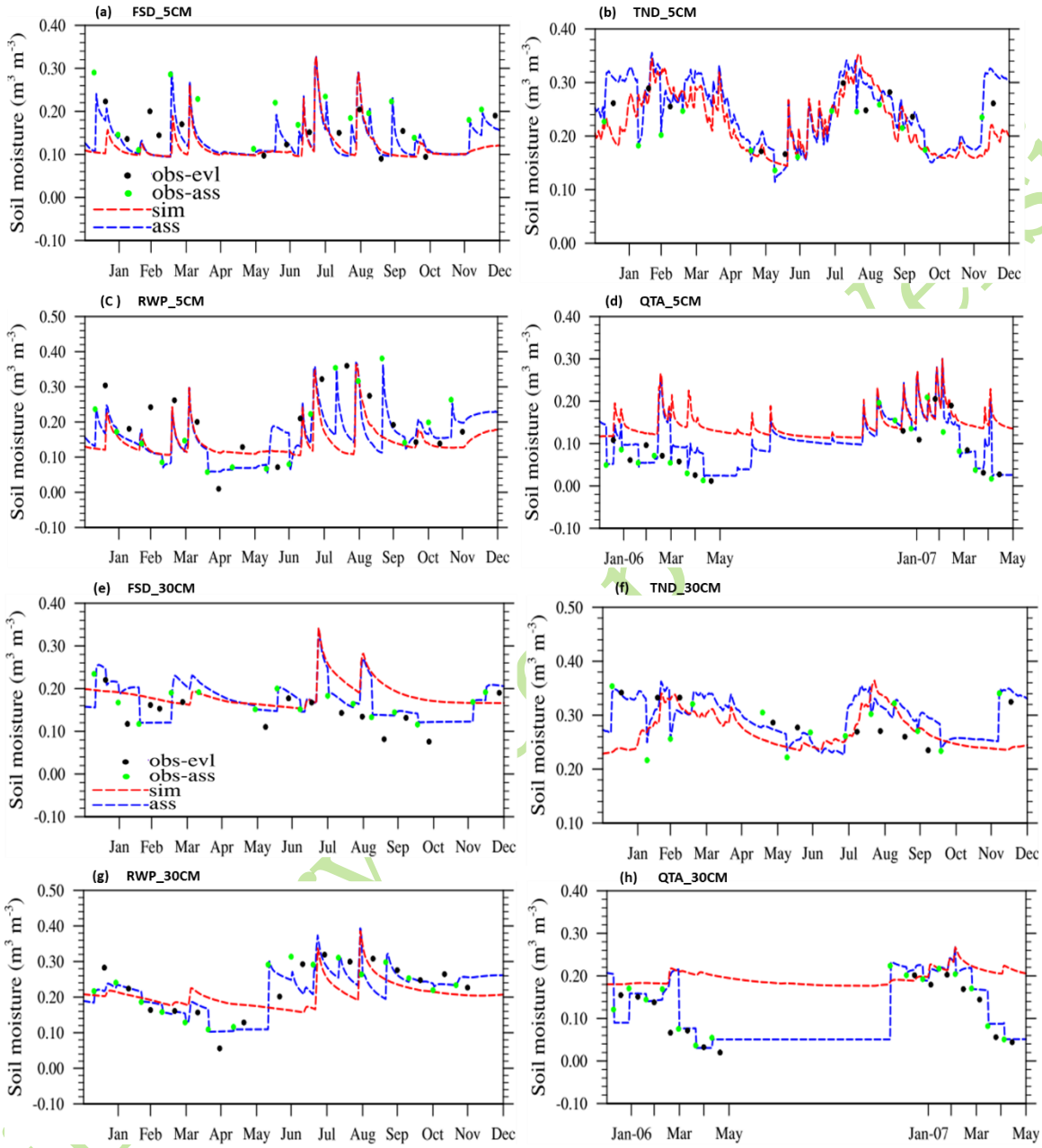

Fig. 3. Assimilation of *in-situ* soil moisture observations for two soil layers (0–5 cm and 20–30 cm) at different sites (red line: simulated soil moisture (without DA), blue line: assimilation, green dot: assimilated observed soil moisture, black dot: observed soil moisture value for evaluation).

