On the Application of Probabilistic
Hydrometeorological Simulation of Soil
Moisture across Different Stations in India

Sarit Kumar Das, Rajib Maity
Department of Civil Engineering, Indian Institute of Technology, Kharagpur, India
Email: saritkdas@iitkgp.ac.in

Received March 2014

Abstract

An application of a proposed hydrometeorological approach for probabilistic simulation of soil moisture is carried out. The time series of in-situ soil moisture and meteorological variables at monthly scale from a few monitoring stations having different soil-hydrologic properties across India are utilized. Preliminary investigation with both precipitation and near-surface air-temperature as meteorological variables to establish that the strength of association between soil moisture and precipitation is more significant as compared to that between soil moisture and temperature. Precipitation-based probabilistic estimation of soil moisture using the proposed hydrometeorological approach is tested with in-situ observed soil moisture, CPC model output and with soil moisture data of the Climate Change Initiative (CCI) project. The parameter of the developed model is linked to the soil-hydrologic characteristics through Hydrologic Soil Group (HSG) classification. Higher values of model parameter (dependence parameter \( \theta \)) for the selected copula correspond to HSG A and B having higher soil porosity, whereas, lower values correspond to HSG B and C having lower soil porosity.

Keywords

Soil Moisture, Probabilistic Modelling, Copula, Hydrometeorology, Hydroclimatologist

1. Introduction

Soil moisture has significant impact on temperature-evaporation-precipitation feedback loop (Seneviratne et al., 2010) and plays a significant role in numerical weather prediction using climate variables at regional scale (Ek & Hostslag, 2004; Drusch, 2007; Panareda et al., 2010). Soil moisture also controls the terrestrial water balance through partitioning precipitation among infiltration, runoff and evapotranspiration. The capillary action that determines the evaporative demand and withdrawal of water through plant roots is driven by soil moisture content (Entekhabi et al., 1996). Thus, soil moisture is also critical for drought, flood prediction, erosion caused by surface runoff, ground water recharge etc. (Dripps & Bradbury, 2007; Seneviratne et al., 2010). Amount of moisture content in the root zone is also significant, especially during the critical growth period of the crops (Narasimhan & Srinivasan, 2005). The temporal distribution of soil moisture over a catchment is useful for
forecasting irrigation requirements (Pandey & Pandey, 2010). Recently, United Nations Framework Convention on Climate Change (UNFCCC) has declared soil moisture as Essential Climate Variable (ECV) in the terrestrial domain and taken steps for monitoring soil moisture during 2010-15 (GCOS, 2010). Climate variables, especially, precipitation and temperature, have strong association with spatio-temporal distribution of soil moisture. It has been revealed that the temporal variation in the relationship between soil moisture and precipitation is non-linear and stochastic (Rodriguez-Iturbe et al., 2001). The factors behind the spatio-temporal variability of soil moisture, especially for topsoil, can be of two types—static and dynamic. Static factors include properties of soil, topography, land use etc. and dynamic factors include hydroclimatic parameters, such as precipitation, evapotranspiration, vegetation growth etc. (Rosenbaum et al., 2012). Sensitivity of soil moisture over changing frequency of precipitation and rise in temperature due to climate change scenarios is found to vary with soil texture (Bormann, 2012).

At point locations, spatial variation of soil moisture primarily depends on soil structure defined by soil texture (Gaur & Mohanty, 2013), plant characteristics (Porporato et al., 2004), whereas the temporal variability depends on hydrometeorological variables, like precipitation (Oyedele & Tijani, 2010). The classification of various soil types based on these soil-hydrologic factors is defined in the US National Engineering Handbook as Hydrological Soil Groups (HSG) as four major groups, viz., A, B, C and D. The rate of infiltration, i.e. the porosity of soil varies high to low from HSG A to HSG D (USDA, 2009).

