Retrieval of Soil Moisture Deficit through Climate Change Initiative (CCI) Soil Moisture Data and Probability Distributed Model

*[1]Prashant K Srivastava, [1]Swati Maurya, [1]Varsha Pandey, [2]Dharmendra Pandey

[1]Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi-221005, India

[2]Space Application Centre, ISRO, Ahmedabad, India

***Correspondence:** prashant.iesd@bhu.ac.in

Abstract:

Soil moisture deficit (or SMD) is very important variable for many applications such as for flood and drought modelling, which is now possible to estimate using the remote sensing data. This study is an attempt to evaluate the climate change initiative (CCI) soil moisture data for SMD estimation at a catchment scale. The SMD (Soil Moisture Deficit) estimated from the Probability Distribution Model, by using the *in situ* station data was used as a benchmark for all the comparisons. Approaches based on generalized linear model and relevance vector machine are provided for the estimation of SMD. The overall analysis reveals that CCI soil moisture is of reasonable quality in estimating the soil moisture deficit at a catchment level. Therefore, this study provides first time comprehensive evaluation of CCI soil moisture in Indian condition and the result provides supportive evidence of the potential value of this product for meso-scale studies and hydrological applications.

Keywords: Soil moisture deficit, Climate Change Initiative soil moisture, Linear and non-linear Modelling, Catchment

1. Introduction

Soil moisture is an important variable for understanding numerous environmental phenomena such as climate change, agriculture, hydrology, meteorology etc. It is also an integral component for predicting weather phenomenon, designing water balance equations and mitigation of natural hazards like floods and drought(Mladenova *et al.*, 2011; Srivastava *et al.*, 2016). *In situ* soil moisture probes are providing accurate measurements of soil moisture, however because of high spatial variations, *in situ* soil moisture data is inadequate for large scale studies (Al-Shrafany *et al.*, 2012). Nowadays with the state-of-arts remote sensing technique, it is now possible to measure near surface soil moisture content by optical and thermal infrared, as well as microwave (active and passive) remote sensing techniques (Owe *et al.*, 2001; Scott *et al.*, 2003). In this direction, the ESA's Climate Change Initiative (CCI) Soil Moisture (CCI-SM) provides global long-term consistent soil moisture dataset, based on the active and passive satellite data. The Soil Moisture CCI project is part of the ESA Programme on Global Monitoring of Essential Climate Variables (ECV), initiated in 2010 (Dorigo *et al.*, 2015).

The Soil Moisture Deficit/Depletion (SMD) is the important variable for estimating the degree of variability in soil moisture content. Soil moisture deficit indicates the amount of water required to raise the soil moisture content to the level of field capacity. SMD is found to be inversely related with the soil moisture(Srivastava *et al.*, 2014). There is chance of drought, if SMD condition is higher in the soil for a long time (Srivastava *et al.*, 2013a), while low SMD in soil may cause flooding situation during the high spell of rainfall for a shorter duration of time. Few studies proved that SMD can be estimated by using the satellite soil moisture information (Srivastava *et al.*, 2013b). It has been found from the previous studies that Probability Distributed Model (PDM) is very useful for estimation of SMD by using the hydro-meteorological datasets. Hence, this study aims to evaluate the usefulness of the CCI soil moisture for SMD estimation, by using the techniques like Generalized Linear Model and Relevance Vector Machines and PDM.

2. Materials and Methodology

2.1. Study Area

The Chakaliya catchment is chosen for this study, as it is suitable experimental site for satellite and hydrological modelling, because of availability of nearly all sorts of data required in this type of studies such as soil type, meteorological variables and gauge data. The catchment covers 3121 sq km area with geographical coordinate's $74^{0}19'14''$ longitude and $23^{0}03'15''$ latitude. The entire Chakaliya catchment comes under the Mahi River Basin of Rajasthan, India. This catchment receives about 738 mm annual average rainfall and sub-tropical climatic condition with mild winter, moderate summer and high humidity. The land use is diverse ranges from highly dense forest to hilly terrain and plateau having soil type of silt loam to clay loam. The soil types are silt loam to clay loam and grayish brown to dark grayish brown. Major crops are wheat, maize, soya bean, gram and garlic.

