Chapter 6 : ASSESSMENT OF VEGETATION WATER CONTENT USING SENTINEL-1 DUAL-POLARIZED SAR DATA FOR THE RETRIEVAL OF SMAP SOIL MOISTURE

6.1 INTRODUCTION

Soil microwave remote sensing is highly controlled by the covering vegetation layer characterized by some biophysical factors such as VWC. It is a vital parameter for parameterizing attenuation caused throughout the vegetation layer in monitoring soil moisture. Numerous algorithms have been established for the calculation of soil moisture content in the microwave region, such as SCA, DCA, MT_DCA, and LPRM. All of these approaches have assimilated VWC to reduce the error caused by vegetation attenuation. Out of these algorithms, SCA is the simplest, most straightforward, and most reliable approach and shows the lowest possible error for the soil moisture estimation [141]. SCA is basically built on the zeroth-order RTM (Tau-Omega model), which utilizes the VWC to calculate the VOD for the canopy layer [54, 55]. Many researchers have used SCA to compute soil moisture content from L-band microwave radiometers (SMOS and SMAP) [10, 44, 142-144]. Also, SCA is a baseline algorithm in SMAP L2 soil moisture product (From 2015 to 2021) [59], and it uses a VWC-NDVI relation provided by Jackson et al., [89] for estimating the VWC. All the researchers mentioned above used either LAI or NDVI as a variable to compute VWC, but these products are constructed on the optical datasets. Since the SMAP radiometer operates in the microwave range, the SMAP baseline algorithm (SCA) requires a comprehensive evaluation of the VWC in the microwave region. Therefore, this study attempts to reduce this error induced caused by the use of optical VWC by evaluating VWC in microwave frequency region and testing it for the SCA to calculate the SMAP soil moisture.

Currently, the VWC can be estimated by mainly two approaches. The method being used is primarily cantered on the optical satellite measurements to generate a universal VWC map using vegetation indices such as NDVI, NDWI (Normalized Difference Water Index), and LAI, canopy EWT [75, 77, 78, 81]. However, the optical measurements are influenced by several atmospheric conditions and solar radiance at the time of data acquisition. Another different approach includes active radar observations in the microwave region. Microwave radar observations have advantages over optical measurements because they can penetrate clouds and vegetation and acquire data in any weather condition. These radar observations can also be taken in two ways; ground measurements (by an airborne platform, drone, automotive radar, etc.) [83-85, 118, 145] and space-borne radar (SAR, RADARSAT, POLSAR) [87, 88]. Estimating VWC relies on ground-based observations that are mainly limited to certain crop types (e.g., wheat, soybean). Validation of these observations over other crops is typically absent or accepted 'as it is.' It may induce a significant uncertainty in these estimates. Therefore, an accurate assessment of VWC on a global basis is required for retrieval algorithms of soil moisture content. In the case of space-borne radar, Szigarski et al., [87] and Mandal et al., [88] have utilized SMAP L-band radar and Sentinel-1 SAR, respectively, to compute vegetation indices for the mapping of VWC. However, SMAP L-band radar stopped operating in 2015 due to technical issues. Therefore, the current study utilizes dual Polarized Sentinel-1 SAR data and the DPRVI developed by Mandal et al., [88] in multiple regression approaches to monitor VWC. The dual-polarization (VV-VH) mentions the wave transmission in vertical polarization and the simultaneous receive in either vertical or horizontal polarization. Thus, the wave obtained after the reception is either co- or cross-polarized (VV-VH), which includes information about the target as a backscatter intensity.

Sentinel-1 data will be used for generating multiple vegetation indices (DPRVI, RVI, and CCR), which further leads to estimating VWC through some regression methods.

The formulation of DPRVI depends on the eigenvalues, calculated by the Eigen-decomposition of the second-order covariance matrix and degree of polarization. DPRVI includes the degree of polarization and the dominant normalized eigenvalue, which bounds it from 0 to 1. Further, another two vegetation indices (RVI and CCR) are also used along with the DPRVI to examine the best representative index for mapping VWC. The quad-pol RVI was firstly presented by Kim and van Zyl [146] and further modified by Trudel et al., [147] for dual-polarized SAR data. RVI is a simple formulation that includes backscattering coefficients with all polarizations of a specified data product. The present study estimates the RVI through the dual-polarized Sentinel-1 SAR data.

Along with RVI, one more vegetation index, the CCR, is also retrieved to monitor VWC. These three vegetation indices are good indicators of VWC. This study used these microwave vegetation indices as input one by one in three regression models, including Random Forest (RF), Support Vector Regression (SVR), and Adaptive Neuro-Fuzzy Inference System (ANFIS), for the approximation of VWC and to acquire the best representative index for VWC by comparing their results. Finally, the estimated values of VWC by different combinations of vegetation indices and the regression algorithms will be used in SCA to test their performance and find the best combination for evaluating soil moisture.

6.2 DATASETS

6.2.1 In-situ measurements

The field sampling was started on 2 January 2020 with the initial plan of ending in May 2020 according to the complete crop cycle of wheat. However, because of the

COVID pandemic, the field sampling remained incomplete and was terminated on 23 February 2020. Still, the data samples were collected for five dates in January and February 2020. The complete detail of the field sampling is given in section 2.3.2 of chapter 2.

6.2.2 Satellite data

This study includes two satellite data sets, Sentinel and SMAP (Table 6.1). Sentinel-1 SLC (dual-polarized) images were assimilated for the monitoring of VWC. The Sentinel-1 mission is a component of SAR and was launched by the ESA. SPL1CTBP_E and SPL2SMP_ E data products from the SMAP mission are also acquired for the monitoring of soil moisture. The data product of brightness temperature (SPL1C) and the data product of Surface Temperature (ST) and clay percentage (SPL2) are employed as the inputs in SCA for the assessment of soil moisture. In contrast, SMAP L2 soil moisture data values

Table 6.1 Description of datasets used in the study.

