Chapter 5 : SPATIAL DOWNSCALING OF SOIL MOISTURE DATA USING A NEWLY DEVELOPED VEGETATION MODULATED SOIL MOISTURE INDEX

5.1 INTRODUCTION

Soil moisture is a necessary variable for many atmospheric and hydrological processes, including the exchange of heat energy between the Earth's surface and near atmosphere, water cycle, and groundwater recharge. Soil moisture information is also widely employed in various domains such as water resource management, weather forecasting, drought analysis, agriculture production, irrigation management, and many other fields. Therefore, it is significant to assess the complete and precise information of soil moisture content spatially and temporally. Remote sensing is a more convenient and proficient way to obtain soil moisture information than in-situ measurements. Remote sensing also has different ways to monitor soil moisture, such as airborne mapping via active or passive sensors [128, 129] and space-borne or satellite observations [9, 130, 131]. The space-borne techniques are mandatory to monitor soil moisture globally on a regular basis [132, 133].

Various studies have been introduced to rescale the soil moisture up to 1 km. Colliander et al. [65] used the Dispatch algorithm to disaggregate soil moisture up to 1 km, initially developed by Merlin et al. [63]. Kim et al. [134] used the triangle method with a polynomial relation of soil moisture with LST and NDVI of order 2 and Support Vector Machine (SVM) for soil moisture disaggregation. In contrast, Peng et al. [133] introduced a different algorithm based on a parameter VTCI (Vegetation Temperature Condition Index) for the same purpose. Meanwhile, Bai et al. [131] and Chen et al. [7] have utilized machine learning approaches. The Dispatch method is a more physical algorithm than the triangle and machine learning approaches [135] and is based on a proxy SEE. But it is very complex to formulate because of the selection of different zones based on the vegetation conditions and the estimation of vegetation temperature.

Apart from this, the Triangle method is simply a polynomial relation of soil moisture with some land surface parameters (LST, NDVI, EVI (Enhanced Vegetation Index), albedo, etc.) [64, 66, 136]. But this method may also include some bias errors because of the regression coefficients. Therefore, this study introduces a relatively simple, computationally proficient, and vegetation modulated disaggregation algorithm for soil moisture to accomplish these issues. Firstly, a parameter that indicates the Earth's surface reflectance formulated in terms of MODIS LST is used in a relation similar to Peng et al. [133] to obtain the soil moisture at high resolution. But this parameter includes the contribution of soil and vegetation, both. Therefore, a new parameter is introduced for the modulation of the vegetation layer, which is a function of NDVI and is termed the vegetation parameter. It is used as a subtraction factor from the previously formulated parameter (function of MODIS LST) to derive a new soil moisture index which is vegetation modulated.

Zeng et al. [137] also developed a different soil moisture index by taking a ratio of two temperature differences for the same value of NDVI. But it may include some errors in the representation of the soil along with the increase of NDVI values because it provides different relations for different values of NDVI but does not exclude the vegetation contributions. This index may be suitable for estimating soil moisture content, but for downscaling, a valid signature of soil is required. Therefore, this study introduced a new and different soil moisture index that can significantly modulate the vegetation contributions and be applied to downscale the soil moisture product. The final equation of downscaling provides a relation between coarse and high-resolution soil moisture through the ratio of coarse and high-resolution Vegetation Modulated Soil Moisture Index (VMSMI).