2.2. The Climate Change Initiative Soil Moisture (CCI-SM) products

In this study, the Climate Change Initiative soil moisture (CCI-SM) product from ESA has been used for estimation of SMD. The Soil Moisture CCI project is part of the ESA Programme on Global Monitoring of Essential Climate Variables (ECV), better known as the Climate Change Initiative (CCI), initiated in 2010. It is started to create a long-term consistent soil moisture time series, based on active and passive satellite data. The CCI-SM combined product has a 25 Km spatial and daily temporal resolutions, with soil moisture measuring units in m^3m^{-3}.

2.3. Probability Distribution Model

The Probability distribution model (PDM) is a conceptual hydrological model, which constitute a combined rainfall-runoff model that converts rainfall and evapotranspiration variables to flow at the catchment outlet (Srivastava *et al.*, 2013c). The model uses soil moisture distribution for soil moisture accounting and a series of linear or non-linear reservoir for the routing components (Moore, 2007). The PDM has been applied all over the world and widely used in operational and designing purposes (Srivastava *et al.*, 2015). The main inputs for PDM model are rainfall and reference evapotranspiration (ET_0) and the model output is the river flow. In the current study, we used PDM for the estimation of SMD (Soil Moisture Deficit), by using the SMD routine as shown below:

$$\frac{E_i^{'}}{E} = 1 - \left\{\frac{(S_{max} - S(t))}{S_{max}}\right\}^{b_e} \tag{1}$$

where, $\frac{E_i^{'}}{E}$ represents the ratio of actual to potential ET; $(S_{max} - S(t))$ is SMD; b_e is an exponent in the actual evaporation function ; S_{max} is the total available storage and S(t) is storage at a particular time t.

2.4 Generalized Linear Model (GLM)

Generally, in linear regression, the relationships between two variables are obtained by fitting a linear equation to observed data. One variable is considered to be an explanatory variable ($\boldsymbol{x}_i$), and the other is considered to be a dependent variable ($\boldsymbol{y}_i$) (Johnson and Wichern, 2002). By considering the familiar linear regression model:

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{b} + \boldsymbol{e}_i, \tag{2}$$

where $i = 1, . . . , n$,; y_i is a dependent variable; $\boldsymbol{x}_i$ is a vector of k independent predictors; $\boldsymbol{b}$ is a vector of unknown parameters; and the e_i is stochastic disturbances.

The Generalized Linear Model (GLM) is characterized by stochastic component, systematic component and link between the random and systematic components (McCullagh and Nelder, 1989). The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. In this case the explanatory variable is soil moisture, while the dependent variable is SMD.

2.5 Relevance Vector Machine

Tipping (2001) proposed the relevance vector machine (RVM) in a Bayesian context. It is based on a probabilistic theory operated by a set of hyperparameters associated with weights, iteratively can be estimated from the data. RVM typically utilizes fewer kernel functions and the training vectors associated with nonzero weights are called as relevance vectors. The RVM model variable can be characterized for forecast y at a given x by the equations (3) and (4):

$$y = f(x) + \varepsilon_n \quad \text{where } \varepsilon_n \sim N(0, \sigma^2_{\varepsilon_n}) \tag{3}$$

$$p(y|w, \sigma^2_{\varepsilon_n}) = (2\pi\sigma^2_{\varepsilon_n})^{-l/2} \exp\left\{ -\frac{1}{2\sigma^2_{\varepsilon_n}} \|y - \Phi_w\|^2 \right\} \tag{4}$$

where, weights, $w = (w_0, w_1, \ldots\ldots, w_l)^T$, $\sigma^2_{\varepsilon_n}$ is noise variance, ε_n are independent samples from some noise process and $\Phi_{(x_i)} = [1, K(x_i, x_1), K(x_i, x_2), \ldots., K(x_i, x_l)]^T$ in which $K(x_i, x_l)$ represents kernel function.

Tipping (2001) recommended the addition of a complex penalty to the likelihood or error function to avoid the over fitting problem of w and $\sigma^2_{\varepsilon_n}$. The advantage with RVM is that its predictions are probabilistic and it doesn't make unnecessarily liberal use of basis functions. Use of fully probabilistic framework is a useful approach in RVMs with a priori over the model weights governed by a set of hyperparameters.