Figure 3 also shows the time series of assimilation and CLM4.5 simulation (without soil moisture assimilation) for the year 2006. It is important to remind that

alternative soil moisture observations are assimilated and remaining observations are used for the evaluation of DA system. It is observed that the assimilation time series for all stations at both soil depths (5cm and 30cm) is much closer to black dots which are the soil moisture observations used for evaluation than the simulation. The closeness of assimilation line to black dots clearly shows that the assimilation improved the estimation of soil moisture.

The statistical indices for all the sites clearly showed that assimilation has significant improvement in soil moisture estimation with higher correlation coefficients, smaller RMSE, and lower BIAS (Fig. 4). The FSD and RWP sites at two soil layers (0-5 cm and 5-10 cm) with negative BIAS (Fig. 4 (i, k)) showing the underestimation whereas the other two stations (Fig. 4(j, l)) overestimated the soil moisture estimation with respect to observations. Overall the simulations showed the overestimation in soil moisture with higher biases at all stations and at maximum number of soil-depths than the assimilation run (Fig. 4(i, j, k, l)). This overestimation in soil moisture for simulation run is consistent with the previous studies (Long et al., 2013; Cai et al., 2014). However, this overestimation of soil moisture was reduced by DA, which decreased the RMSE (Fig. 4(e, f, g, h)) and produced higher correlation coefficients (Fig. 4(a, b, c, d)). At the QTA site, which was a rain fed and high elevation agricultural field site, soil moisture data was only collected during the wheat season because wheat was the major crop. The soil moisture data for two wheat seasons for year 2006 and 2007 were used for the assimilation (Fig. 3(d, h)).

Thus the statistical analysis indicated that soil moisture estimation improved (Fig.4) when the in-situ soil moisture data was assimilated at four top soil layers (0–5 cm,

5–10 cm, 10–20 cm, and 20–30 cm) and hence the performance of DA system was reasonable.