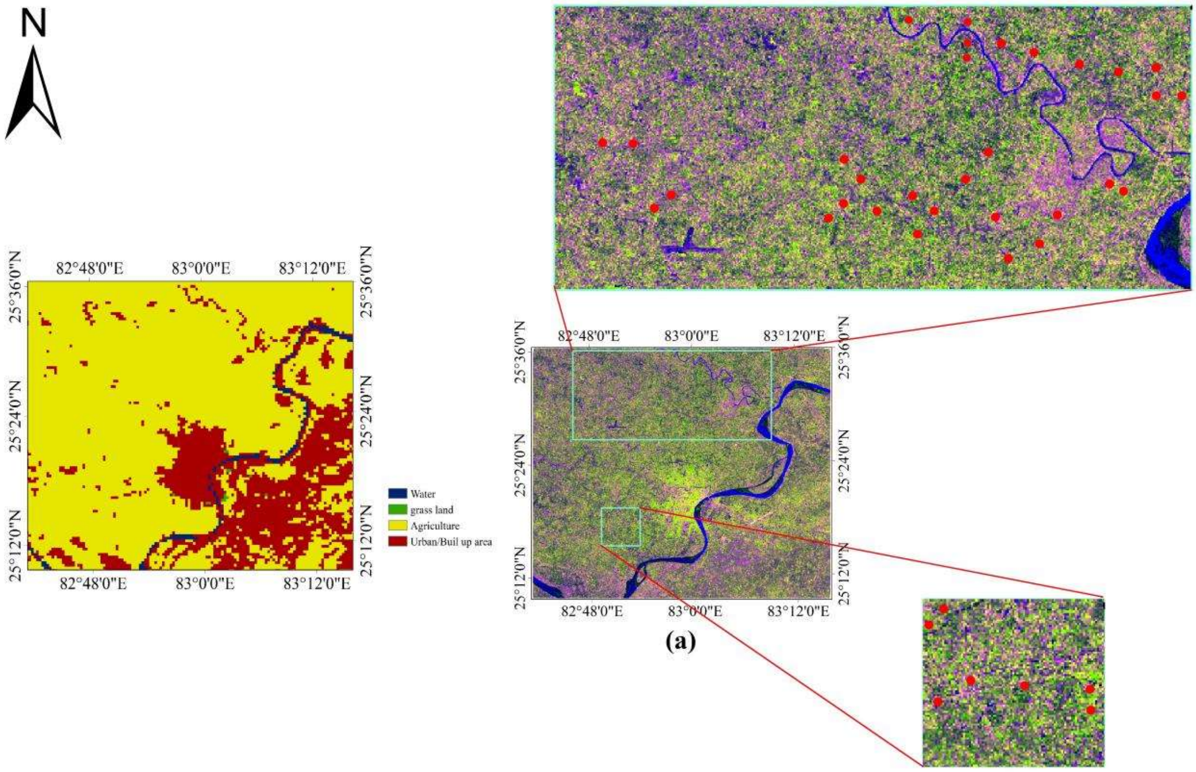

Figure 5.1 Description of the ground sampling plan (a) The true colour image with the location of sampling points (red dots), (b) MODIS land cover map of the selected region.

5.2 DATASETS

5.2.1 In-situ measurements

Two green rectangles over the RGB image in Figure 5.1(a) denote the selected region for the ground data sampling, and solid red circles represent the data collection sites. The area of the field sampling was fully covered by agriculture, and wheat was the dominating crop. The field sampling campaign was intended for seven months, from October 2017 to April 2018, and the ten dates were chosen at the interval of 10-15 days. The months (May to September), which belong to the monsoon season or have some effects of clouds, were excluded from the sampling plan. The dates were also chosen according to the clear sky condition. A soil sensor named Hydra Go was used to measure soil moisture up to the depth of 5 cm. It is designed with a portable soil moisture probe with a length of 5 cm and wireless technology to stabilize a smartphone connection for data values.

5.2.2 Space-borne data

This study used two space-borne (SMAP and MODIS) datasets to monitor and downscale soil moisture. The SMAP data product SPL2_SM_P_E [30] provides soil moisture values daily at the 9 km scale, which is used to rescale up to 1 km. This product is an enhanced form of the SPL2_SM_P data product, which provides soil moisture daily at 36 km. The operational frequency of SMAP ranges from 1 to 2 GHz (L-band of microwave region) due to the transparency of low frequency through clouds and vegetation, which leads to the highest accuracy for monitoring soil moisture.

The data products MYD11A1 and MOD09GA of MODIS provide daily LST at 1 km and daily surface reflectance for seven bands at a spatial scale of 500 m, respectively. The band-1 (Red) and band-2 (NIR) obtained from the product MOD09GA were utilized to calculate the daily NDVI values. Though the dates of this study are chosen according to the cloud-free condition still to eliminate a small effect of clouds from the data products provided by the MODIS LST, the quality flag information provided by the MODIS LST data product was used. The quality flag information is used to correct the MODIS LST images by excluding the low-quality pixels. The low-quality pixels have errors higher than the acceptable error in the LST values (±2 K).

The satellite datasets described above were downloaded for the 10 dates according to the ground sampling plan, starting from 14 October 2017 to 18 April 2018, as described in Section 2.5.1.

5.3 METHODOLOGY

- **Proposed downscaling method**

Based on the features of LST and NDVI, a new technique has been developed for the downscaling of satellite soil moisture, which is simple and dependent on the coarse resolution soil moisture, LST and NDVI. The low-resolution soil moisture is converted into high-resolution soil moisture by incorporating the values of LST and NDVI. Basically, LST indicates the Earth's surface's energy balance and is estimated from the MODIS thermal frequency data for the entire Earth's surface [138]. So, it can be a good signature of land surface cover through.