Hydrometeorological approaches for modelling the variation of different hydrologic variables are gaining popularity (Maity & Kashid, 2011). With the availability of extensive hydroclimatic database, there is a scope for developing such a probabilistic model for soil moisture simulation prompted by the established physical interactions among available soil moisture and hydrometeorological variables responsible for its spatio-temporal variability. Such a model would cater the stochastic behaviour of these relationships and may yield better estimation with quantification of associated uncertainty. Moreover, developed approach can also be used to estimate the future soil moisture using the climate variables, simulated by different climate models.

Joint probability distribution between soil moisture series and influencing hydroclimatic variable is adopted for probabilistic model development. Copulas are found to be the best choice to develop the joint distribution by joining the individual marginal distribution of any two or more variables (Nelsen, 2006). These functions are recently used in several studies to obtain joint distribution among hydrological or climatological variables (Kao & Govindaraju, 2008; Maity & Nagesh Kumar, 2008; Das & Maity, 2013).

This applies a hydrometeorological approach, which can be considered as an alternative method for estimating soil moisture using hydrometeorological variables. The scope of this paper is to investigate the association between surface soil moisture and two hydrometeorological variables—precipitation and temperature. A copula-based bivariate probabilistic modelling approach for simulation of soil moisture is adopted.

The modelled soil moisture data are compared with in-situ data to check the accuracy of proposed simulation approach during both model development and testing periods. The performance of the proposed approach is also tested with the surface soil moisture data set for depth 0 - 5 cm developed under the Climate Change Initiative (CCI) project funded by ESA (ESA CCI). The Climate Prediction Centre’s (CPC’s) leaky bucket model (Huang et al., 1996) based soil moisture estimates (Fan & Dool, 2004) are also considered to check the accuracy of the model output. Finally, an attempt is also made to check the relationship of the model properties (in terms of model parameter) with the soil hydrologic characteristics in terms of HSGs at the monitoring locations.

2. Data Used and Preliminary Analysis

2.1. Data

Thirteen locations across different HSGs in India, as shown in the HSG map for India (Figure 1) are selected. The weekly in-situ surface (for depth 0 - 7.5 cm) soil moisture data for bare soil during the years 1991 to 2006 is procured from National Data Centre (NDC) of Indian Meteorological Department (IMD), Pune, India. The daily precipitation and mean daily temperature data are also obtained from IMD. Precipitation data are converted into the cumulative monthly values, and soil moisture as well as temperature data into monthly average values. There exist some missing data periods for some of the variables considered. These time instances are excluded from the analysis.

ESA CCI data, as mentioned before, is obtained from
The ESA CCI soil moisture data is available at 0.25° by 0.25° geo-spatial grid (Albergel et al., 2013a), and so the Inverse Distance Weighting (IDW) method is adopted to estimate the values at the station locations. Finally, the information on HSG classification is obtained from Central Ground Water Board of India (CGWB, 2007).

### 2.2. Preliminary Analysis

A preliminary analysis is carried out to investigate the association of surface soil moisture data with meteorological variables—temperature and precipitation, individually. Scatter plots (not presented) between soil moisture and precipitation or near surface air temperature indicate a stronger association between soil moisture and precipitation than that between soil moisture and temperature, which associations may not necessarily be linear.

### 2.3. Assessment of Linear Association

Partial correlation coefficients are calculated to check the strength of the linear association of soil moisture on precipitation (or near surface air temperature), while the effect of temperature (or precipitation) is partial out. The first order partial correlation coefficient between $x$ and $y$, while the effect of $z$ is partial out is expressed as:

$$r_{xy1} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

where, $x$, $y$ are primary variables and $z$ is the variable, which is partial out; $r_{xy}$, $r_{xz}$ and $r_{yz}$ are the correlation coefficients between the variables given in the subscript (Ley, 1973). Values of first order partial correlation coefficients along with their significance level (95%) are presented in Figure 2(a). It is found that at all stations, first order partial correlation coefficient between soil moisture and precipitation is positive, statistically significant (at 95% significance level) for all the stations and varies over a range of 0.507 to 0.861. However, the partial correlation coefficient for soil moisture and temperature is significant only at Anakapalle, Durgapura, Nalgur, Pillamedu and Vellayani. Thus, the linear association between soil moisture and precipitation is
found stronger than that between soil moisture and temperature. The strength of association is also investigated through a scale-free measure of association.