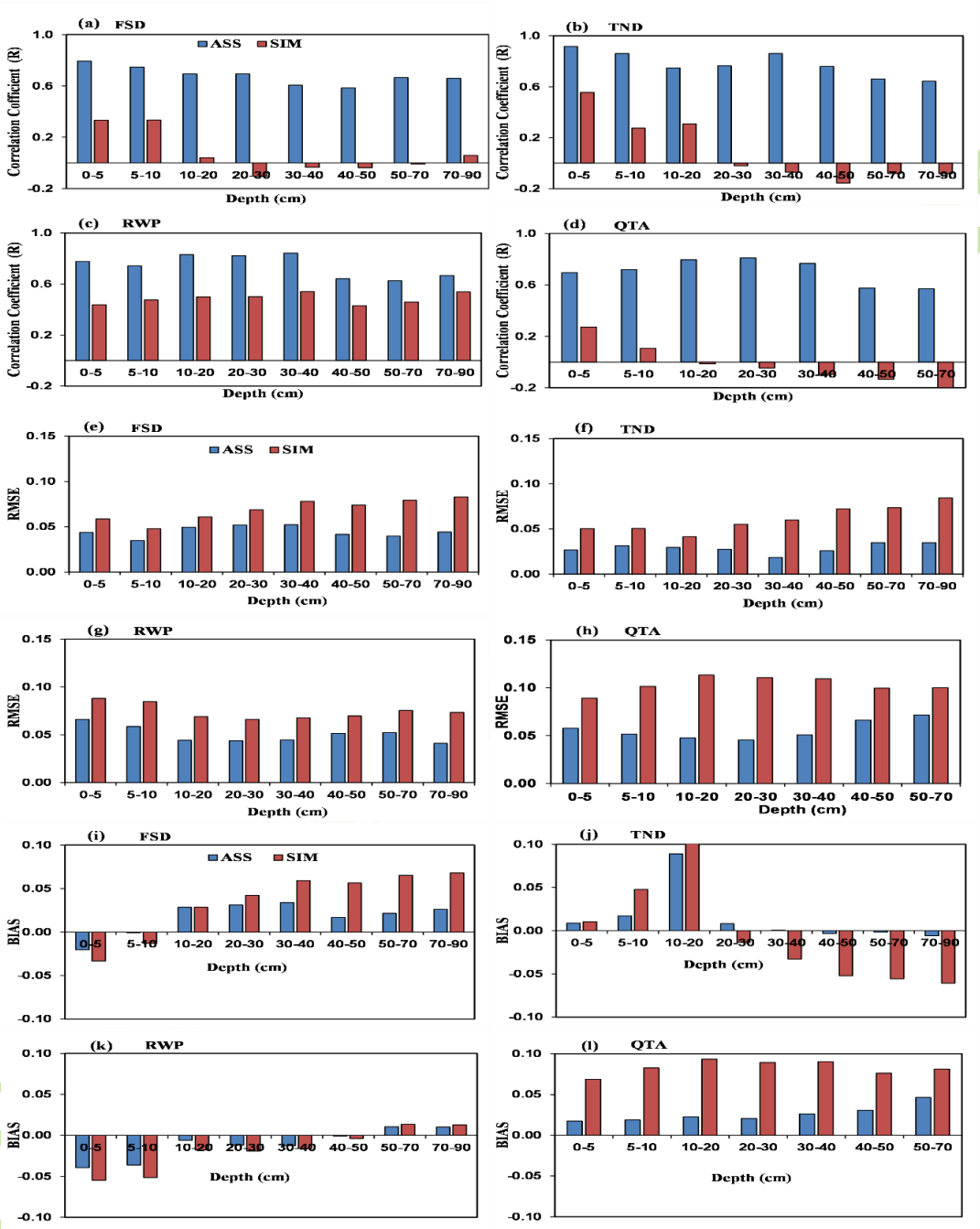

Fig. 4. Statistical analysis (R, RMSE, and BIAS) of simulated (without DA) and assimilated soil moisture against in-situ observations for different soil layers at different sites in Pakistan.

3.3.1.2 Effects of assimilation on the deep layers

Another aim of this research was to explore the effects of the top soil layers soil moisture assimilation over the deeper soil layers soil moisture estimates. The evaluating soil layers information is shown in Table 2. Figure 5 shows the assimilation effects at soil- depths 40–50 and 50–70 cm, respectively, for all the experimental sites. The results for soil-depths 30–40 and 70–90 cm are not shown because of similar results. Soil moisture observations for QTA were not available for soil layer 70–90 cm, and therefore, the results at QTA_90CM are missing for evaluation.

Figure 5 shows that the assimilation time series is closer to the observations than the simulation (without DA), which suggests that soil moisture estimations for the deeper soil layers have improved, even when there is no assimilation of soil moisture done in these soil layers of model. Figure 4 also represents the statistical indices for the deeper soil layers (30–40, 40–50, 50–70, and 70–90 cm). It is observed that simulation had a lower correlation coefficient and higher RMSE as compared to assimilation for all the deeper soil layers at all the experimental sites which shows that assimilation has improvement in soil moisture estimation at deeper soil layers (Fig. 4(a, b, c, d)). Higher biases for simulation were also recorded for the deeper soil layers than the top soil layers, except for RWP where the biases difference was smaller at the deeper soil layers than the other sites (Fig. 4(i, j, k, l)). These higher biases might be due to systematic biases of CLM (Long et al., 2013; Cai et al., 2014). Overall the statistical analysis with higher correlation coefficient, less RMSE and BIAS showing that the assimilation in the top soil

layers improved the soil moisture estimates in the deeper soil layers as well, which means that the DA system can be reliably used for soil moisture assimilation.