Data	Satellite product	Satellite	Spatial resolution	Temporal resolution
Dual Polarized backscatter intensity	S1A_IW_SLC_1SDV	Sentinel-1A	5 x 20 m	6 days
Brightness Temperature	SPL1CTB_E (Version - 003)	SMAP	9 km	Daily
Surface Temperature	SPL2SMP_E (Version - 004)	SMAP	9 km	Daily
Clay Percentage	SPL2SMP_E (Version - 004)	SMAP	9 km	Daily
Soil moisture	SPL2SMP_E (Version - 004)	SMAP	9 km	Daily

are included for the comparison with the assessed soil moisture. Details of these satellites are described in sections 2.1.1 and 2.1.3.

6.3 METHODOLOGY

The present study's focus is to monitor the VWC in two ways. First, by investigating the potential of three microwave indices (DPRVI, RVI, and CCR) from Sentinel-1 to calculate the VWC through three regression approaches (Random Forest Regression, Support Vector Regression [SVR], and Adaptive Neuro-Fuzzy Inference System [ANFIS]). Secondly, for validation, these results are used in SCA to assess the SMAP soil moisture. Figure 6.1 demonstrates the complete workflow of the paper.

6.3.1 Pre-processing of Sentinel-1 Polarimetric SAR data

The SLC Sentinel-1 level-1 data products are acquired to retrieve VWC. It is a dual-polarized phase-preserving SAR system and can collect different images through an identical sequence of pulses with its antenna in specific polarizations. The images are pre-processed in SNAP 8.0 (Sentinel Application Platform) for extracting the Polarimetric components through the following steps: (1) TOPSAR splits (2) Apply orbit file (3) Radiometric calibration (twice: (a) as a complex output (b) sigma naught as the output) (4) TOPSAR Deburst (5) Multilooking (5) Matrix generation (only for DPRVI) (6) speckle filtering (7) Terrain correction (8) polarimetric features extraction to generate Polarimetric indices.

The primary acquisition mode of Sentinel-1 SAR over land is IW swath mode with a swath of 250 km in length and spatial resolution of 5 x 20 m2. The sub-swaths (IW1, IW2, IW3) are captured with the Terrain Observation and Progressive Scan SAR (TOPSAR) in IW swath mode. Each individual sub-swath is available with the array of some bursts, and

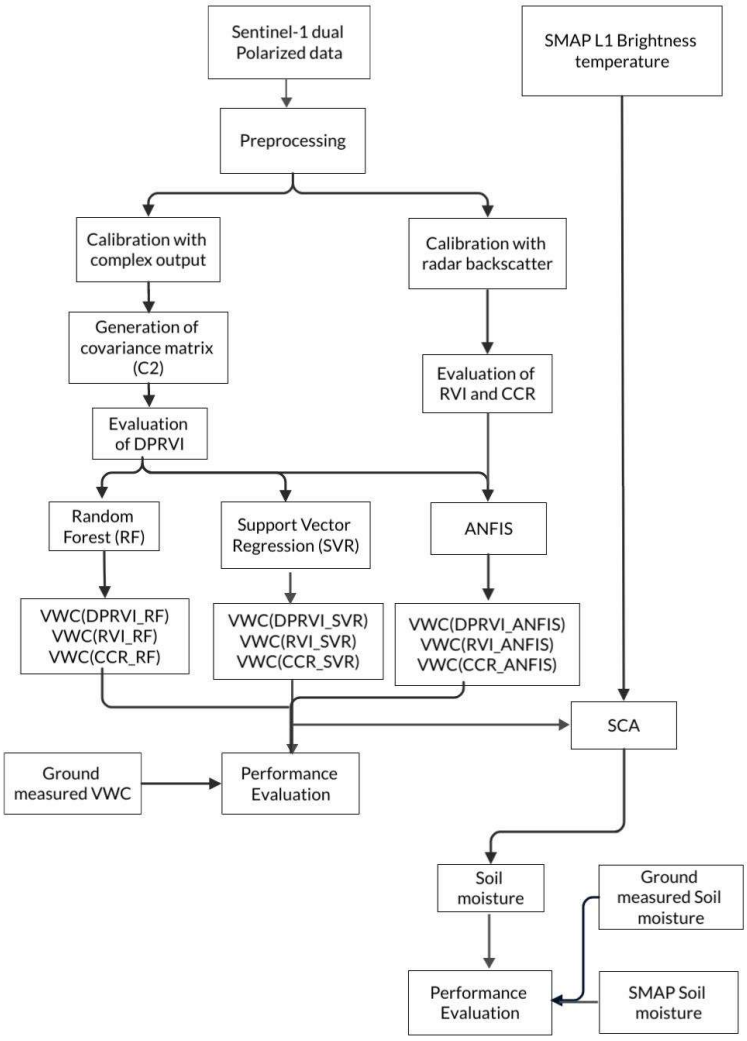

Figure 6.1 Flow chart for the estimation and validation procedure of VWC and soil moisture in this study.

these bursts are considered as a single SLC image. The first pre-processing step is the TOPSAR split, used for extracting a separate image from the Sentinel-1 product by splitting the sub-swaths and selecting the desired bursts and polarizations. Then, the next module applies an orbit file which downloads the precise orbit file to update the state vectors. Later, radiometric calibration is employed on the images to calculate the radar backscatter values from the digital number of each pixel. This step is performed two times; Firstly, the saved output is in a complex format needed to find the correlation and the relative phase between VH and VV polarizations for covariance matrix generation. Secondly, the saved output contains radar cross-section power (σ^0) to calculate the RVI

and CCR ($\sigma^0_{VH}/\sigma^0_{VV}$). The calibrated images are processed in TOPSAR Deburst to merge all bursts and sub-swaths into a single SLC image. According to the test site, the Deburst image is then subset to clip a small image, and then multilooking was performed by 4 x 1 to convert the pixels into ground range square pixels. After multilooking, the complex output product is used to produce a 2 x 2 covariance matrix (C_2) for approximating the DPRVI. The Polarimetric speckle filtering with the 'refined lee' filter is applied on both the images (images calibrated with σ^0 as an output and the matrix elements calibrated with complex output). Then, the Range Doppler Terrain Correction was performed on these images to moderate the distortions introduced by the topographical variations and to geocode the images into a WGS84/UTM coordinate system. The Shuttle Radar Topography Mission (SRTM) was automatically downloaded by the SNAP toolbox and utilized as a Digital Elevation Model (DEM) for height information.

