

# COMPARISON OF SEMI-EMPIRICAL AND XGBOOST MACHINE LEARNING AIDED APPROACH FOR CYGNSS SOIL MOISTURE ESTIMATION

Ravi Prakash Kumar and Dr. Balaji Devaraju

Department of Civil Engineering, Indian Institute of Technology, Kanpur, India

## ABSTRACT

SMAP and SMOS space missions are reliable sources of soil moisture, providing coarse resolution data (a temporal resolution of more than three days and spatial resolutions of 36 km and 25 km, respectively) [9]. The resolution of soil moisture can be improved by using one of the successful GNSS-R missions, CYGNSS (Cyclone GNSS). CYGNSS, with eight micro-satellites in its constellation, captures a large volume of observations in the equatorial regions. Each micro-satellite can receive four reflections per second with varying spatial resolutions from  $0.5 \text{ km} \times 3.5 \text{ km}$  to  $25 \text{ km} \times 25 \text{ km}$  and a temporal resolution of 1 day [9-12]. In this paper, a soil moisture estimation model was made for CYGNSS satellite data so that the CYGNSS observations could be used differently to improve the soil moisture resolution. Since vegetation has always remained a problem for estimating soil moisture, this paper aims to account for vegetation attenuation to improve the Model's performance. The CYGNSS surface reflectivity was highly correlated with the SMAP soil moisture (correlation of 81.6% on barren land and correlation of 81.2% on vegetated land). Ulaby's Water Cloud Model worked well with the CYGNSS data (1.5 GHz) by ignoring the vegetation attenuation because it depends on frequency. The Model was more biased towards the higher frequency signals (more than 8 GHz). By adding the leaf area index as the fractional contribution of backscatter from vegetation to Ulaby's Water Cloud Model, a lot of changes were made that helped solve this problem. The modified water cloud model has performed well compared to Ulaby's Model in the vegetated and built-in regions with more than 75% correlations with RMSE less than  $0.01 \text{ m}^3/\text{m}^3$  on average. Further XGBoost Regression algorithm has been applied and achieved correlations between the predicted and SMAP soil moisture up to 87% with RMSE of  $0.001 \text{ m}^3/\text{m}^3$  and relative error of less than 0.2%.

**Index Terms**— SMAP, SMOS, GNSS-R, CYGNSS, WCM

## 1. INTRODUCTION

A fundamental requirement in the agricultural field is soil moisture. The satellites can be used for estimating soil moisture at a global scale [1-30]. The current successful missions of SMAP and SMOS give soil moisture data in low spatial and temporal resolutions (36 km and 25 km), which may need to be revised for agricultural applications [12]. The GNSS-R technique, CYGNSS, can be used to improve

the resolution of soil moisture. CYGNSS (Cyclone GNSS) captures large observations in equatorial regions. It is a group of eight micro-satellites in low Earth orbit at an angle of 35° with the equator. It is a type of bi-static radar in which GNSS constellations act as the transmitter and CYGNSS micro-satellites act as the receivers. Each micro-satellite is fitted with the 4 DDMI(Delay Doppler Mapping Instruments) and GPS receivers [10-12]. The GPS receivers use the direct signal to find the observatory location, and DDMI receives the surface properties in the form of DDM from the reflected signal. Each micro-satellite can collect four reflections per second, with a varying spatial resolution of  $0.5 \times 3.5 \text{ km}$  to  $25 \times 25 \text{ km}$ . With a 6-hour revisit time, CYGNSS covers the equatorial region well within 38°N to 38°S latitudes. The large volume of CYGNSS observations in the equatorial region can be re-purposed to retrieve soil moisture using the Water Cloud Model [10,12].

In previous research, according to Dong and Zin, more than 80% of the reflections of the ground surface are thought to be from the specular points [1]. Clara Chew [10] and Venkat Lakshmi [12] have achieved an 80% correlation between CYGNSS-derived surface reflectivity and SMAP soil moisture by neglecting the non-specular contribution in the effective surface reflectivity. Assuming the changes in soil moisture are due to the changes in surface reflectivity, they have developed a change detection model to estimate the soil moisture for the CYGNSS. The main demerit of this approach was that they needed to account for the vegetation attenuation contribution [10,12]. F.T.Ulaby created the Water Cloud Model in 1992. The Model was tested for frequencies above 9 GHz [28]. He concluded that Multi-polarized L-band observations could be utilized to reliably estimate soil moisture in the top 5 cm of soil and surface roughness for bare-soil surfaces [18]. In 2012 Kweon and Soon-Koo Hwang estimated soil moisture using Water Cloud Model for COSMO-SkyMed SAR data [21]. Marco Lavalle and Clara Chew used the Water Cloud Model for bi-static backscatter modeling with CYGNSS satellite measurements to create a dynamic map of the world's tropical wetlands in 2018 [25]. Developing the soil moisture estimation model for the CYGNSS satellite is still challenging.