2.4. Assessment of Scale-Free Association

Rank-based, scale-free measures of association such as, Kendall’s Tau (τ) or Spearman’s Rho (ρ) are independent of the individual marginal distribution of the variables, and thus, useful for determining joint distribution without compromising the dependence structure during simulation (Nelsen, 2006). In this study, Kendall’s Tau (τ) is adopted as a measure of such association since it directly determines the dependence parameter for the copula.

Partial Kendall’s Tau can be treated as the association between two variables while the effect of third variable is partial out. This is expressed as (Kendall, 1962):
where, $\tau_{xy}$, $\tau_{xz}$ and $\tau_{yz}$ are the Kendall’s Tau ($\tau$) between the variables shown in the subscript. Partial Kendall’s Tau values, between soil moisture and precipitation while the effect of temperature is partial out, and between soil moisture and temperature while the effect of precipitation is partial out, are computed for all stations and are shown in Figure 2(b). At all stations, partial Kendall’s Tau for soil moisture and precipitation relationship is positive and significant at 95% confidence level, however the soil moisture-temperature relationship is significant only at three (Nagpur, Pune, Vellayani) (Figure 2(b)). Thus, based on the linear association and scale-free measures of association, only precipitation is considered hereinafter for simulation of soil moisture.

3. Methodology

To remove the effect of seasonality, anomaly values of both monthly precipitation and soil moisture are used in the simulation model. Anomalies for monthly average of soil moisture and monthly cumulative values of precipitation are obtained by deducting the long-term monthly mean from the observed values of the variable for each month in a calendar year. Kolmogorov-Smirnov test is applied to check the fitted distribution and to select the best one from the tentatively eligible pool of distributions.

3.1. Estimation of Joint Probability Distribution

Marginal distributions are used to determine the joint distribution by employing a copula based approach. In this study, three different bivariate copula functions viz., Clayton, Frank, and Gumbel-Hougaard (belong to the Archimedean Class of copulas) are tested. Here, another commonly used bivariate copula, Ali-Mikhail-Haq, is not adopted as the acceptable range of dependence for this copula is narrow ($\tau = -0.182$ to 0.333) (Chowdhary et al., 2011), and not suitable for most of the monitoring stations. Details of these copula functions can be found in Nelsen (2006). First, the tentatively eligible copulas are selected based on the Kendall’s Tau ($\tau$) value calculated for association between soil moisture anomaly and precipitation anomaly, separately for each station. When there exists more than one feasible copula family that satisfy the range of dependence for the given data, the final selection of a suitable copula is based on the best fit to the observations. This fit is assessed statistically by goodness-of-fit (GOF) tests prescribed by Genest et al. (2009).

The copula showing best fit (denoted as $C^b$) is selected to obtain the joint distribution between soil moisture and precipitation anomaly. The developed joint distribution between soil moisture anomaly and precipitation anomaly, denoted as $F_{S,P}(s,p)$. This is expressed as:

$$F_{S,P}(s,p) = C^b(F_S(s), F_P(p)).$$

where, $S$ and $P$ are the soil moisture anomaly and the precipitation anomaly respectively; $s$ and $p$ are values.

3.2. Determination of Conditional Distribution Function

The joint distribution, $F_{S,P}(s,p)$ is used to get a conditional cumulative distribution for the soil moisture anomaly conditioned on the precipitation. The formulation as given by Schmidt (2006):

$$F_{S|P=p}(S \leq s \mid P = p) = C^b_1(F_S(s), F_P(p)).$$

For a particular value of observed precipitation anomaly, this conditional cumulative distribution function (Equation (4)) is used for probabilistic simulation of soil moisture anomaly. This simulated soil moisture anomaly is then added to the long-term monthly mean of soil moisture to obtain the probabilistic distribution of simulated soil moisture.