After the pre-processing, the data products are ready to calculate the DPRVI, RVI, and CCR

- DPRVI is evaluated through the following expression [88]

$$DPRVI = 1 - m\beta \tag{6.1}$$

$$\text{Where } m = \sqrt{1 - \frac{4|C_2|}{(Tr(C_2))^2}} \tag{6.2}$$

$$\beta = \frac{\lambda_1}{\lambda_1 + \lambda_2} \; or \; \frac{\lambda_2}{\lambda_1 + \lambda_2} \tag{6.3}$$

λ_1 and λ_2 are the eigenvalues of the covariance matrix (C_2), derived by vectorizing the dual-polarized scattering matrix [148].

$$[S_{dual}] = \begin{bmatrix} 0 & S_{VH} \\ S_{VH} & S_{VV} \end{bmatrix}$$

$$[C_2] = \begin{bmatrix} < |S_{VV}|^2 > & < S_{VV}S_{VH}^* > \\ < S_{VH}S_{VV}^* > & < |S_{VH}|^2 > \end{bmatrix}$$

- RVI and CCR are estimated through [147],

$$RVI = \frac{4\sigma_{VH}^0}{\sigma_{VH}^0 + \sigma_{VV}^0} \tag{6.4}$$

$$\text{Cross} - \text{ and co} - \text{polarized ratio} = \frac{\sigma_{VH}^0}{\sigma_{VV}^0} \tag{6.5}$$

6.3.2 Assessment of VWC using DPRVI, RVI, and Cross- and co-polarized ratio through multiple regression approaches

The pre-processing of Sentinel-1 SLC images and the evaluation of DPRVI, RVI, and the CCR are described in detail in the preceding section. This section will describe three machine learning methods, RF, SVR, and ANFIS, utilized for the estimation of VWC.

- ***Random Forest***

Random Forest (RF) is a Machine learning algorithm that is related to the family of ensemble learning methods. It utilizes the weak learners (poor accuracy models) for training and then aggregates their prediction to build a model with better accuracy. In RF, these vulnerable learners are considered the decision trees [149, 150]. The RF regression method uses various samples from the training datasets to obtain different nodes of the tree, and these nodes are some conditions given to the decision trees from root to leaves. The RF approach combines each tree's results to predict a better result. The RF model has an advantage because it doesn't depend upon the assumptions for the correlation between

the predictors and dependent variables and can identify the non-linear relations. It doesn't give overfit results and always converges and gives bounded generalized error on increasing the number of trees. It uses multiple decision trees, which are trained for response variables. A maximum of 200 decision trees was chosen for training using 55 % of sample datasets. The remaining datasets are considered the Out of Bag (OOB) samples [151]. The OOB samples are then utilised to validate the RF model. RF model can be described as [152],

$$RF\ predictions = \frac{1}{k}\sum_{k=1}^{k} k^{th}\ tree\ response \tag{6.6}$$

Where k index runs over each tree.

- ***Support Vector Regression***

Support Vector Regression (SVR) uses the concepts of Support Vector Machine (SVM) to perform regression. It is basically based on the boundaries between the classes to map the input space of independent variables with the help of a non-linear transformation built on a kernel function. There are various types of kernel functions, such as Polynomial, linear, Sigmoid, Radial basis functions, and many more. Out of these kernel functions, the radial basis function was the most powerful [153] because it is straightforward and reliable and can deal with the complex dimensional space and the separation factors. The SVR uses the set of some data points, which are dependent on the predefined error, to perform a regression between dependent and predictor variables. These sets of data points are identified as support vectors.

Let's consider for the given independent variable X = {x_1, x_2, x_3….., x_i}; the target variable is Y = {y_1, y_2, y_3, ……. y_i}, X, Y∈R. The SVR maps the original space of input points to a higher dimensional space through a non-linear radial basis function,

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{6.7}$$

Generally, the fitting function in SVR is approximated in the linear form,

$$f(x) = a + w.K(x) \tag{6.8}$$

Where *w* is the vector of weights, and a is the bias.

To reduce the model complexity, the fitting function is sought by minimizing $\|w^2\|$

$$Min\ \frac{1}{2}\ \|w^2\| + C\ \sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{6.9}$$

$$\text{With constraints, } \begin{cases} y_i - f(x_i) \le \epsilon + \xi_i, \\ f(x_i) - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

Where ξ_i and ξ_i^* are the positive and negative errors for the i^{th} data point, respectively. Thus, the optimization problem is now equivalent to the minimization of the Lagrange function, and the final predicted variable in terms of the input domain is,

$$f(x) = a + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\ K(x, x_i) \tag{6.10}$$

Where α_i and α_i^* are Lagrange multipliers.

- ***Adaptive Neuro-Fuzzy Inference System (ANFIS)***

ANFIS is a combined implementation of Artificial Neuro Network (ANN) and the fuzzy logic approaches. It utilized the benefits of both the techniques, such as qualitative aspects of fuzzy logic and quantitative aspects of ANN. Therefore, ANFIS is a more effective modelling tool than these two independent models (ANN and fuzzy logic). It is multi-layered and feeds forward ANN with the adoptive nodes, where the output is obtained on the parameters of the adoptive nodes. These parameters are adjusted by the error specified by the learning instructions. Generally, hybrid learning is chosen as the form of learning method in ANFIS [154].

ANFIS performs more accurately when no suitable method is available to select the membership functions and parameters [155]. ANFIS applies a combined form of error backpropagation with the least square method to adjust the fuzzy logic membership function [156]. In the end, the derivative of the squared error of every node's output is taken as the error signal and fed back with the input of the model to adjust the membership function parameters [157, 158]. The detail of ANFIS modelling is described in Dehnavi et al., [159] and Mohanty et al., [160].