This paper aims to develop a soil moisture estimation model for CYGNSS that considers the effects of vegetation attenuation, hence facilitating the mapping of the ever-changing state of tropical wetlands and other areas prone to floods. Here, a novel soil moisture estimation algorithm is proposed

for CYGNSS. The proposed Model showed an outstanding performance over all types of land surfaces, exhibiting its effectiveness. This paper is organized as follows. Section 2 describes the GLDAS, LULC, LAI, SMAP data, and CYGNSS observables. Section 3 presents the design of the algorithm. Section 4 shows the results and analysis. Section 5 summarizes the conclusions.

## 2. DATA SETS AND OBSERVABLE

### 2.1. CYGNSS DATA

According to various calibration levels, CYGNSS includes four distinct levels of data sets. A fundamental observable of CYGNSS, DDMs of received Power, and scattering cross-section are calibrated at Level 1. The DDMI produced the dataset that includes DDMs of specular points on the Earth's surface calibrated into power Received (Watts) and Bi-static Radar Cross Section (BRCS) of that specular point expressed in units of  $m^2$ . When downloading the data, to overcome the problem of sub-setting in OPeNDAP, a python script was written, where the selected variables could be given as a list for the given period and the extent of the region as inputs. The script can be used to download level-1 data of CYGNSS from OPeNDAP for any year from 2019 to 2021, but it can be further extended for any level or year by doing some modifications in the sample array. In this paper, the CYGNSS level-1 version 3.0 data were downloaded for 2020 and 2021 within the sub-basin of the Chambal of the Ganga Catchment in the Indian region. The latitudinal extent of the area is 21°N to 31°N. The longitudinal extent of the region is 73°E to 89°E.

The DDM is the primary surface information observed by the CYGNSS satellite, which maps diffuse scattering power reflected as surface attributes to the receiver between the spatial and Delay-Doppler domains [1]. The land-diffused scattering power consists of coherent and incoherent components. These two components depend on the two types of scatterings, coherent and incoherent [1,5,9,10,12]. After integrating over the glistening elliptical zone around the scattering point, the scattering point's DDM can be written as:

$$|P(\hat{\tau}, \hat{f}_D)|^2 = |P_{coh.}(\hat{\tau}, \hat{f}_D)|^2 + |P_{Incoh.}(\hat{\tau}, \hat{f}_D)|^2 \quad (1)$$

Where  $\hat{\tau}$  and  $\hat{f}_D$  are the delays in the received signal, attenuation due to doppler misalignment. In the case of coherent scattering, the surface roughness does not alter the scattered signal's phase [1,3,10]. Observatory-mounted DDMI collects information on scattering points in the form of diffused power or DDM, which looks like a point size due to its perfect mirror reflection [Fig.1] [10]. In the case of incoherent scattering, the scattered signal has different phases. DDM looks like a horse-shoe shape due to its imperfect mirror reflection [Fig.2]. The spatial footprint of CYGNSS is a  $3.5 \text{ km} \times 0.5 \text{ km}$  diffused area in the case of coherent scattering [1,3,10] and  $25 \text{ km} \times 25 \text{ km}$  diffused area in the case of incoherent scattering. DDM is caused by the different ways signals can be delayed in the space domain. This delay could be caused by things in the air