$$T^* = \frac{LST_{max} - LST}{LST_{max} - LST_{min}} \qquad (5.1)$$

Similar to the formula given by Peng et al. [133], T^* is used here to obtain the high-resolution soil moisture from the low-resolution soil moisture as follows,

$$SM_{high} = SM_c \frac{T^*_{high}}{T^*_c} \qquad (5.2)$$

Here, SM stands for Soil Moisture, subscript *high*, and *c* denotes the higher and coarse spatial resolution. Since T^* represents all the existing land cover over our study area and therefore the downscaled soil moisture using the above parameter may have some bias/error in the output due to the signal attenuation by the vegetation. Thus, this study introduced another term to modify the formula for $T^*$to minimize the vegetation effects and better estimate the soil moisture. Which is as follows,

$$F_v = \frac{NDVI_{max} - NDVI}{NDVI_{max} - NDVI_{min}} \qquad (5.3)$$

$$SM_{high} = SM_c \frac{T^*_{high} - F_{v\,high}}{T^*_c - F_{v\,c}} \quad (5.4)$$

Where

$$VMSMI\ (Vegetation\ Modulateded\ Soil\ Moisture\ Index) = \frac{T^*_{high} - F_{v\,high}}{T^*_c - F_{v\,c}} \quad (5.5)$$

Here, VMSMI is obtained after subtracting the vegetation parameter from T^*. It reduces the vegetation effect so VMSMI can represent only the soil surface. The equation (5.4) is then used for downscaling the soil moisture.

5.4 RESULTS AND DISCUSSION

5.4.1 Relationship of SMAP soil moisture with MODIS LST and NDVI

Figure 5.2 shows the correlation matrix plot for MODIS LST and NDVI with SMAP soil moisture for the selected region and time series. MODIS LST and NDVI are extracted from the MYD11A1 and MOD09GA data products of MODIS, respectively, having a pixel size of 1 km and 500 m, respectively. Therefore, these variables were previously upscaled to 9 km, equivalent to the SMAP pixel. The r and rs in the Figure 5.2, represent the linear and non-linear correlation coefficient and p denotes the level of significance. The diagonal images show the histogram of each variable. The upper-triangular images show the scatter plot between corresponding horizontal and vertical data product whereas, the lower-triangular images show the statistical analysis of these variables. The bar plots demonstrate that the minimum and the maximum values of SMAP soil moisture are 0.05 m3/m3 and 0.25 m3/m3, respectively. The values of MODIS LST lie between 288 K to

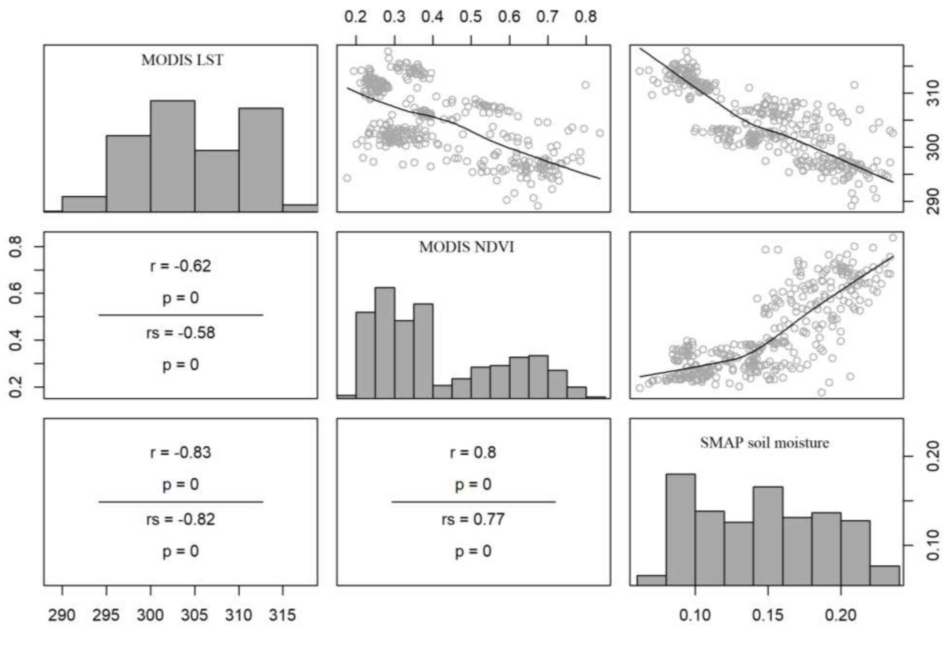

Figure 5.2 Correlation matrix plot for MODIS LST and NDVI with SMAP soil moisture.

318 K, and MODIS NDVI ranges from 0.19 to 0.82 for the chosen study area and throughout the selected months. Usually, LST shows a negative, and NDVI shows a positive correlation with the soil moisture, as found in the results. Both products show an

excellent correlation to the SMAP soil moisture with the correlation coefficients of - 0.83 and 0.80, respectively. The LST and the NDVI show a higher correlation value for linear behaviour with SMAP soil moisture than the non-linear behaviour. The correlation analysis illustrates that LST and NDVI are relevant factors to predict or downscale the soil moisture content.