The ANFIS structure with single input and the single output is presented in Figure 6.2. The ANFIS structure combines two sections, which are antecedent and conclusion. These two sections are related through some fuzzy rules. Let’s consider a set of m rules in the arrangement of [161],

Where Ai are the fuzzy membership functions and Pi and Qi are the acceptable real parameters required to determine output function. The layers of the ANFIS model are elucidated below [161]

$$\textit{if } u \textit{ is } A_i \qquad \textit{then} \qquad f_i = P_i u + Q_i \;\; \textit{for } i = 1, 2, 3, \ldots\ldots\ldots m$$

Layer 1 This layer includes membership functions corresponding to each input. This paper used a triangular membership function, described as,

$$\mu_{A_i}(u) = \text{trimf}\,(u: a_i, b_i, c_i)$$

$$= \max(\min\left(\frac{u - a_i}{b_i - a_i}, \frac{c_i - u}{c_i - a_i}\right), 0) \tag{6.11}$$

Where $i = 1, 2, \ldots\ldots\ldots m$

$$O_i^1 = \mu_{A_i}(u) \tag{6.12}$$

The a_i, b_i, c_i are the antecedent parameters and help to decide the shape of membership functions of each input.

Layer 2 This layer characterizes the firing strength of rules developed in the previous layer. For a multi-input system, the output of layer 2 is the product of all sets of fuzzy inputs coming from layer 1. Whereas, for a single input system, there is only one input, so the output of layer 2 is the same as layer 1

$$O_i^2 = w_i = \mu_{A_i}(u) \tag{6.13}$$

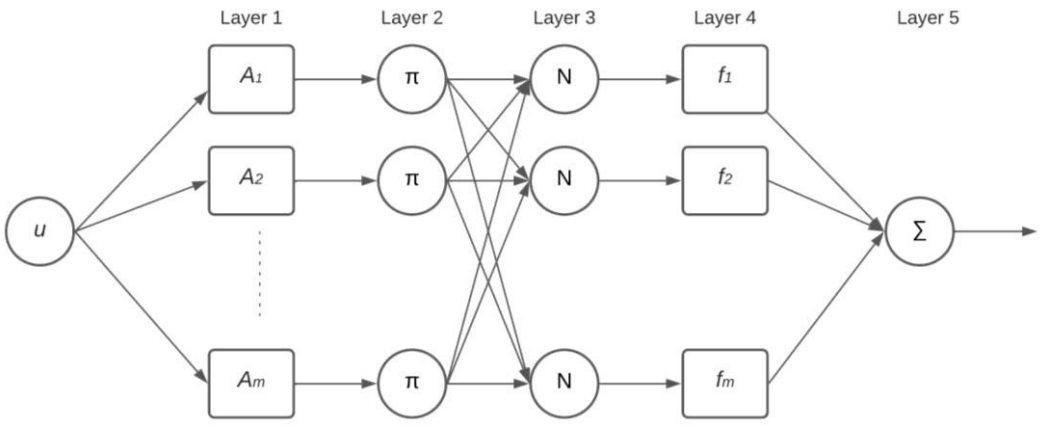

Figure 6.2 Structure of ANFIS model with single input and single output.

$i = 1, 2, \ldots\ldots m$

Layer 3 This layer normalized the firing strength of all inputs and provide a ratio of the firing strength of ith rule and the total firing strength of all rules

$$O_i^3 = \overline{w}_i = \frac{w_i}{\sum_{i=1}^{m} w_i} \tag{6.14}$$

Layer 4 This layer provides the product of the preceding layer's output and the parameter set

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (P_i\, u + Q_i) \tag{6.15}$$

Where the parameter Q_i is known as the conclusion parameter.

Layer 5 This layer is used to compute the complete output through the sum of the entire incoming signal

$$O_i^5 = overall\ output = \sum_{i=1}^{m} \overline{w}_i f_i \tag{6.16}$$

The vegetation indices (DPRVI, RVI, CCR) explained in the earlier section are incorporated in the three machine learning techniques described above to retrieve VWC. These indices are used one by one as a single input in these approaches, and VWC is obtained as an output.

6.3.3 SCA for the retrieval of soil moisture content

The SCA basically originated from the zeroth-order RTM (Tau-Omega model) and is described in section 2.5.1 of chapter 2.

6.4 RESULTS AND DISCUSSION

6.4.1 Temporal variation of Sentinel-1 and SMAP-derived data products along with ground measurements

The ground sampling was executed on the five dates of 2020 according to the revisit cycle of Sentinel-1 SAR data, the Sentinel-1 SLC dual-polarized, and the SPL1CTB_E and SPL2SM_P_E data products of SMAP were also downloaded for the exact selected dates. The Sentinel-1 data was utilized to calculate three microwave vegetation indices for VWC estimation. Therefore, Figure 6.3(a) shows the temporal behaviour of these indices for the chosen time series along with the ground-measured VWC. All the variables show rising behaviour after the first day of observation, indicating crop growth. Figure 6.3(b)