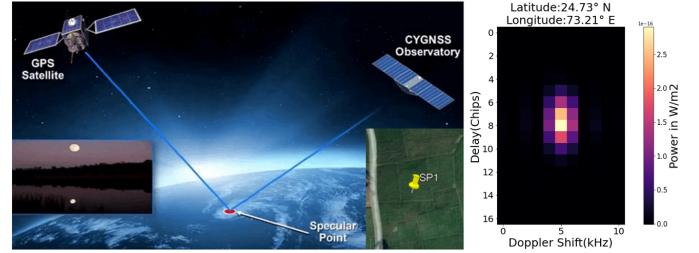


Fig. 1. Delay Doppler Map corresponding to the smooth feature

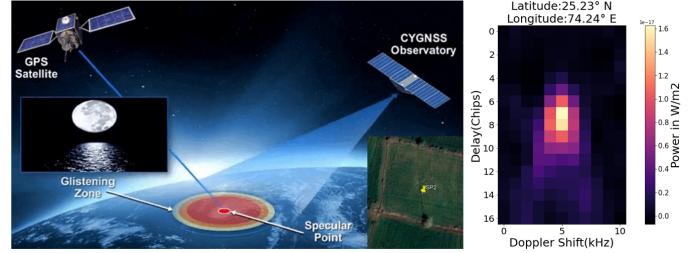


Fig. 2. Delay Doppler Map corresponding to the rough feature

or by vegetation between the transmitting and the receiving antenna. This attenuation of the signal is directly related to the frequency of the signal. So the L1 frequency signal does not lose much power because of the atmosphere [3]. If the peak of the DDM generated from a specular point, that peak could be less affected by vegetation attenuation and surface roughness. Taking the peak of DDM corresponding to each scattering point can reduce the effect of vegetation and all of which reduce the intensity of the reflected signal reaching the CYGNSS DDMI sensor [1,3,10]. Also, it is directly related to the dielectric constant,  $k$ , of the surface, and  $k$  is directly related to the soil moisture within the upper 1-5 centimeters of the soil layer. Soil moisture can be derived using surface reflectivity (corrected DDM due to delays) [1-13]. The bi-static radar equation is derived, ignoring the incoherent scattering contribution in equation (1) (More than 85% signals come from the specular points [1]). Surface reflectivity  $P_{refl}$  is derived using the bi-static radar range equation as:

$$P_{refl} = \frac{(4\pi)^2 P_{peak}(R_r + R_t)^2}{\lambda^2 P_t G_t G_r} \quad (2)$$

Where,  $P_{peak}$ ,  $P_t$ ,  $\lambda$ ,  $G_r$ ,  $G_t$ ,  $R_r$  and  $R_t$  are the Peak of the DDM, Power of the transmitted signal by the GPS mounted active sensor, the wavelength of the L1 frequency signal (0.19 m), Gains of the receiving and transmitting antenna, Ranges between the transmitting antenna and scattering point and range between the scattering point and receiving antenna respectively.

CYGNSS captures the surface properties of the soil as the backscatter after the signal has been reflected from the surface. The amount of radiation reflected from vegetation and soil layers determines the backscatter coefficient [13]. The radar backscatter coefficient, indicated by  $\sigma^0$ , is the radar cross-

section per unit area ( $A$ ) of the specular point [13]. The radar cross-section ( $\sigma$ ) describes how well a specular point on Earth's surface reflects the transmitted signal towards the receiver. The radar backscatter coefficient estimates the amount of the transmitted signal returned to the receiver from the Earth's surface. The backscattering coefficient corresponding to the single footprint of CYGNSS is the effective surface reflectivity measured for that point [10,29].

$$P_{r,eff} = \sigma = \sigma^0 A \quad (3)$$

Ulaby's Water Cloud Model can simulate the total backscatter from the CYGNSS footprint with the soil moisture [28].

## 2.2. GLDAS DATA

The field observations of the soil moisture data are needed to set up the non-linear model of WCM when it is being adjusted. The calibration level-4 version 2.2 daily soil moisture data is available in NASA's Global Land Data Assimilation System in the NetCDF format with a spatial resolution of 25 km and a temporal resolution of 1 day. In this paper, GLDAS 2020 soil moisture data has been taken for initialization.

## 2.3. LAI Data for the Biomass

When applying Water Cloud Model, biomass data was extracted from the daily Leaf Area Index (LAI). LAI measures the one-sided leaf area per unit of the surface area of the vegetated region on the Earth's surface. Daily data sets for the LAI are available in the Climate Data Record on a 5 km grid. The data have been downloaded for the year 2020 within the Ganga Catchment within the Indian region from National Centers for Environmental Information (NCEI).

## 2.4. SMAP DATA

The SMAP (Soil Moisture Active Passive) data were taken as the reference soil moisture data. At 1.5 GHz, the passive radiometer on SMAP measures the microwave radiation that the Earth's surface naturally emits. The radiometer measures how the soil moisture changes the microwave frequencies. The daily data is available at a coarse spatial resolution of 36 km on NSIDC (National Snow and Ice Data Center). In this paper, for the validation of the algorithm and the training purpose of the XGBoost regression model, 2020 and 2021 data have been taken, respectively.