3.3. Model Performance Evaluation

The simulated soil moisture using the methodology proposed here is first compared with in-situ observed data. Further, the proposed model output is compared with the ESA CCI surface soil moisture data set, and CPC model soil moisture data. The statistical comparison, between Proposed model output and CCI soil moisture date
is made in terms of standard and most updated measures for model performance, namely, Correlation Coefficient (CC), normalized unbiased Root Mean Square Error (ubRMSE) (Albergel et al., 2013b), and Refined Index of Agreement ($d_r$) (Willmott et al., 2012) etc. The ubRMSE is normalized using the range of in-situ or observed data.

4. Results and Discussions

4.1. Development of Simulation Model

A few tentatively eligible probability density functions (pdf) with relatively high Kurtosis, such as, Normal and Logistic distribution, which are applicable for the data having both positive and negative values, are tested for the anomaly of average monthly soil moisture and precipitation. Goodness-of-fit tests are carried out using Kolmogorov-Smirnov test. These p-values indicate (not presented) that at majority of stations Logistic distribution is best among the distribution functions tested. At nine out of 13 stations for precipitation anomalies and 12 out of 13 stations for soil moisture anomalies Logistic distribution is passing through the KS test. Thus, the probability distributions for both soil moisture and precipitation are accepted to follow Logistic distribution.

The results of copula selection procedure reveal that Clayton is the best copula for deriving the joint distribution of soil moisture and precipitation anomalies. Simulation of soil moisture using Clayton copula is carried out for all the stations, across different HSGs. For discussion, the results (time series plots) obtained at four stations, one from each HSGs (Bellary, Udaipur, Anakapalle and Nagpur) are presented in Figure 3 (left panel) respectively. Probabilistic estimation of soil moisture is represented through 90% confidence interval (the interval area is shown in grey) along with the expected values (50th quantile) in black line. Observed in-situ soil moisture (red), satellite based ESA CCI soil moisture values (green) and CPC model estimated values (dark blue) are also shown in these figures.

4.2. Checking of Model Performance

The simulated soil moisture (50th quantile value), based on the proposed methodology, is first compared with in-situ observed data. As noted earlier, the proposed model provides a probabilistic output and availability of uncertainty information is a strong point of the proposed model. From the probabilistic output of proposed approach, 5th quantile is used as Lower Limit (LL), 95th quantile as Upper Limit (UL). Comparison plots between estimated soil moisture (LL, UL and expected value), in-situ soil moisture data, ESA CCI data and CPC data are prepared. It is important to mention here that CPC model is taken as an existing model which simulates soil moisture based on precipitation and temperature, whereas, Satellite based ESA-CCI data are considered another observed data set for soil moisture to be compared with. Model performances are assessed through three statistics during both model development and testing periods. Results are presented in Table 1. The results reveal that at all monitoring stations, the proposed model performs reasonably well during both model development and testing period. While comparing with in-situ values during model development period, CC varies between 0.744 (New Delhi) and 0.904 (Pillamedu); normalized ubRMSE varies between 0.231 (Karnal) and 0.084 (Vellayani); $d_r$ ranges from 0.626 (Anakapalle) to 0.805 (Pillamedu). However, during testing period the CC values ranges between 0.999 (Bellary) and 0.752 (Pillamedu); normalized ubRMSE varies from 0.211 (Pillamedu) to 0.001 (Bellary); $d_r$ ranges between 0.871 (Vittal) and 0.707 (Nagpur). This relative improvement in model performance for some stations, like Bellary, during testing period can be attributed to the range of in-situ data during that time period. The performance of the proposed model against CPC model is assessed through the performance of the both against in-situ values. This assessment as shown in Table 1 reveals that the proposed model performs relatively better at all the stations during both model development and testing periods, except at station, Karnal during model development period. The significantly poor performance of the CPC model at a few stations—Anakapalle, Bellary, Hebbal etc. can be attributed to the fact that this model assumes a constant soil porosity values across the globe which can be far different from the in-situ value as in these cases. The performance of the proposed model against the ESA-CCI data is also found to be reasonably good at all stations, except Anakapalle. The CC values varies between 0.647(Hebbal) and 0.909 (Vittal), normalized ubRMSE varies from 0.243 (New Delhi) to 0.130 (Vittal). However, $d_r$ values found to be negative at almost all the stations, indicating the larger variability in the simulated soil moisture data compared to the ESA-CCI data set.
4.3. Model Parameter and Soil Hydrologic Properties