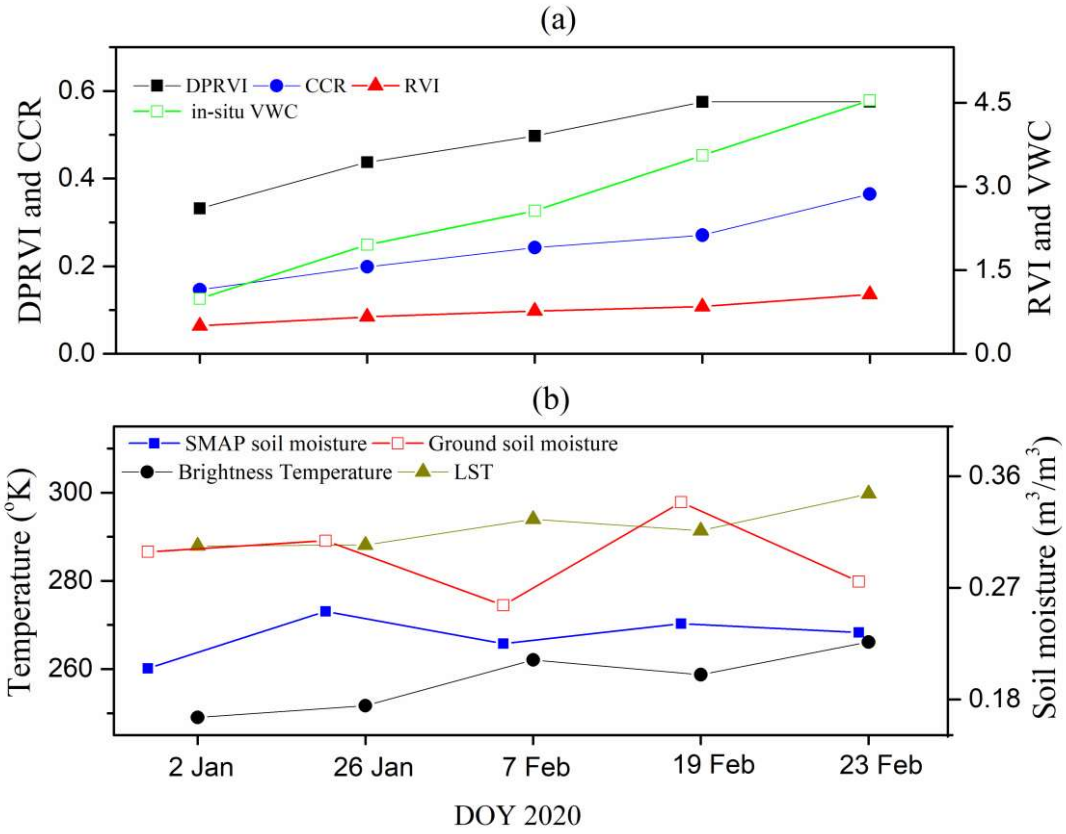

Figure 6.3 Temporal behaviours of (a) different vegetation indices (DPRVI, RVI, CCR) along with in-situ VWC and (b) SMAP derived data products along with the in-situ soil moisture.

shows the temporal behaviour of SMAP-derived soil moisture, ST, and brightness temperature along with the ground-measured soil moisture. The ST and brightness temperature show matching behaviour for the chosen time series, whereas the SMAP soil moisture is always lower than the in-situ soil moisture. The difference between the minimum and the maximum values of in-situ soil moisture values is also greater than that of SMAP soil moisture

6.4.2 VWC estimation using DPRVI, RVI, and CCR by machine learning techniques

As the DPRVI, RVI, and CCR are the vital indicators of VWC, three machine learning approaches are used with these vegetation indices to retrieve VWC. For validation, retrieval accuracy is examined by calculating the correlation coefficient (R^2), Root Mean Square Error (RMSE), slope, and bias error (Table 6.2). The assessed VWC

Table 6.2 Statistical analysis of VWC retrieval using different vegetation indices and machine learning algorithm.

Regression approach	Vegetation index	R^2	RMSE	Bias	Slope
RF	DPRVI	0.59	0.42	0.06	0.47
	CCR	0.33	0.49	0.19	0.29
	RVI	0.36	0.49	-0.06	0.28
SVR	DPRVI	0.53	0.52	-0.10	0.54
	CCR	0.51	0.46	-0.20	0.41
	RVI	0.53	0.45	-0.16	0.41
ANFIS	DPRVI	0.56	0.32	0.48	0.26
	CCR	0.51	0.27	0.35	0.19
	RVI	0.46	0.30	0.37	0.19

shows the highest correlation (R^2) for the RF regression method using DPRVI, whereas the RF regression method also provides the lowest correlation with CCR. In the case of RMSE, the ANFIS regression approach performs better with all vegetation indices with the values of 0.32, 0.27, and 0.30 using DPRVI, CCR, and RVI, respectively. In case of bias, DPRVI and RVI both show the lowest numerical value (0.06) with the RF regression approach with the opposite sign. The former is overestimating, whereas; the latter is underestimating VWC. In terms of slope, the highest slop value is obtained for the SVR approach using DPRVI, while; the second-highest slope value is obtained with the DPRVI but with the RF approach.

The validation in the form of scatter and Taylor plots are presented in Figures 6.4 and 6.5, respectively. Figure 6.4 is in the form of a 1 x 3 matrix. The first column represents the RF model, the second corresponds to SVR, and the third corresponds to the ANFIS regression approach. The underscore between vegetation indices and the regression approaches defines the use of vegetation index as the input in that regression method. Comparing the scatter plot results, the DPRVI performs well compared to others and

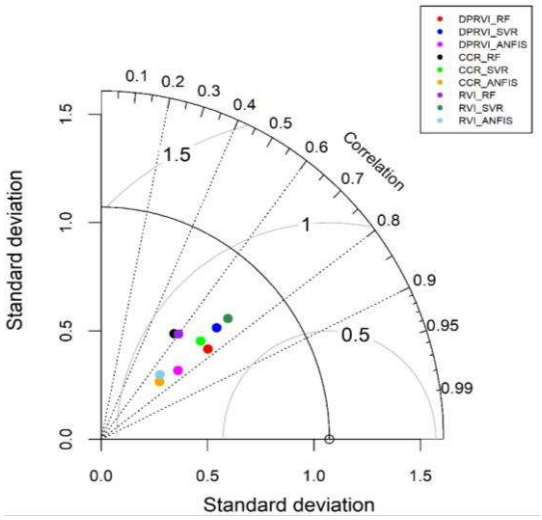

Figure 6.4 Comparison of different combinations for VWC estimation by the Taylor diagram.

shows a better distribution of data points with both RF and SVR models with correlation coefficients of 0.59 and 0.53, respectively.