## 2.5. LULC DATA

To visualize the algorithm's performance on different types of land covers within the catchment, the LULC (Land Used Land Cover) annual data for 2019 were taken from Copernicus, having a spatial resolution of 100 m. Data contain 15 types of land cover.

# 3. SOIL MOISTURE ESTIMATION ALGORITHM

## 3.1. Ulaby's Water Cloud Model

The total back-scatter  $\sigma$  for a CYGNSS footprint includes two-way attenuation for propagation between the air-canopy boundary and the canopy-soft boundary and direct back-scattering by soil. Ulaby assumes that a water cloud is composed of identical water particles that scatter signals in a constant phase. The water-cloud assumption comes from the fact that the dielectric constant of water is more significant than vegetation's, which is made up of vegetation and water. Ignoring the higher degree contributions, the back-scatter contributions made by the soil for a given footprint of CYGNSS are as follows [28]:

$$\sigma = \sigma_{veg} + \sigma_{soil} \quad (4)$$

Where  $\sigma_{veg}$  and  $\sigma_{soil}$  represent the contribution of the back-scatter from the vegetation volume and the soil surface, respectively. Ulaby has given the soil back-scattering contribution by a simple expression as follows:

$$\sigma_{soil} = (CM_v + D)\Gamma^2 \quad (5)$$

Where,  $C$  is a frequency and roughness dependent constant,  $M_v$  is the soil moisture and  $\Gamma^2$  is two-way vegetation attenuation in ( $m^2/m^2$ ). Assuming that the water particles are distributed uniformly throughout the canopy volume, they were able to derive an expression for  $\sigma_{veg}$  by integrating the back-scattering contributions of thin strata located between the air-vegetation boundary and the vegetation-soil boundary as follows [28]:

$$\sigma = f(N)\cos\theta(1 - \Gamma^2) + (CM_v + D)\Gamma^2 \quad (6)$$

Assuming that each particle is the same size as the others,  $f(N)$ , the function of the particle size is directly proportional to some power of  $L$ , the biomass of the canopy ( $Kg/m^2$ ) as:

$$f(N) = AL^{E_1}, \Gamma^2 = e^{-2kh\sec\theta} \quad (7)$$

Where  $h$  is the depth of the canopy and  $k$  is the canopy extinction coefficient. The term  $kh$  was found to be directly proportional to some power of the biomass as:

$$kh = BL^{E_2} \quad (8)$$

If the vegetation partially contributes the footprint, then  $\sigma$ , the total back-scatter ( $P_{r,eff}$ ), can be written as:

$$\sigma = AL^{E_1}\cos\theta(1 - e^{\frac{-2BL^{E_2}}{\cos\theta}}) + (CM_v + D)e^{\frac{-2BL^{E_2}}{\cos\theta}} \quad (9)$$

Where  $A$  and  $B$  denote the absorption coefficient and albedo of the vegetative cover in ( $m^2/kg$ ) $^{E_1,2}$  units.  $C$  and  $D$  are the calibration constant and the signal's sensitivity to changes in moisture ( $m^2/m^2$  or dB).  $E_1$  and  $E_2$  are the canopy geometry and water content parameters (unitless). The parameters of the water cloud model are site-dependent. The non-linear least squares regression method, which fits the model to experimental data sets, estimates the parameters  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E_1$ , and  $E_2$  for a specific footprint of CYGNSS [16-30].

### Biomass from the Leaf Area Index:

$$L = 2.22(Foliar_C) + 2.22(Wood_C) + Veg_W \quad (10)$$

$$Foliar_C = LAI \left( \frac{LAI_F}{SLA} \right) \quad (11)$$

$$Wood_C = 1.25(Wd_{fract})(LAI_{MX}) \quad (12)$$

$$Veg_W = 2.22[(Foliar_C)(Foliar_W) + (Wood_C)(Wood_W)] \quad (13)$$

Where, All the biomass is in  $\text{kg}/\text{m}^2$ , Biophysical parameters depending on the land cover [24], are,  $LAI_F$  is all-sided LAI to the one-sided LAI ratio (2.1),  $SLA$  is the Canopy leaf area specific in  $\text{m}^2 \text{kg}^{-1}$  (30),  $Wd_{fract}$  is the proportion of wood to foliar biomass (0.25),  $LAI_{MX}$  is the Annual, monthly maximum LAI in  $\text{m}^2/\text{m}^2$ ,  $Foliar_W$  is the Fraction of foliar biomass (1.3) and  $Wood_W$  is the Wood water content, the proportion of dry biomass (0.54). Since the Ganga catchment is classified as wooded grassland, the corresponding biophysical parameters have been taken from M. Rodell's study, Global Biomass Variation and Its Geodynamic Effects [24].