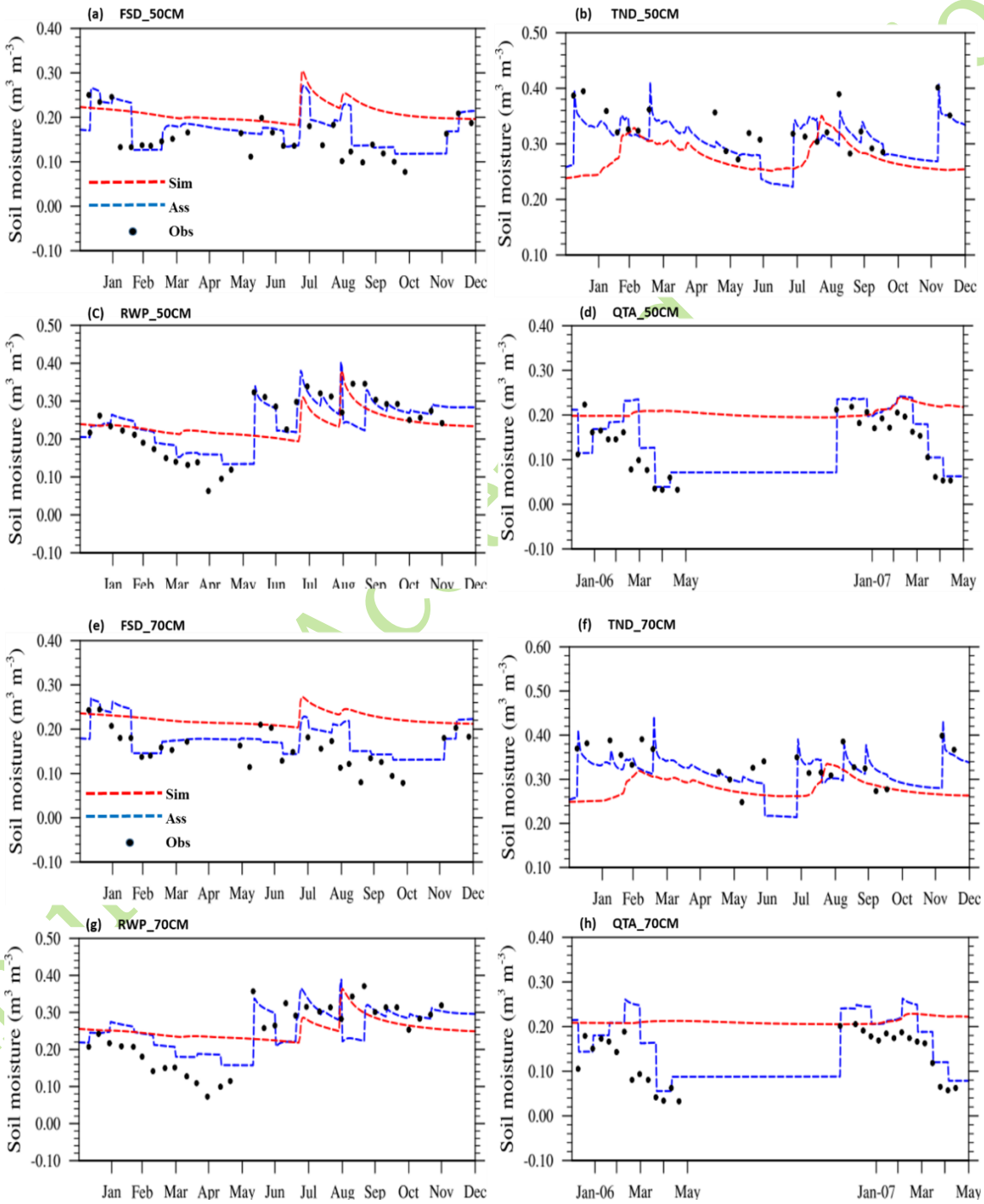

Fig. 5. Effects of assimilation on two soil layers (40–50 cm and 50–70 cm) at different sites (red line: simulated soil moisture (without DA), blue line: assimilation, and black dot: observed soil moisture).

3.3.1.3 Soil temperature and surface heat fluxes

Figure 6 shows the differences in soil temperature between the simulated (without DA) and assimilated values produced by CLM4.5 for the top four soil layers. The difference is obtained from the simulated minus assimilated values for daily soil temperature. However, the difference was only calculated for the assimilation days at the four experimental sites. The magnitude of the soil temperature difference varies from site to site and layer to layer. The maximum temperature difference (14 K) was observed at QTA whereas the minimum difference was 0.9 K at the FSD station. These variations in soil temperature were substantial, which indicated that assimilation of soil moisture observations in CLM4.5 produced different results for soil temperature as well.