The Taylor diagram compares various approaches, methodologies, or models by three statistical variables; centred RMSD (Root Mean Square Difference), standard deviation, and correlation coefficient. The white circle on the x-axis denotes the point for the observed data (Ground measurements), and other circles with several colours represent the results of different models. The concentric radial lines with a white circle as a common center denote the centred RMSD. The solid radial line through the reference data point indicates the standard deviation from the observed dataset. The arc of the diagram is considered for the correlation coefficient. The data point nearest to the observed data point (white circle) represents the best fit among all the data points. This study utilizes the Taylor diagram to identify the best vegetation index and regression approach for evaluating VWC (Figure 6.5). In the Taylor diagram, most data points lie between the correlation lines of 0.6 and 0.8. Only the black dot, corresponding to the CCR_RF model, lies slightly above the correlation line of 0.6. The red and magenta dot corresponds to DPRVI_RF and DPRVI_ANFIS models, respectively, showing the RVI_SVR model (sea green highest correlation compared to all data points. Still, in terms of centred RMSD, the

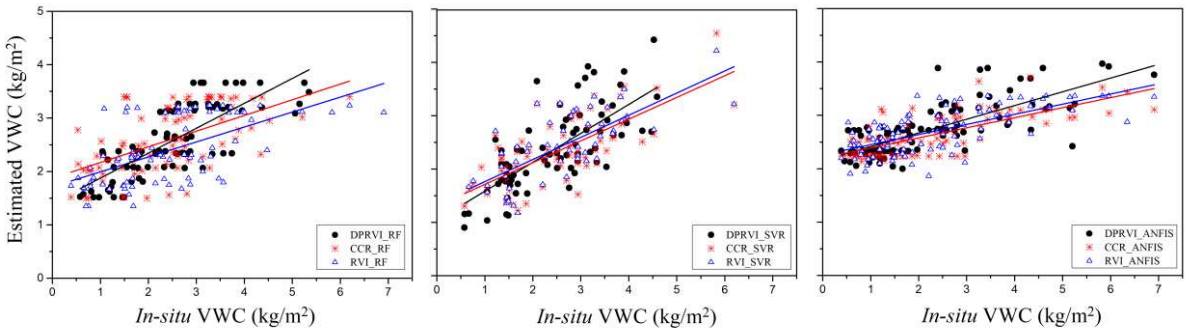

Figure 6.5 Comparison between estimated and Ground truth VWC for different combinations of vegetation indices and machine learning algorithm.

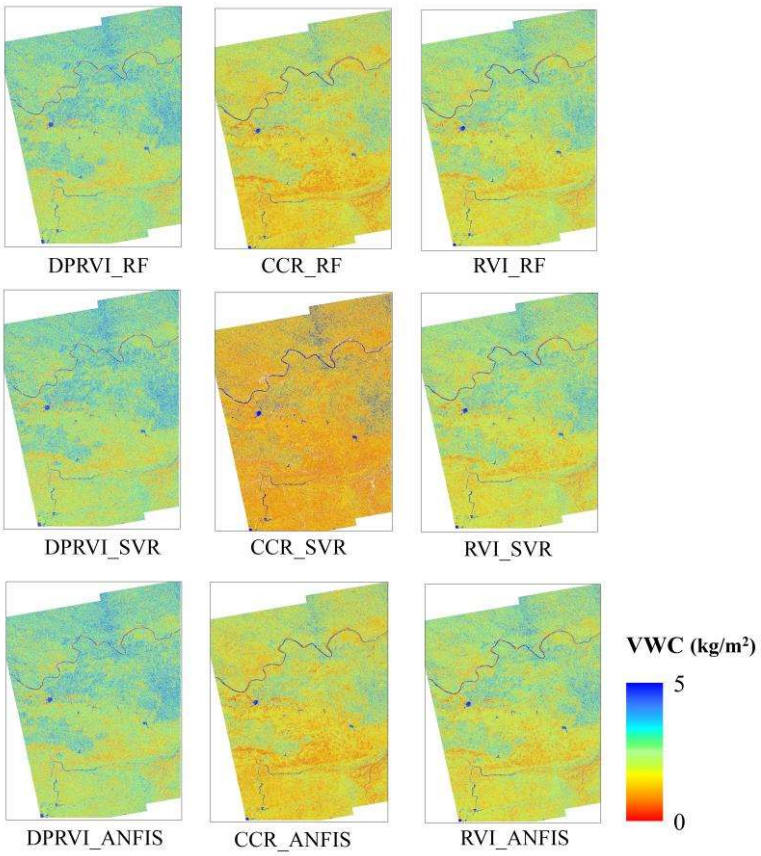

Figure 6.6 Spatial distribution of estimated VWC using different combinations of vegetation indices and machine learning algorithms.

red dot performs better. Regarding the standard deviation, the dot) is found closest to the line of standard deviation. According to the conditions of the Taylor diagram, the red dot is seen most relative to the reference data point, which illustrates that the DPRVI_RF regression model gives the best fit.

The spatial distribution maps of estimated VWC using different vegetation indices and different machine learning algorithms are presented in Figure 6.6. The maps are generated for the date of 23 February 2020. The image corresponding to RF with DPRVI provides the highest estimated VWC values, while the image corresponding to SVR with CCR provides the lowest estimated VWC values. The study area is mainly covered by cultivated land and so can be seen in the maps of VWC. According to field observations

of the study region and the croplands, the DPRVI produces more accurate results than the RVI and CCR using all machine learning techniques, although the highest accuracy is attained by the RF-based approach.