**Limitation:** Ulaby's Water Cloud Model was developed to monitor the crop field. The model provides a high level of accuracy with higher frequency signals (more than 9 GHz) by reducing the contribution ( $\Gamma^2$ ) of the backscatter from the soil to a level that is negligible in comparison to the contribution of the backscatter from the vegetation [28]. For shorter frequency signals (1.5 GHz), the two-way attenuation,  $\Gamma^2$ , will become close to 1 by neglecting the vegetation's backscatter contribution and increasing the backscatter contribution coming from the soil to 100%. The model is found to be more biased toward barren land with CYGNSS's backscatter, which uses GPS's L1 frequency (1.5 GHz). In this paper, a change was made to Ulaby's water cloud model to make it work well for monitoring crop fields with CYGNSS.

### 3.2. Extension of Water Cloud Model

The CYGNSS backscatter,  $\sigma$  includes the contribution from the soil, water, and vegetation, all of which are weighted by the fraction of each contained within the footprint of the CYGNSS (Fig. 3) [25,28].

$$\sigma = \sigma_s \eta_{os} + \sigma_s \eta_{vs} e^{-\alpha} + \sigma_v \eta_{vs} (1 - e^{-\alpha}) + \sigma_v \eta_{vw} (1 - e^{-\alpha}) + \sigma_w \eta_{ow} + \sigma_w \eta_{vw} e^{-\alpha} \quad (14)$$

$$\eta_{os} + \eta_{vs} + \eta_{ow} + \eta_{vw} = 1 \quad (15)$$

Where  $\sigma_s, \sigma_v$  and  $\sigma_w$  are the back-scatter contributions from the soil, vegetation, and water,  $\eta_{os}$  and  $\eta_{vs}$  are the fractions of open and vegetated soil,  $\eta_{ow}$  and  $\eta_{vw}$  are the fractions of open and vegetated water, respectively,  $\alpha$  is the two-way attenuation factor.

$$\alpha = \frac{2k(h - d)}{\cos\theta} \quad (16)$$

where  $h$  is the vegetation height within the CYGNSS footprint,  $k$  is the canopy extinction coefficient,  $d$  is the depth of

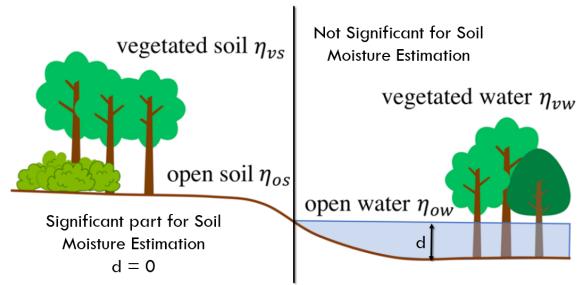


Fig. 3. Contribution of Back-scatter from Heterogeneous Land Surface within CYGNSS footprint

the water, and  $\theta$  is the incidence angle of the signal with Nadir. In this simplified scenario, the footprint is the soil and vegetation's contribution to monitoring soil moisture by CYGNSS. Under this assumption, the total scattering is equal to the sum of the contributions to scattering from vegetated and open soil, respectively. Now the expressions (14) and (15) become as follows:

$$\eta_{ow} = 0, \eta_{vw} = 0, d = 0 \quad (17)$$

$$\eta_{os} + \eta_{vs} = 1, \alpha = \frac{2kh}{\cos\theta} \quad (18)$$

$$\sigma = \sigma_s (1 - \eta_{vs}) + \sigma_s \eta_{vs} e^{-\alpha} + \sigma_v \eta_{vs} (1 - e^{-\alpha}) \quad (19)$$