Model parameter ($\theta$) of the proposed copula-based soil moisture simulation model represents the dependence between soil moisture and precipitation anomalies. This parameter is applied to the joint distribution of soil moisture and precipitation anomalies. The values of this model parameter at all soil moisture monitoring stations are investigated for any possible link with the soil hydrologic characteristics at monitoring stations. This comparison (Table 2) provides the necessary evidence of the influence of soil properties on the variation of soil moisture due to variation in precipitation. The comparison, as parameter value is higher for stations belongs to HSG A and B, whereas it is lower for stations belongs to HSG C and D. The soil porosity is maximum for HSG A and gradually drops through HSGs. Thus it emphasizes the fact that the soil-hydrologic properties e.g. soil porosity has a control on the response of soil moisture variability due to hydrometeorological forcing.

5. Conclusions

In this paper, dependence of monthly soil moisture variation on hydroclimatic variables (precipitation and temperature) is explored. It is found that the strength of association is higher between soil moisture and precipitation...
Table 1. Model Performance during Model Development (top rows) and Model Testing (bottom rows).

<table>
<thead>
<tr>
<th>HSG</th>
<th>Station</th>
<th>In-situ vs. Proposed model</th>
<th>In-situ vs. CPC Model</th>
<th>CCI Data vs. Proposed Model</th>
<th>In-situ vs. Proposed model</th>
<th>In-situ vs. CPC model</th>
<th>CCI Data vs. Proposed Model</th>
<th>Refined Index of Agreement (dr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Bellary</td>
<td>0.821</td>
<td>0.209</td>
<td>0.678</td>
<td>0.143</td>
<td>0.367</td>
<td>0.222</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>Hebbal</td>
<td>0.999</td>
<td>0.471</td>
<td>0.665</td>
<td>0.001</td>
<td>0.700</td>
<td>0.29</td>
<td>0.734</td>
</tr>
<tr>
<td></td>
<td>Pillamedu</td>
<td>0.904</td>
<td>0.555</td>
<td>0.73</td>
<td>0.117</td>
<td>0.237</td>
<td>0.196</td>
<td>0.805</td>
</tr>
<tr>
<td></td>
<td>Durgapura</td>
<td>0.752</td>
<td>0.332</td>
<td>D.N.A.</td>
<td>0.211</td>
<td>0.442</td>
<td>D.N.A.</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>Karnal</td>
<td>0.810</td>
<td>0.18</td>
<td>0.723</td>
<td>0.146</td>
<td>0.737</td>
<td>0.222</td>
<td>0.711</td>
</tr>
<tr>
<td></td>
<td>New Delhi</td>
<td>0.744</td>
<td>0.273</td>
<td>0.671</td>
<td>0.174</td>
<td>0.271</td>
<td>0.243</td>
<td>0.688</td>
</tr>
<tr>
<td></td>
<td>Udaipur</td>
<td>0.850</td>
<td>0.491</td>
<td>0.793</td>
<td>0.137</td>
<td>0.269</td>
<td>0.188</td>
<td>0.773</td>
</tr>
<tr>
<td></td>
<td>Anakapalle</td>
<td>0.834</td>
<td>0.291</td>
<td>0.253</td>
<td>0.13</td>
<td>0.246</td>
<td>0.411</td>
<td>0.626</td>
</tr>
<tr>
<td></td>
<td>Vellayani</td>
<td>0.932</td>
<td>0.031</td>
<td>0.372</td>
<td>0.105</td>
<td>0.378</td>
<td>0.395</td>
<td>0.784</td>
</tr>
<tr>
<td></td>
<td>Vittal</td>
<td>0.968</td>
<td>0.601</td>
<td>0.774</td>
<td>0.084</td>
<td>0.287</td>
<td>0.212</td>
<td>0.889</td>
</tr>
<tr>
<td>B</td>
<td>Nagpur</td>
<td>0.865</td>
<td>0.778</td>
<td>0.909</td>
<td>0.139</td>
<td>0.228</td>
<td>0.13</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Pune</td>
<td>0.941</td>
<td>0.839</td>
<td>D.N.A.</td>
<td>0.11</td>
<td>0.234</td>
<td>D.N.A.</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>Adhartal</td>
<td>0.873</td>
<td>0.746</td>
<td>0.776</td>
<td>0.141</td>
<td>0.197</td>
<td>0.158</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>Vellayani</td>
<td>0.91</td>
<td>0.698</td>
<td>0.766</td>
<td>0.138</td>
<td>0.174</td>
<td>0.255</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>Vittal</td>
<td>0.758</td>
<td>0.681</td>
<td>0.869</td>
<td>0.166</td>
<td>0.201</td>
<td>0.166</td>
<td>0.686</td>
</tr>
<tr>
<td></td>
<td>Adhartal</td>
<td>0.841</td>
<td>0.519</td>
<td>0.762</td>
<td>0.149</td>
<td>0.302</td>
<td>0.207</td>
<td>0.707</td>
</tr>
<tr>
<td></td>
<td>Pune</td>
<td>0.861</td>
<td>0.643</td>
<td>0.816</td>
<td>0.128</td>
<td>0.193</td>
<td>0.21</td>
<td>0.761</td>
</tr>
</tbody>
</table>