2.6 Statistical parameters

In this study, we compared the SMD obtained from GLM and RVM with those obtained from PDM by using the locally measured station based datasets. Although there are many statistics available, three of them are used in this study: *root mean square error (RMSE), Correlation (Corr) and Bias* as shown below:

Root Mean Square Error (Eq. 5)

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i-1}^{n}[y_i - x_i]^2\right)} \tag{5}$$

Correlation (r) (Eq.6),

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}} \tag{6}$$

Bias (Eq.7):

$$Bias = [(\overline{y} - \overline{x})] \tag{7}$$

where n is the number of observations; x_i is the ground based measurements and y_i is the estimated measurements; $\overline{x}$ is the mean of ground based measurements.

3. Results and Discussion:

3.1 Evaluation of soil moisture and SMD derived from PDM

This approach represents a more direct comparison between the CCI soil moisture and SMD from PDM. The time series obtained for soil moisture and SMD is shown in **Figure 1**. The pattern in soil moisture derived from CCI soil moisture indicated a closer relationship with the PDM SMD. As expected, the soil moisture content was higher during the monsoon and winter season and lower during the July month. On the contrary, SMD showed an inverse trend as compared to the soil moisture with marked fluctuations over the entire period and followed a strong seasonal cycle. The decrease in soil moisture can be attributed to the increasing temperature, with high evaporative demands during the July month period leading to a progressive drying of the soil. The high soil moisture during August can be attributed to the high rainfall during these months and low temperature during the winter months.