Which further can be simplified as:

$$\sigma = \sigma_s [1 - \eta_{vs}(1 - e^{-\alpha})] + \sigma_v \eta_{vs} (1 - e^{-\alpha}) \quad (20)$$

$$\sigma = \sigma_s (1 - \eta_{veg}) + \sigma_v \eta_{veg} \quad (21)$$

$$\eta_{veg} = \eta_{vs}(1 - e^{-\alpha}) \quad (22)$$

Assuming the Leaf Area Index data as the function of  $k$ , the canopy extinction coefficient, and  $h$ , the canopy height, two-way attenuation can be written as:

$$\alpha = \frac{2(LAI)}{\cos\theta} \quad (23)$$

The time series of  $\eta_{veg}$  was found to have almost similar behavior as the leaf area index (Figs. 4 and 5). When the contribution of vegetation fraction  $\eta_{vs}$  is 50%, the time series curve of the term does not well fit the January-April month because of the high backscatter contribution from the vegetation (more than 50%). However, as the  $\eta_{vs}$  approached 70%, it nearly superimposed the LAI time series curve, neglecting the effect of the incidence angle on it.  $\eta_{veg}$  can be substituted with the LAI safely from equation (9) and (21).

$$\sigma = (CM_v + D)(1 - LAI) + \sigma_v LAI \quad (24)$$

A higher value of LAI in a highly vegetated region ensures an increase in the contribution of the backscatter from vegetation to the model while simultaneously reducing the contribution of the backscatter from the soil. So, using this modified Water Cloud Model, it is possible to model the backscatter contribution from the vegetation and soil to estimate the soil moisture for CYGNSS.

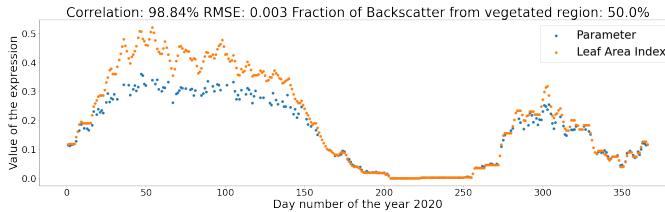


Fig. 4. Time Series of LAI and  $\eta_{veg}$  for a CYGNSS footprint at 22.18°N latitude and 73.18°E longitude when  $\eta_{vs}=0.5$

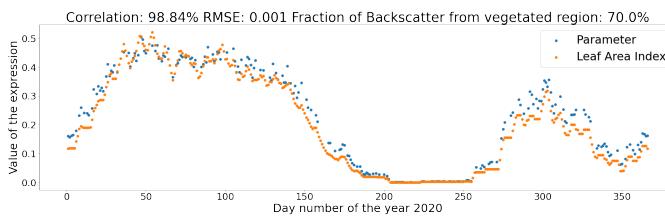


Fig. 5. Time Series of LAI and  $\eta_{veg}$  for a CYGNSS footprint at 22.18°N latitude and 73.18°E longitude when  $\eta_{vs}=0.7$

### 3.3 XGBoost Regression Model

Boosting is an ensemble technique that combines several weaker learners to create a stronger one. Training data creates a model. After that, the second model is built to fix the first's flaws. This process is repeated until the training data set can be predicted correctly or until the maximum number of models can be added. XGBoost, the Extreme Gradient Boosting algorithm, is highly flexible. It uses the power of parallel processing and is faster than other boosting algorithms. With XGBoost, the residual trees are made by determining which variables are used as the roots and nodes by determining the similarity scores between the leaves and the nodes that came before them. It supports regularization. It is designed to handle missing data with its in-built features [31].

CYGNSS observations and the backscatter were resampled using grid averaging inside the SMAP 36 km grid cell for all the CYGNSS samples from 2020 and 2021. Then for every single grid cell, CYGNSS features, incidence angle, gain of the receiving and transmitting antenna, transmitting power of the GPS antenna, DDM SNR, DDM Noise floor, ranges between the specular point to the receiver and the transmitting antenna, peak of power analog, surface reflectivity, and the month of the year, were trained by taking the target variable as the SMAP soil moisture for 2021. For the n estimators, the number of trees was taken as 1000, and then for the best learning rate, hyper-parameter tuning was done by giving 100 learning rates equispaced between 0.1 and 1. Figs (6), (7), and (8) explain the whole procedure of the XGBoost algorithm.