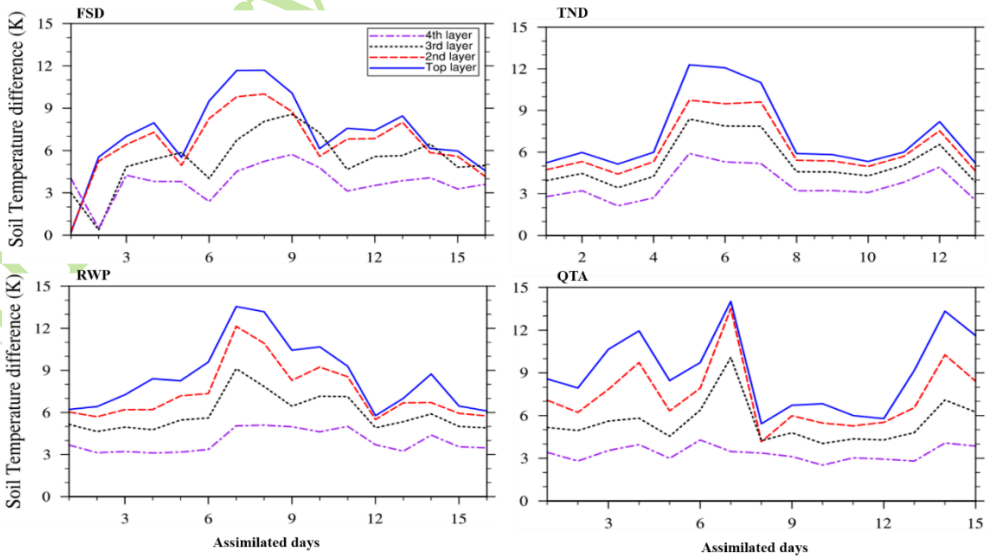

Fig. 6. Differences between simulated (without DA) and assimilated daily soil temperatures in the top four soil layers of CLM4.5 at four different experimental sites.

The soil moisture difference can change the simulation of surface latent flux, whereas the sensible heat flux may show adverse performance because of the change in soil temperature (Tian et al., 2008a). Figure 7 shows the simulated minus assimilated differences in latent heat and sensible heat fluxes at the different sites. The latent heat flux difference varies from –21.5 to 152.5 W/m^2 among the experimental stations, whereas –0.04 to 115.5 W/m^2 for the sensible heat flux difference. These higher differences in surface heat fluxes could produce striking impacts on the land-atmosphere interaction at all the stations.

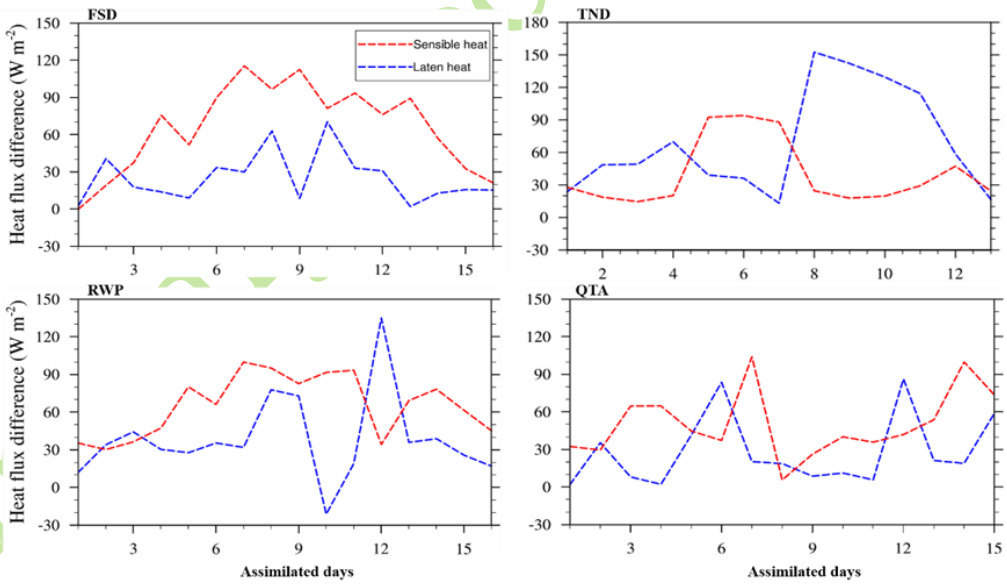

Fig. 7. Differences between simulated (without DA) and assimilated daily surface latent and sensible heat fluxes at four experimental sites.

3.3.2 Observing system simulation experiments (OSSEs)

Figure 8 shows the results of OSSEs experiment, carried out for Pakistan. In these experiments, daily soil moisture observations were assimilated only for the rainy season, which is from June–August (JJA) in Pakistan to evaluate the DA system. The constant error of 0.012 was added to the "true" fields to generate the daily soil moisture observations and these artificial observations were assimilated into the system. Figure 8 shows the evaluation results of daily assimilation for only four soil layers whereas assimilation was carried out for all the soil layers of CLM. The results for the other soil layers are not shown. Daily assimilation produced significantly good performance during the rainy season (JJA).