Figure 5.3 shows the temporal variation of average MODIS LST and NDVI from October 2017 to April 2018 for the chosen study area. The MODIS LST and NDVI were averaged to a single value for each date and study area. The LST started decreasing in November 2017 and is found to be lowest on 16 January 2018 due to the winter season, then again rising up to April 2018. Whereas the value of NDVI first decreases, then slightly increases due to the harvesting of paddy crop and sowing of wheat crop in November 2017. The highest observed value of NDVI was found in February because of the fully matured wheat crop and then again decreased up to its lowest value in April due

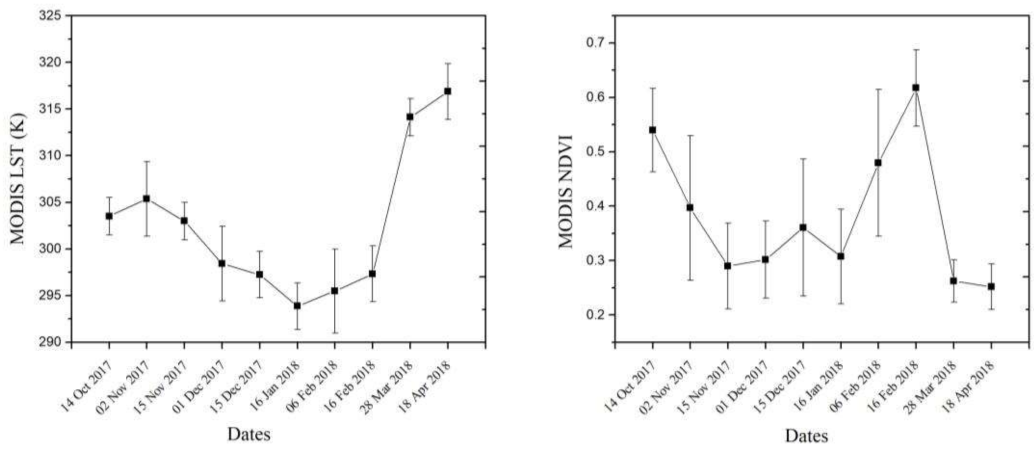

Figure 5.3 Temporal variation of average MODIS LST and NDVI for the chosen study area and time series.

to the harvesting of the wheat crop.

5.4.2 Evaluation of SMAP soil moisture with in-situ measurements

Before applying the disaggregation approaches to the SMAP soil moisture data product, the SMAP soil moisture is first validated against ground truth soil moisture values over the selected study site. Figure 5.4 shows the scatter plot between SMAP soil moisture and in-situ soil moisture observations. SMAP soil moisture is available on a daily basis, and the in-situ observations used here are according to the clear sky conditions. The in-situ observations are averaged over a 9 km pixel according to the SMAP spatial resolution to validate soil moisture data values. The SMAP soil moisture agrees well with the ground truth soil moisture observation, and the data points are scattered well across the one-one line with a correlation coefficient of 0.56, and obtained ubRMSE is 0.070.

In addition, the time series of average SMAP soil moisture is also analysed along with the average ground measurements. Figure 5.5 shows the variation of average SMAP and ground truth soil moisture from October 2017 to April 2018. The SMAP soil moisture data values seem to be analytically less than the ground measured soil moisture values. The SMAP and the ground measured soil moisture show a good seasonal variation with minimum soil moisture in summer (March and April) and maximum soil moisture in winter (January and February). The variation in soil moisture values for both datasets is almost identical except for the December month. The SMAP soil moisture gives similar values to the in-situ soil moisture in October, November, March, and April. The overall

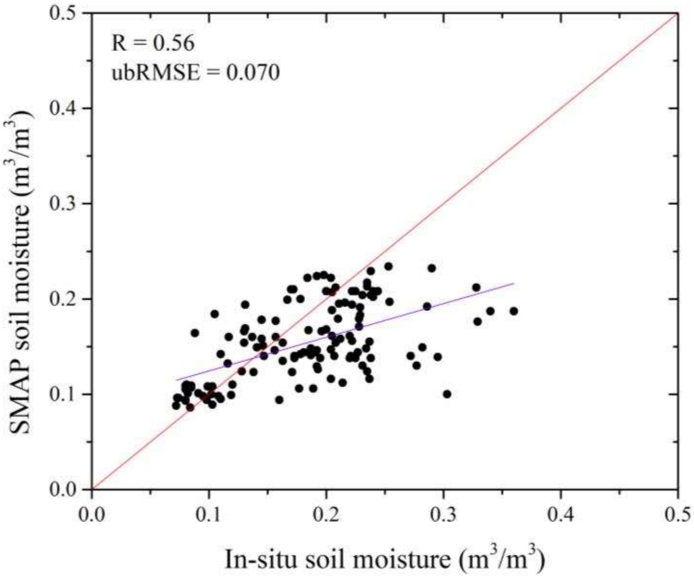

Figure 5.4 Scatter plot between SMAP soil moisture and the in-situ soil moisture observations.