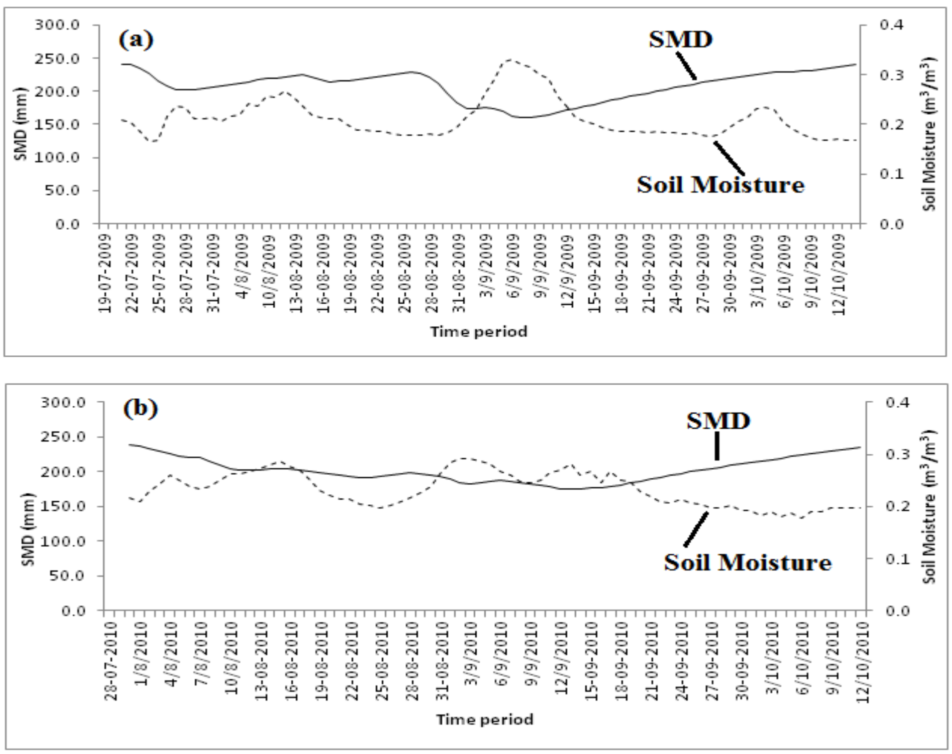


Figure 1 Time series of CCI soil moisture and SMD a-Calibration and b-Validation

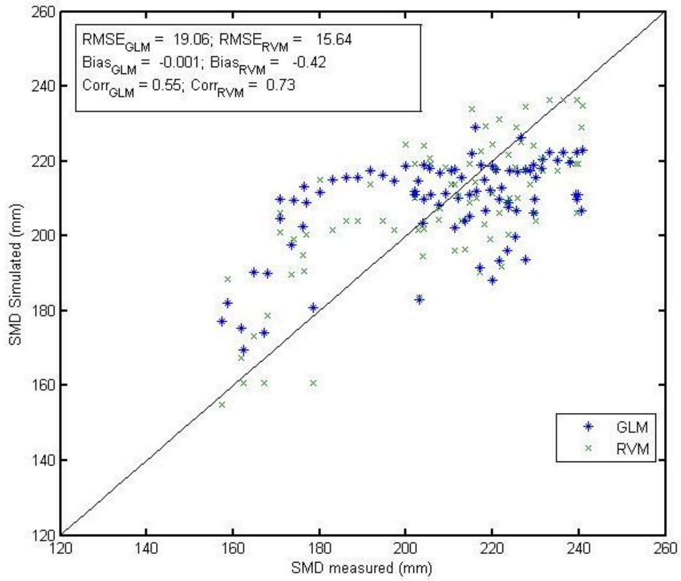
3.2 Parameterization of GLM and RVM techniques

The GLM model used in this study was comprised of the soil moisture as explanatory variables and the PDM SMD as a dependent variable. The binomial regression family has been used with probit as a link function and binomial as a variance. The RVM techniques need to be optimised for its parameters for a satisfactory result. A preliminary analysis of parameters was performed before using it for SMD prediction following the method given by Srivastava et al. in (2013). The model selection of RVM was based on subsuming hyperparameter adaptation and feature selection with respect to different model selection criteria. For the RVM model, the Gaussian radial based kernel function was used, because of its high performance and general approximation ability. The sigest module supported by R language was used in this study for performing an automatic hyperparameter estimation to calculate an optimum sigma value for the Gaussian radial based function.

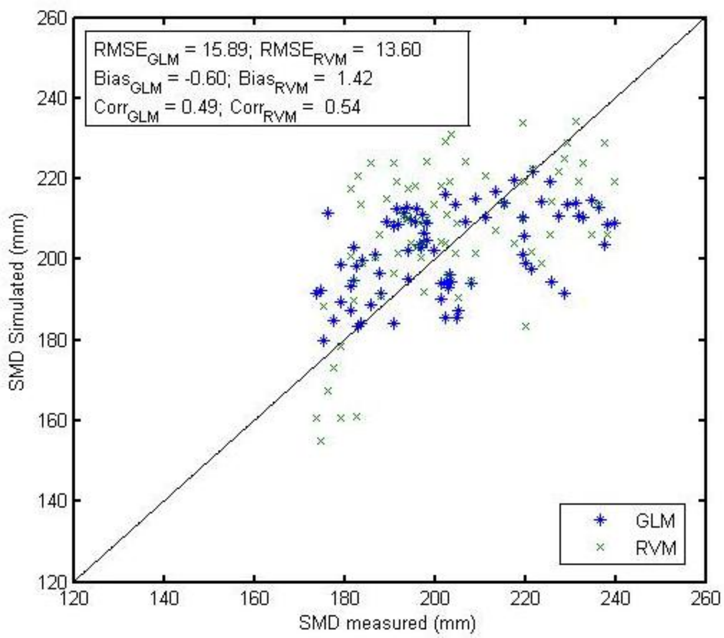
3.3 Performance of GLM and RVM for SMD estimation

For performance evaluation of the model, three performance statistics- *RMSE, Bias* and *correlation* were used to compare the GLM and RVM results with the PDM SMD (**see figure 2**). Looking at the distribution of the bias, the SMD was generally well simulated by using the RVM technique. The performance of RVM for SMD prediction during the calibration/validation exhibited much higher *r* (cal=0.73; val=0.54) in comparison to the GLM simulated SMD with *r* (cal=0.55; val=0.49). The *RMSE* value for RVM during calibration and validation were found as 15.64 and 13.60 respectively, while *Bias* statistics were found as -0.42 and 1.42 for the calibration and validation respectively. This statistics indicate that the value of the Bias was much higher in case of GLM simulated data, when compared to the RVM. The *RMSE* values reported from the GLM based technique were found during the calibration as cal=19.06 and val=15.89. Although, the RVM based technique for SMD estimated showed an improved performance over GLM simulated data, the *Bias* value indicated that the GLM have lower departure from the reference SMD as compared to the RVM.