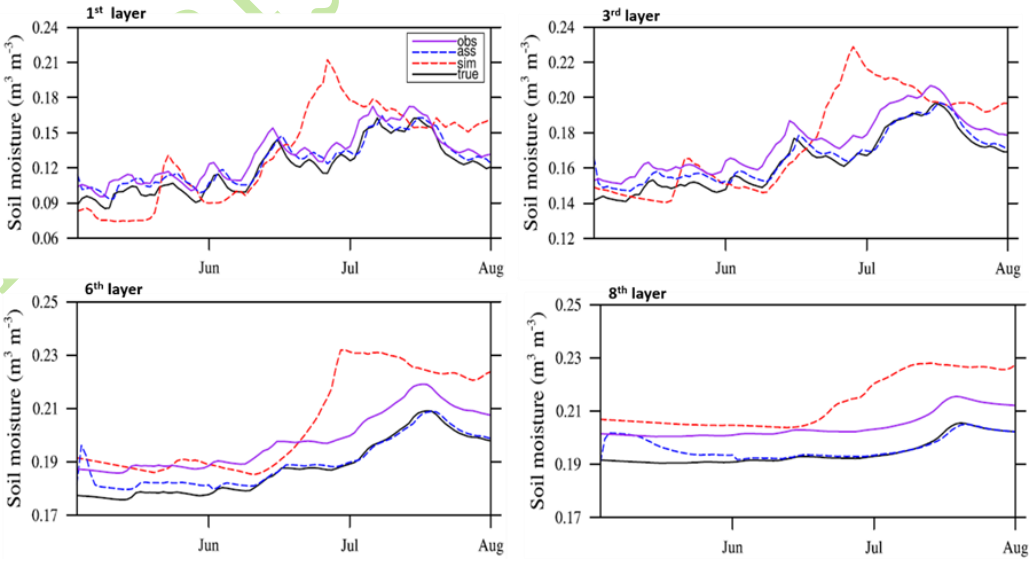

Fig. 8. Time series for the soil moisture observations, true fields, assimilation and simulation (without DA) for the 1st, 3rd, 6th, and 8th soil layers of CLM4.5 for Pakistan.

From the Fig. 8, it is cleared that the assimilation time series is more consistent and closer to observation and "true" time series than the simulation for the whole time period of experiment, which clearly indicates that there is an improvement in soil moisture estimation. Table 3 shows the statistical analysis for the daily OSSEs experiments. The results with higher correlation coefficient and lower RMSE values for assimilation suggest that the DA system has improved the estimation of daily soil moisture. The simulation overestimates soil moisture, which is similar to previous results (Long et al., 2013; Cai et al., 2014), and this overestimation is reduced in all soil layers by assimilation. Overall, the OSSEs results showed that the data assimilation system can reliably estimate soil moisture in Pakistan.

Table 3.Comparison of the root mean square errors (RMSE) and the correlation coefficients (R) for assimilation and simulation (without DA) in OSSE experiments

	1st layer	**3rd layer**	**6th layer**	**8th layer**
RMSE (sim)	0.30	0.21	0.49	0.23
RMSE (ass)	0.075	.046	.035	0.030
R (sim)	0.70	0.72	0.80	0.74
R (ass)	0.98	0.97	0.96	0.94

4. Conclusions

In current study, a general framework for soil moisture data assimilation (DA) has been developed for Pakistan. In this soil moisture DA system, PODEn4DVar was used as the assimilation algorithm whereas the Community Land Model CLM4.5 was selected as the forecasting operator. For the performance evaluation of the DA system, preliminary analysis and evaluation experiments were conducted at four agricultural sites from different agro-climatic zones across Pakistan, and the in-situ information of soil moisture were assimilated. The alternative soil moisture observations for four top soil-depths i.e., 0–5 cm, 5–10 cm, 10–20 cm, and 20–30 cm were assimilated in CLM4.5 for evaluation. The correlation coefficients, root mean square errors, and the biases after assimilation significantly improved for the top four soil layers. The results indicated that the DA system can produce more accurate and precise soil moisture estimations.