6.4.3 Assessment of Soil moisture by using SCA incorporated with estimated VWC and its comparison with SMAP L2 soil moisture

The VWC estimation using three approaches and vegetation indices, shown in the preceding section, the computed VWC are further utilized in SCA to estimate the soil moisture content. A comparison has been performed between estimated and in-situ soil moisture, presented in the scatter plots (Figure 6.7). The appraised soil moisture via enhanced SCA shows a worthy correlation with the in-situ soil moisture compared to the SMAP L2 soil moisture data product. The data-points corresponding to SMAP soil moisture doesn't significantly vary with the mean value and are scattered between 0.2 and 0.25 m^3/m^3. Consequently, it doesn't show a good correlation with the ground measurements of soil moisture. On comparing the results of SCA, the inclusion of estimated VWC in SCA significantly increases the accuracy of soil moisture estimation, and the estimated VWC by SVR with DPRVI gives the highest correlation among other combinations of vegetation indices and regression approaches. The statistical analysis in the form of correlation coefficient (R^2), RMSE, bias, and the slope is given in Table 6.3.

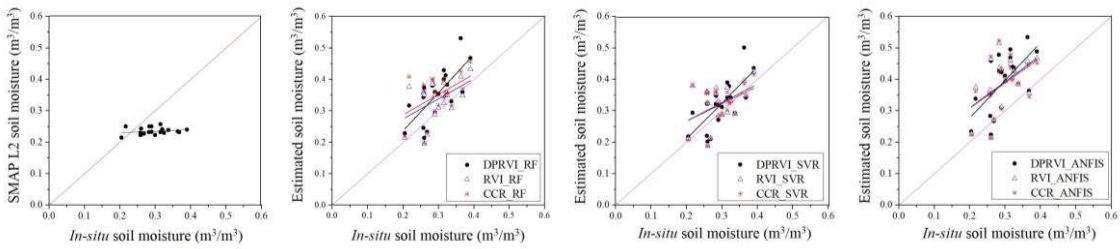

Figure 6.7 Comparison of estimated soil moisture and SMAP L2 soil moisture with in-situ soil moisture.

The correlation coefficient for the SMAP soil moisture is 0.05, whereas the highest and lowest correlation values for the predicted soil moisture are 0.53 with DPRVI_SVR and 0.19 with CCR_SVR, respectively. Which specifies that the assessed soil moisture is performing well. Among three vegetation indices, the DPRVI performs better, and the DPRVI_SVR shows better results with all statistical variables. The predicted soil moisture shows a positive bias error of 0.02 to 0.10, and the SMAP soil moisture shows a negative bias error of 0.06. which indicates that the appraised soil moisture is overestimated, and SMAP L2 soil moisture underestimates its values compared to the in-situ soil moisture. In the case of RMSE, the assessed soil moisture ranges from 0.06 to 0.12, while for SMAP soil moisture, it is 0.08. In terms of slope, the predicted soil moisture always shows a greater slope than SMAP soil moisture, and the best-obtained slope value is 1.14 for

Table 6.3 Statistical analysis of retrieved soil moisture and SMAP L2 soil moisture with in-situ soil moisture.

		R^2	RMSE	Bias	Slope
Estimated soil moisture through SCA incorporated with VWC computed by	DPRVI_RF	0.51	0.08	0.06	1.22
	CCR_RF	0.21	0.08	0.05	0.65
	RVI_RF	0.23	0.07	0.04	0.64
	DPRVI_SVR	0.53	0.06	0.02	1.14
	CCR_SVR	0.19	0.06	0.02	0.55
	RVI_SVR	0.20	0.06	0.03	0.59
	DPRVI_ANFIS	0.40	0.12	0.10	1.20
	CCR_ANFIS	0.21	0.12	0.09	0.80
	RVI_ANFIS	0.24	0.06	0.09	0.83
SMAP L2 soil moisture		0.05	0.08	-0.06	0.05

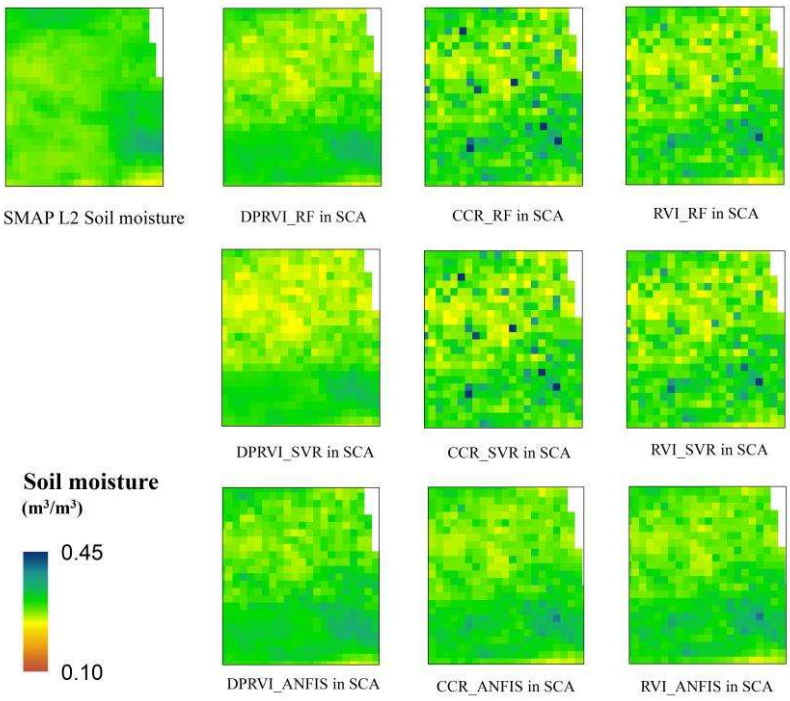

Figure 6.8 Spatial distribution pattern of SMAP L2 soil moisture and estimated soil moisture using different combinations of vegetation indices and machine learning approaches.

DPRVI_SVR. It indicates that the SCA algorithm with microwave VWC improves soil moisture assessment.

Figure 6.8 shows the spatial map of SMAP L2 soil moisture and the calculated soil moisture for the date of 23 February 2020. A broad area has been picked for the generation of a soil moisture spatial map due to the coarse pixel resolution of SMAP (9 km). The spatial variability of all maps is almost identical, and DPRVI performs better than other indices. The spatial map for CCR_RF, CCR_SVR, RVI_RF, and RVI_SVR doesn't provide a symmetrical pattern. The comparison of vegetation indices illustrates that DPRVI provides better soil moisture maps with all regression approaches.