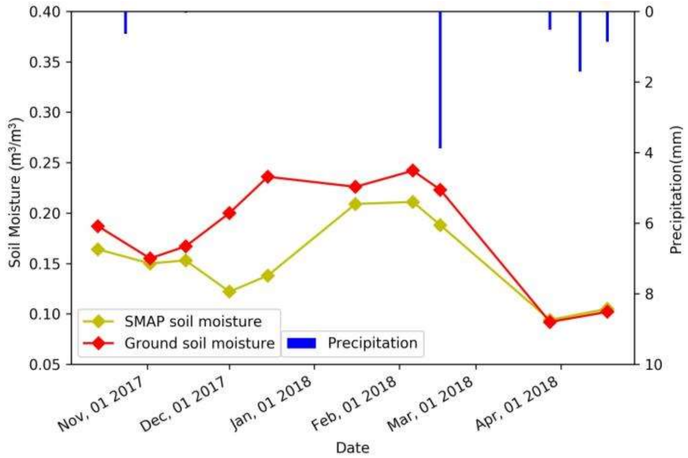

Figure 5.5 Temporal variation of average SMAP and ground truth soil moisture with precipitation over selected study area from October 2017 to April 2018.

results demonstrate the validity of SMAP soil moisture for the selected study region and in-situ measurements, leading to SMAP data products' applicability for the downscaling algorithms.

5.4.3 Evaluation of downscaled results with the in-situ measurements

The performance of four downscaling methods is displayed in the form of the Taylor diagram [139] in Figure 5.6. It summarized the statistics for the disaggregated soil moisture using different downscaling methods compared to the ground soil moisture. The white circle on the x-axis is the reference point and shows the observed data values. The closest point to this reference point indicates the absolute fit between the approach's results and the observed data. The radial line goes through the reference point denotes the standard deviation for the observed datasets; if the data point lies after this radial line, it shows the overestimation and below the line shows the underestimation concerning the observed values, and the closest point denotes the more relative variability to the reference data. The concentric radial lines with the reference point as a center indicate the Root Mean Square Difference (RMSD), and the arc of the plot shows the correlation coefficient. It goes from vertical radius to horizontal radius. A notable observation throughout the Taylor diagram is that all data points lie under the standard deviation line and the red mark (approach with VMSMI) is the closest to it. In the case of RMSD, all data points lie between the RMSD lines of 0.05 and 0.1; the red and blue dots are closer to the white circle than the other, while the former shows the lowest RMSD. In terms of the correlation

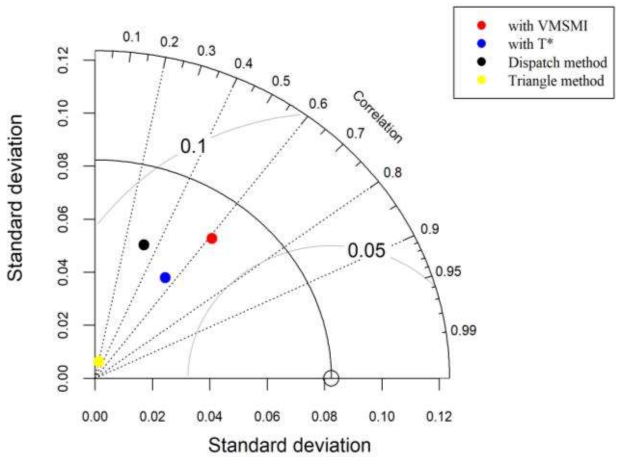

Figure 5.6 The Taylor diagram presenting the comparison of different downscaling algorithms.

coefficient, the data point belonging to the VMSMI-based technique is observed to be above the 0.6 value, while the other of the points belonging to the other ways are found below the 0.6 value, and the Triangle method shows the lowest correlation coefficient. From the overall results, the red data point is found the closest to the referenced datasets; therefore, it can be stated that the results acquired by the VMSMI-based technique perform substantially better than others.

The performance of downscaling approaches and the 9 km SMAP soil moisture product is further examined by several statistical variables such as correlation coefficient (R), ubRMSE, bias, and slope (Table 5.1) [140]. The VMSMI-based approach shows the uppermost value of the correlation coefficient (0.61) and the lowermost values of bias (-0.004) and ubRMSE (0.068). On the other side, VMSMI also provides a slope with a value closer to 1 than other methods. The statistical analysis results show that VMSMI based approach performs better than the other methods and also coarse resolution SMAP soil

Table 5.1 Results comparison summary for the 1 km downscaled soil moisture and SMAP 9 km soil moisture vs. in situ soil moisture.