The effects of top layers soil moisture assimilation over lower soil-depths, i.e., 30–40 cm, 40–50 cm, 50–7 0cm, and 70–90 cm, were also investigated. The statistical indices (R, RMSE, and BIAS) improved at all sites and for all soil-depths. These results are a clear indication that assimilation in the top soil layers can also improve CLM4.5 soil moisture estimations in deeper soil layers. Thus, the evaluation results showed that DA system based on PODEn4DVar and CLM4.5 can improve soil moisture simulation. For further assessment of the soil moisture DA system, we conducted OSSEs experiments for Pakistan. However, these experiments were only undertaken in, only the rainy season (JJA).To validate the performance of the DA system, daily soil moisture information were used for assimilation. The evaluation results from the OSSEs experiments clearly showed that assimilation can improve soil moisture estimation.

In this study, the evaluation experiments were conducted at only four stations in Pakistan by assimilating the in-situ soil moisture data because of sparse data. However, this soil moisture DA system is expected more detailed and comprehensive validation after more soil moisture information are obtained in Pakistan. The findings from these small scale assimilation experiments and the OSSEs showed that the biases exist in the CLM4.5. The current results indicated that the DA system has the potential to improve land surface conditions in Pakistan, which may improve weather and climate studies in Pakistan. Our further studies will emphasis on parameter estimation, addressing the biasness in other state variables and calibration of the land surface model CLM4.5 before data assimilation.

References

Cai, X. T., Z. -L. Yang, Y. L. Xia, et al ., 2014: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *J. Geophys. Res., Atmos.*, **119**, 751-13,770, doi: 10.1002/2014JD022113.

Chahine, M. T., 1992: The hydrological cycle and its influence on climate. *Nat.*, **359**, 373-380.

Chaudhry, Q.-u.-Z., and G. Rasul, 2004: Agro-climatic classification of Pakistan. Quarterly Science Vision, **9**, 59-66.

Crow, W. T., A. A. Berg, M. H. Cosh, et al., 2012: Upscaling sparse ground- based soil moisture observations for the validation of coarse - resolution satellite moisture products. Rev. Geophys., **50**. RG2002, doi:10.1029/2011RG000372.

Dai, A. G., K. E. Trenberth, and T. Qian, 2004: A Global Dataset of Palmer Drought Severity Index for 1870–2002: Relationship with Soil Moisture and Effects of Surface Warming. *J. Hydrometeorol.*, **5**, 1117-1130, doi : /10.1175/JHM-386.1.

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, et al., 2007: State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency. *Water Resour. Res.*, **43**, W06401, doi: /10.1029/2006WR005100.

Dirmeyer, P. A., A. J. Dolman, and N. Sato, 1999: The Global Soil Wetness Project: A pilot project for global land surface modeling and validation, *Bull. Amer. Meteor. Soc.*, **80**, 851– 878, doi:10.1175/152 0477(1999)080<0851:TPPOTG>2.0.CO;2.

Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean. Dyn.*, **54**, 539-560, doi:10.1007/s10236-004-0099-2.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, et al., 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, **34**, 3405-3420, doi : 10.1029/1998WR900001.

Hurrell, J. W., M. M. Holland, P. R. Gent, et al., 2013: The Community Earth System Model: A Framework for Collaborative Research. *Bull. Amer. Meteor. Soc.*, **94**, 1339-1360, doi:10.1175/BAMS-D-12-00121.1.

Koster, R. D., P. A. Dirmeyer, Z. C. Guo, et al., 2004: Regions of Strong Coupling Between Soil Moisture and Precipitation. *Sci. Rep.*, **305**, 1138-1140, doi :10.1126/science.1100217.

Kumar, S. V., R. H. Reichle, R. D. Koster, et al., 2009: Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations. *J. Hydrometeorol.*, **10**, 1534-1547. doi : 10.1175/2009JHM1134.1.

Liu, D., and A. k. Mishra, 2017: Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale. *J. Hydrol.*, **547**, 67-79, doi: 10.1016/j.jhydrol.2017.01.036.

Long, D., B. R. Scanlon, L. Longuevergne, et al., 2013: GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophys. Res. Lett.*, **40**, 3395-3401, doi : 10.1002/grl.50655.

Mao, J., P. E. Thornton, X. Y. Shi, et al., 2012: Remote Sensing Evaluation of CLM4 GPP for the Period 2000–09*. J. Climate, **25**, 5327-5342, doi: 10.1175/JCLI-D-11-00401.1.

Mao, J., X. Y. Shi, P. E. Thornton, et al., 2013: Global Latitudinal-Asymmetric Vegetation Growth Trends and Their Driving Mechanisms: 1982–2009. *Remote. Sens.*, **5**, 1484-1497, doi : 10.3390/rs5031484.