The overall analysis of GLM and RVM for SMD estimation showed that the RVM was better than GLM during the both calibration and validation phases. This showed that the probabilistic method based scheme of RVM is more effective than point based method of GLM. Hence, RVM could be a better choice than point-based GLM for SMD estimation.



(a)



(b)

Figure 2 (a-b) GLM and RVM calibration and validation plots for SMD simulation

4. Conclusions

This study provided for the first time a comprehensive models evaluation for the SMD estimation by using the CCI soil moisture datasets. The RVM was found efficient for SMD estimation in comparison to the GLM. It was also observed that RVM outperformed GLM in terms of the performance statistical used in this study, except Bias. Results from this study will potentially help in the estimation of more accurate SMD information from satellite soil moisture data, and will also help to increase the efficiency of such data applicability in hydrological modelling. This work provides hydrologists with valuable information on CCI soil moisture dataset and its applicability for hydrological application such as SMD in this case. However, further exploration of this potentially valuable data source by the hydro-meteorological community has been recommended for extensive analysis of the techniques and datasets for other geographical regions.

Acknowledgments

The first authors would like to thank the Indian Meteorological Department for providing the ground hydro-meteorological dataset for this research. The authors would like to acknowledge the Central Water Commission for providing the flow gauge measurements. Authors would like to thanks the UGC for providing funding for this research.

References

Al-Shrafany, D., Rico-Ramirez, M.A., Han, D., 2012. Calibration of roughness parameters using rainfall–runoff water balance for satellite soil Moisture retrieval. Journal of Hydrologic Engineering 17, 704-714.

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sensing of Environment 162, 380-395.

Johnson, R.A., Wichern, D.W., 2002. Applied multivariate statistical analysis. Prentice hall Upper Saddle River, NJ.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Chapman & Hall/CRC.

Mladenova, I., Lakshmi, V., Jackson, T.J., Walker, J.P., Merlin, O., de Jeu, R.A., 2011. Validation of AMSR-E soil moisture using L-band airborne radiometer data from National Airborne Field Experiment 2006. Remote Sensing of Environment 115, 2096-2103.

Moore, R., 2007. The PDM rainfall-runoff model. Hydrology and Earth System Sciences Discussions 11, 483-499.

Owe, M., de Jeu, R., Walker, J., 2001. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. IEEE Transactions on Geoscience and Remote Sensing 39, 1643-1654.

Scott, C.A., Bastiaanssen, W.G., Ahmad, M.-u.-D., 2003. Mapping root zone soil moisture using remotely sensed optical imagery. Journal of Irrigation and Drainage Engineering 129, 326-335.

Srivastava, P.K., 2013. Soil moisture estimation from SMOS Satellite and mesoscale model for hydrological applications. University of Bristol, Bristol, UK.

Srivastava, P.K., Han, D., Ramirez, M.A.R., Islam, T., 2013a. Appraisal of SMOS soil moisture at a catchment scale in a temperate maritime climate. Journal of hydrology 498, 292-304.

Srivastava, P.K., Han, D., Ramirez, M.R., Islam, T., 2013b. Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. Water Resources Management 27, 3127-3144.

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., Al-Shrafany, D., Islam, T., 2013c. Data fusion techniques for improving soil moisture deficit using SMOS satellite and WRF-NOAH land surface model. Water Resources Management 27, 5069-5087.

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., O'Neill, P., Islam, T., Gupta, M., 2014. Assessment of SMOS soil moisture retrieval parameters using tau–omega algorithms for soil moisture deficit estimation. Journal of Hydrology 519, 574-587.

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., O’Neill, P., Islam, T., Gupta, M., Dai, Q., 2015. Performance evaluation of WRF-Noah Land surface model estimated soil moisture for hydrological application: Synergistic evaluation using SMOS retrieved soil moisture. Journal of Hydrology 529, 200-212.

Srivastava, P.K., Petropoulos, G.P., Kerr, Y.H., 2016. Satellite soil moisture retrieval: techniques and applications. Elsevier Press, United States. ISBN: 9780128033883

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. The Journal of Machine Learning Research 1, 211-244.