Statistical variable	Dispatch method	Triangle method	Using T^*	Using VMSMI	SMAP 9 km soil moisture product
Correlation coefficient (R)	0.32	0.20	0.55	0.61	0.56
ubRMSE	0.084	0.082	0.075	0.068	0.070
Bias	-0.024	-0.022	-0.02	-0.004	-0.030
Slope	0.20	0.16	0.39	0.60	0.58

moisture. The results of the scatter plot (Figure 5.7) for different approaches are ordered as the lowest to highest slope value. This demonstrates that the minimum slope value is obtained for the Triangle method, whereas the maximum slope value is obtained for the VMSMI-based method. It shows that the Triangle method doesn't scatter much and is spread only in the range of 0.14 to 0.185. The scatter plots corresponding to Dispatch, *T**, and VMSMI show a better distribution of data points across the one-one line but in the case of slope and correlation, the Dispatch and *T** based method provides only satisfactory results, whereas VMSMI gives better values for slope and correlation coefficient. Also, the VMSMI-based method shows a higher slope and performs well compared to the scatter plot of coarse resolution SMAP soil moisture (Figure 5.4).

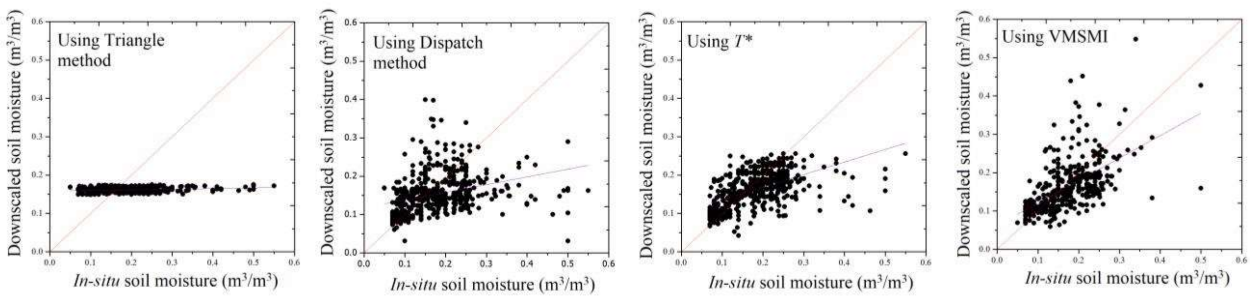

Figure 5.7 Scatter plots of downscaled soil moisture estimates based on different downscaling methods at 1 km vs. in situ soil moisture measurements.

The downscaled results are again tested by plotting a Cumulative Probability Function (CPF) against volumetric soil moisture (Figure 5.8). This function scaled the probability of soil moisture values from 0 to 1, and the probability of a data value may be calculated by subtracting the probability of the preceding data value. The probability distribution plots for different downscaling methods and the ground truth soil moisture illustrate that the VMSMI-based approach shows the closest variation with the ground soil moisture compared to other methods. The Dispatch and the *T** based methods also perform well with probability distribution but cannot reach the maximum limit and are also found to be

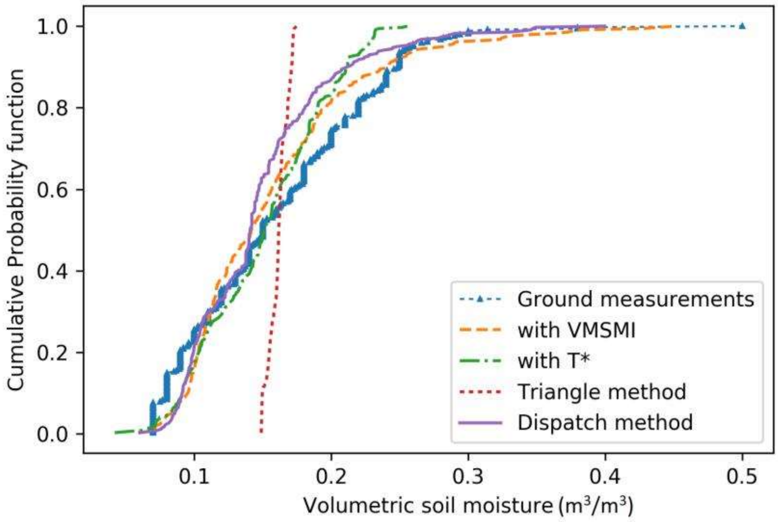

Figure 5.8 CPFs of downscaled SMAP soil moisture (with Triangle, Dispatch, *T** and VMSMI methods), and ground observations.

more deviated than VMSMI based approach. The triangle method is found unable to match the variation of the probability distribution and gives a straight line due to the small range of soil moisture values. On seeing the overall results, the downscaled soil moisture obtained by VMSMI based method shows good accuracy with the ground-measured soil moisture.