Mitchell, K. E., D. Lohmann, P. R. Houser, et al., 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res. Atmos.*, **109**, D07S90, doi : 10.1029/2003JD003823.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693-712, doi : 10.1002/joc.1181.

Oleson, K. W., Y. J. Dai, G. Bonan, et al., 2004: Technical description of Community Land Model (CLM). NCAR Tech. Note/NCAR/TN-461+STR, Boulder Colorado, 173pp, doi : 10.5065/D6N877R0.

Oleson, K. W., D. M. Lawrence, G. B. Bonan, et al., 2010: Technical description of version 4.0 of the Community Land Model (CLM). NCAR Tech. Note/NCAR/TN-478+STR, Boulder Colorado, 266pp, doi: 10.5065/D6FB50WZ.

Oleson, K. W., D. M. Lawrence, G. B. Bonan, et al ., 2013: Technical description of version 4.5 of the Community Land Model (CLM). NCAR Tech. Note/NCAR/TN-503+STR, Boulder Colorado, 171pp, doi: org/10.5065/D6RR1W7M .

Piao, S., A. Ito, S. Li, et al., 2012: The carbon budget of terrestrial ecosystems in East Asia over the last two decades. *Biogeo. Discuss.*, **9**, 4025-4066, doi : 10.5194/bg-9-3571-2012.

Qian, T. T., A. G. Dai, K. E. Trenberth, et al., 2006: Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations. *J. Hydrometeorol.*, **7,** 953-975, doi : 10.1175/JHM540.1.

Robinson, D. A., C. S. Campbell, J. W. Hopmans, et al., 2008: Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vad.Zon. J.*, **7**, 358-389.

Robock, A., K. Y. Vinnikov, G. Srinivasan, et al., 2000: The Global Soil Moisture Data Bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281-1299, doi:10.2136/vzj2007.0143.

Sheffield, J., and E. F. Wood, 2008: Global Trends and Variability in Soil Moisture and Drought Characteristics, 1950–2000, from Observation-Driven Simulations of the Terrestrial Hydrologic Cycle. J. climate, **21**, 432-458, doi :10.1175/2007JCLI1822.1

Shi, C. X., Z. H. Xie, H. Qian, et al., 2011: China land soil moisture EnKF data assimilation based on satellite remote sensing data. *Sci. China Earth Sci.*, **54**, 1430-1440, doi: 10.1007/s11430-010-4160-3.

Shi, X. Y., J. F. Mao, P. E. Thornton, et al., 2013: Spatiotemporal patterns of evapotranspiration in response to multiple environmental factors simulated by the Community Land Model. *Environ. Res. Lett.*, **8**, 024012, doi:10.1088/1748-9326/8/2/024012.

Sun, Q., Z. H. Xie, and X. J. Tian, 2015: GRACE terrestrial water storage data assimilation based on the ensemble four-dimensional variational method PODEn4DVar: Method and validation. *Sci. China Earth Sci.*, **58**, 371-384, doi: 10.1007/s11430-014-4978-1.

Tian, X. J., Z. H. Xie, and A. G. Dai, 2008a: A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model. *J. Geophys. Res.*,, **113**, D14127, doi : 10.1029/2007JD009650.

Tian, X. J., Z. H. Xie, and A. G. Dai, 2008b: An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.: Atmos.*, **113**, D21124, doi :10.1029/2008JD010358.

Tian, X. J., Z. H. Xie, and Q. Sun, 2011: A POD-based ensemble four-dimensional variational assimilation method. *Tellus A.*, **63**, 805-816, doi : 10.1111/j.1600-0870.2011.00529.x.

Wang, B., J. J. Liu, S. D. Wang, et al., 2010: An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **27**, 715-727, doi: 10.1007/s00376-009-9122-3.

Zhang, S. L., J. C. Shi, and Y. J. Dou, 2012: A soil moisture assimilation scheme based on the microwave Land Emissivity Model and the Community Land Model. *Int. J. Remote. Sens.*, **33**, 2770-2797, doi : 10.1080/01431161.2011.620032.

Zreda, M., W. J. Shuttleworth, X. Zeng, et al., 2012: COSMOS: the cosmic-ray Soil Moisture Observing System. *Hydrol. Earth Syst. Sci.*, **16**, 4079-4099, doi : 10.5194/hess-16-4079-2012

602