D.N.A. = Data not available at the time step.

Table 2. Model parameter with different HSGs.

<table>
<thead>
<tr>
<th>HSG</th>
<th>Station</th>
<th>Model parameter (θ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Bellary</td>
<td>1.254</td>
</tr>
<tr>
<td></td>
<td>Hebbal</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>Pillamedu</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>Durgapura</td>
<td>1.086</td>
</tr>
<tr>
<td></td>
<td>Karnal</td>
<td>1.252</td>
</tr>
<tr>
<td></td>
<td>New Delhi</td>
<td>1.143</td>
</tr>
<tr>
<td></td>
<td>Udaipur</td>
<td>1.535</td>
</tr>
<tr>
<td></td>
<td>Anakapalle</td>
<td>0.803</td>
</tr>
<tr>
<td>B</td>
<td>Vellayani</td>
<td>0.391</td>
</tr>
<tr>
<td></td>
<td>Vittal</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>Adhartal</td>
<td>1.767</td>
</tr>
<tr>
<td></td>
<td>Nagpur</td>
<td>0.496</td>
</tr>
<tr>
<td></td>
<td>Pune</td>
<td>0.288</td>
</tr>
</tbody>
</table>
comparing to that between soil moisture and temperature. Considering the associated uncertainty, a probabilistic model is applied for soil moisture estimation along with uncertainty quantification. The monthly soil moisture computed with the proposed model is found to be promising. Observed in-situ soil moisture variation is found to be well captured between 5th (upper limit) and 95th quantile (lower limit) provided by the proposed approach. 50th quantile of the probabilistic estimate is found to well correspond with the in-situ values for all the stations. The comparison is also carried out between soil moisture estimate by the proposed model and CPC model as well as the ESA CCI soil moisture data. A fairly good correspondence is indicated for all the stations. Finally the relationship between model parameter value and the hydrologic soil properties at monitoring location indicates the possible spatial transferability of the proposed model.

Acknowledgements

We acknowledge financial support from ISIRD fund of Indian Institute of Technology, Kharagpur through the project, code FSH.

References