5.4.4 Spatial and Temporal distribution of downscaled soil moisture

Figure 5.9 depicts the temporal aspects of averaged downscaled soil moisture over the study region, as well as a comparison to average ground soil moisture over the study area. As observed previously, the Triangle method is not showing much variation from its mean value and gives almost a straight line from October 2017 to April 2018. Apart from it, the other three methods perform accurately in the month of October, November, March, and April according to the performance of SMAP soil moisture (Figure 5.5). From December to February (Winter), the Dispatch and the *T** based approach show identical behaviour to each other as well as to the SMAP soil moisture. They performed poorly in December while delivering better accuracy in January and February. Where the VMSMI-based

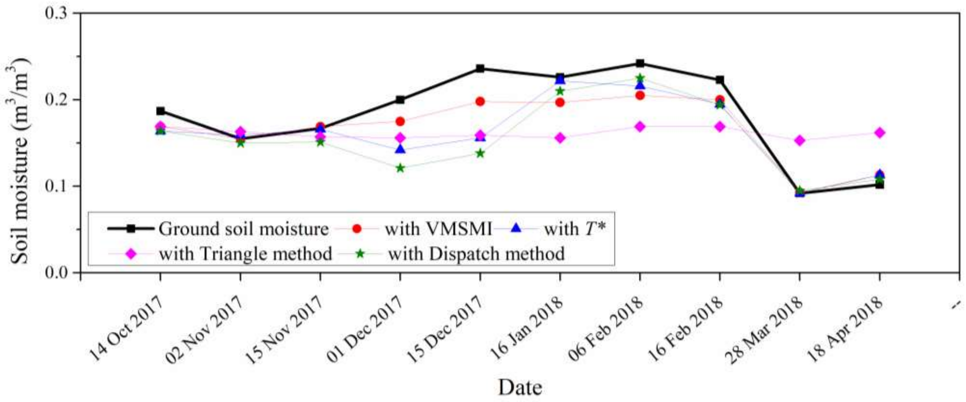

Figure 5.9 Temporal variation of ground soil moisture and the downscaled soil moisture using different algorithms.

method shows an average performance and lies in between them. On seeing the complete variation of soil moisture values, the VMSMI-based method shows a more accurate pattern with the ground soil moisture for all the dates. This proposed method also offers a closer variation to the ground soil moisture than the coarse resolution SMAP soil moisture (Figure 5.5). Thus, on seeing the overall results, the VMSMI-based method enhanced the accuracy of the downscaling of SMAP soil moisture data product.

The downscaling algorithms described in the methodology section were applied to the SMAP and MODIS data for 10 dates between 2017 and 2018 over the chosen study area. The spatial maps of the 9 km and 1 km SMAP soil moisture using existing downscaling algorithms and the proposed downscaling algorithm for 28 March 2018 are shown in Figure 5.10. The overall spatial pattern of 1 km soil moisture is approximately identical to the 9 km SMAP soil moisture variations with more fine distributions. Most of the area is covered by agricultural fields or croplands along with the Ganges River in the center. One can observe the different variations near the river in different images of 1 km soil moisture. The spatial comparison of 1 km soil moisture obtained from different downscaling algorithms reveals that the soil moisture variation increases from Figure 5.10(b - e). The