6.4.4 Variation of estimated VWC and assessed soil moisture for the whole time series of the study

The previous section illustrates that DPRVI_SVR in SCA enhances soil moisture appraisal from SMAP. Therefore, the spatial maps of VWC estimated by DPRVI_SVR and soil moisture estimated by DPRVI_SVR in SCA for the complete time sequence of this study have been presented in Figure 6.9. Figure 6.9(a) shows the variation of VWC for the five days of 2020. One satellite of Sentinel-1 offers a revisit frequency of 12 days for the same region, and the two satellite constellations repeat their cycle after every 6 days. The first four maps are generated by using one satellite swath, and the fifth image is generated by utilizing another satellite's swath. Therefore, its alignment differs from others but has the same center latitude and longitude. The maps specify that the values of VWC are increasing with the dates and become high in the last image because of the

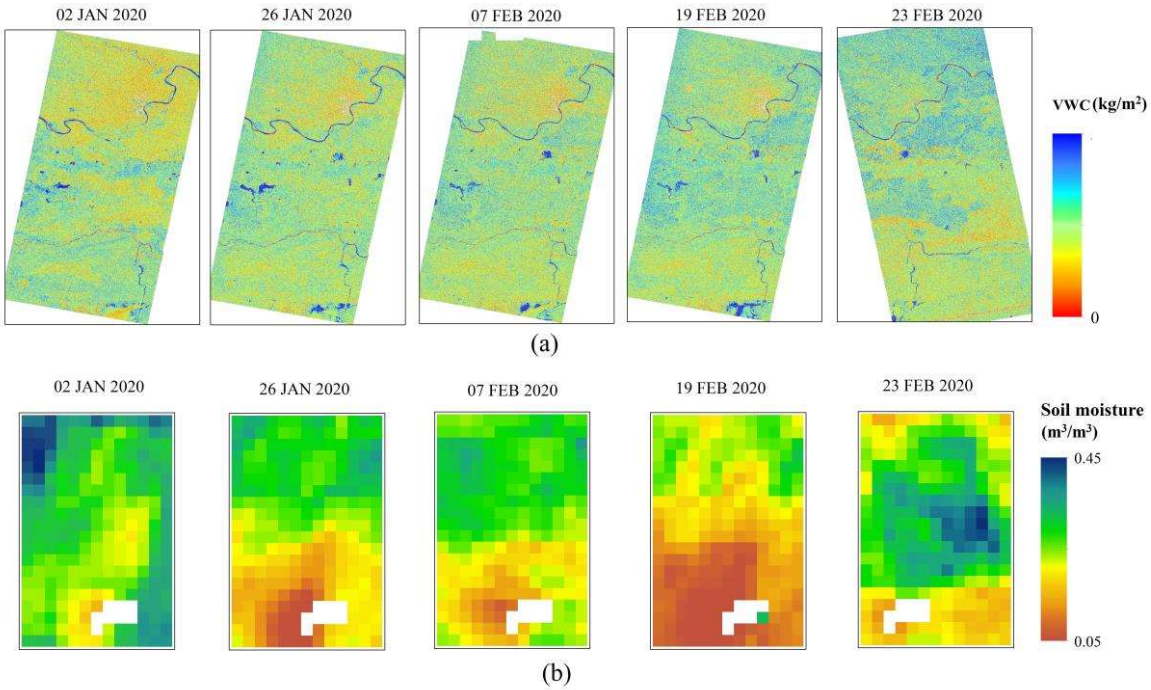

Figure 6.9 Spatial maps of (a) VWC estimated by DPRVI_SVR and (b) Soil moisture estimated by DPRVI_SVR in SCA for the five dates of 2020.

growing period of the wheat crop. Generally, the wheat is sowed in the months of December and January, due to which the image for 02 January displays the lowest value of VWC, and the image for 23 February shows the highest value of VWC.

The images of VWC have been then utilized in SCA for the evaluation of soil moisture. The spatial maps of the estimated soil moisture for the complete time series are displayed in Figure 6.9(b). Since the image of VWC for 23 February is generated by a different Sentinel satellite, therefore, the soil moisture map generated for the exact date is also slightly different. The amount of moisture content in soil shows a decreasing pattern from 02 January to 19 February because of the seasonal change from winter to spring. At the same time, the spatial map for 23 February shows a higher moisture content than on previous days due to the rainfall on 22 February 2020.

6.5 CONCLUSION

This study intended to improve the accuracy of soil moisture estimation by utilizing the microwave VWC in SCA. Firstly, dual-polarized data has been used to compute three vegetation indices; DPRVI, RVI, and CCR. These indices are then used one by one in three regression approaches for estimating VWC. The comparison results confirm that DPRVI is effectively a good indicator of VWC and performs well statistically among the three vegetation indices. And among the three regression approaches, SVR shows the highest average correlation coefficient, and ANFIS shows the lowest average RMSE for the three indices. However, on seeing individually, the RF regression method with DPRVI shows the highest correlation, lowest RMSE after SVR, lowest bias error, and the second-highest slope value with the in-situ measurements of VWC. Therefore, it concludes that DPRVI in the RF regression techniques is the best combination among the nine for VWC retrieval.

The values of estimated VWC are then employed in the approach of soil moisture estimation. The comparison of calculated and SMAP L2 soil moisture reveals that the former soil moisture values are more highly correlated with in-situ measurements of soil moisture than the latter one and perform well in each statistical variable. But the comparison of these nine combinations shows that the DPRVI_SVR can enhance SCA's accuracy.

As in soil remote sensing, it is required to exclude the influence of the vegetation layer from the ground surface emissions, and VWC is a crucial variable for accounting for the attenuation through vegetation. The SMAP radiometer operates in the microwave frequency region. Thus, the estimated microwave VWC can be combined with the SMAP baseline approach instead of optical VWC to estimate soil moisture more accurately.

****