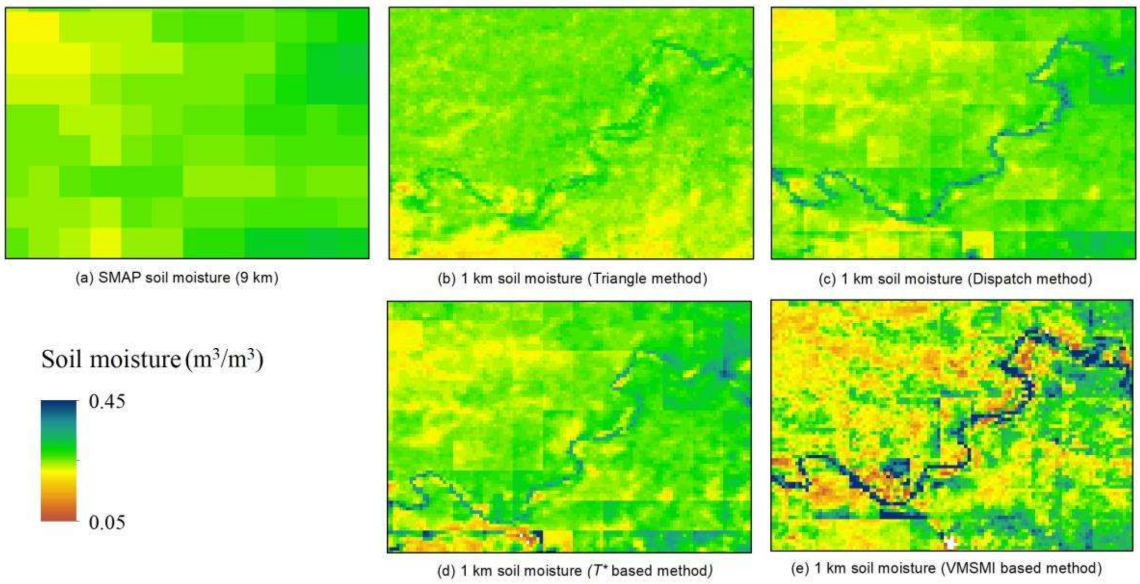

Figure 5.10 Spatial maps of 9 km SMAP soil moisture data (a), and 1 km downscaled soil moisture data (b) – (e).

triangle and the VMSMI-based methods show the lowest and highest variations in soil moisture values, respectively. The spatial patterns of the image of the Triangle, Dispatch, and VMSMI-based method rely on the distribution of MODIS LST and NDVI, whereas that of the T^* based approach depends only on the distribution of MODIS LST. The SMAP soil moisture doesn't show wet patterns in the center portion of the study area (Ganges River), while the downscaled images show high soil moisture values in this region. The spatial map of downscaled soil moisture using the Triangle method is also unable to deliver some wet patterns for the Ganges River due to the low variability of soil moisture values from its mean value. The soil moisture variation in VMSMI based method increased due to the inclusion of vegetation modulation factor. It reduces the impact of vegetation on soil radiance, which leads to a valid signature of the soil surface.

Many methods have already been developed for downscaling SMAP soil moisture, but accurate measurements of high-resolution SMAP soil moisture are still needed. The existing algorithms discussed in this paper either have some complexity to formulate or are not so appropriate. Such as, the Triangle method may induce some error in the measurements of soil moisture due to the dependency of its regression coefficients on the wet and dry edge; therefore, the measured values do not show much variation from the mean value. However, the Dispatch method is more physical and theoretical than the Triangle method, but it is more complicated to formulate due to the dependency on different zones. The new method introduced in this chapter is more straightforward and computationally simple. As the LST denotes the entire energy balance of the Earth's surface, it is used in a normalization form to derive a parameter for the representation of the land surface. Since LST denotes the entire energy balance of land surface, it includes both vegetation and soil reflectance. Hence, it requires minimizing vegetation contribution from the parameter T^*. Therefore, a new vegetation parameter derived by normalizing the NDVI is used to exclude the vegetation effects. It is used as a subtractor from T^*. After subtracting the vegetation parameter, the obtained vegetation modulated soil moisture index can provide the precise representation of the soil surface and be used to model soil moisture disaggregation.

5.5 CONCLUSIONS

This research proposes a relatively simple and computationally efficient approach based on vegetation modulation to improve the spatial resolution of SMAP soil moisture. Previously, a parameter formulated by using MODIS LST has been used with the coarse resolution SMAP soil moisture to obtain the high-resolution soil moisture content at 1 km, which is termed as T^*. But this parameter relates to the total reflectance of the land surface; therefore, to minimize the effect of vegetation, it is modified by using a vegetation modulation parameter, which is a function of NDVI. The resultant obtained index is then used for downscaling the SMAP soil moisture and named the Vegetation Modulated Soil Moisture Index (VMSMI).

The downscaled SMAP soil moisture using existing and proposed techniques and coarse resolution SMAP soil moisture are then compared to the ground measurements of soil moisture. The comparison is performed in the form of a scatter plot, Taylor plot, statistical analysis, and CPF plot. The obtained results indicate that the VMSMI-based method improves accuracy among all methods and provides a better correlation and closer variability with in-situ measurements of soil moisture. The accuracy of downscaled soil moisture is also enhanced than the coarse resolution soil moisture product. It shows a more comparative temporal variation to the ground soil moisture due to the modulation of the vegetation in soil reflectance.

Based on the aforementioned findings, it can be concluded that the suggested VMSMI-based downscaling approach may be used to improve the spatial representation of satellite soil moisture since it performs better than the traditional downscaling techniques. The proposed method not only improves the spatial scale of coarse resolution soil moisture but also provides an accurate measurement than the SMAP 9 km product. This vegetation modulated index performs well for the chosen study area, which is mainly covered by agricultural fields and low vegetation; therefore, it is more suitable for the low vegetation area and can be helpful in obtaining the fine resolution of soil moisture data products over agricultural or sparse vegetation regions

****