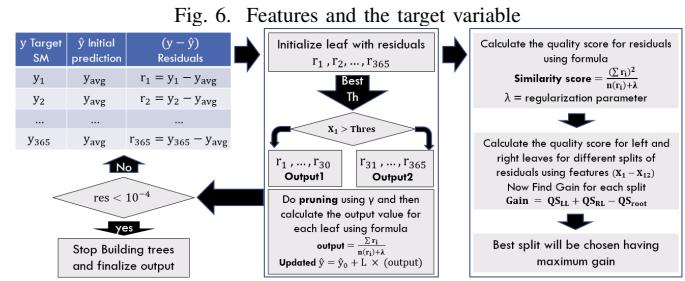
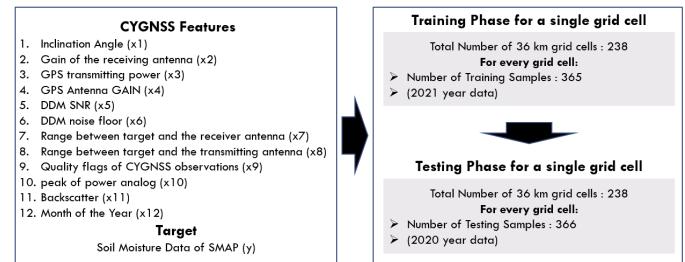


Fig. 6. Features and the target variable

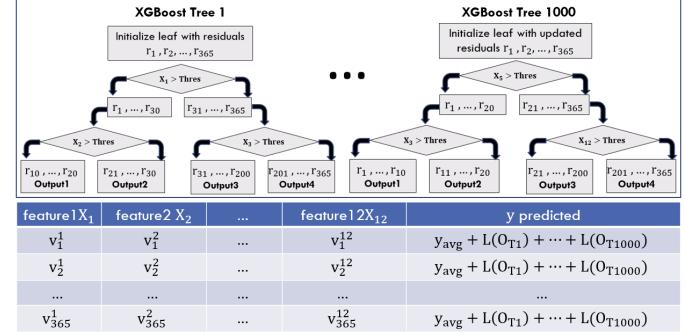


Fig. 7. Training Phase of XGBoost Regression Model



Fig. 8. Testing Phase of XGBoost Regression Model



Fig. 9. Google Earth Image: Centroid of 36 km grid-cell:  
Latitude:24°N and Longitude:77°E near Husaigarh (Vegetated Land)  
Latitude:24°N and Longitude:78°E near Kishanganj(Bare Land)

## 4. RESULTS

### 4.1. Sensitivity analysis between CYGNSS Backscatter and SMAP soil moisture

The CYGNSS backscatter was resampled along with the standard deviation within the SMAP 36 km grid cells to analyze the temporal sensitivity analysis with SMAP soil moisture. The correlation for barren land devoid of vegetation or canopy was quite strong (81.6%). However, the correlation was reduced on the surface of the vegetated land (81.2%).

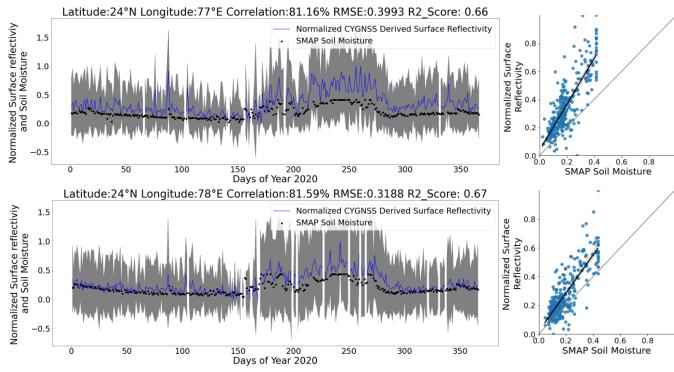


Fig. 10. Temporal Correlation with CYGNSS surface reflectivity and SMAP soil moisture data for a pixel surrounded with the vegetated and barren land features

Also, the standard deviations are stable until July, but in November and December, they change a lot [figs 9,10]. Each of the 238  $36^\circ \times 36^\circ$  grid cells in the Chambal catchment area was subjected to temporal correlation analysis. A grid with 238 grid cells was used to store the correlations and the coefficient of determination ( $R^2$ ) for each pair. Using the spatial interpolation technique kriging, they were interpolated inside the grid size of  $3^\circ \times 3^\circ$ . After that, the results were summarized to show how well the CYGNSS backscatter correlated with the SMAP soil moisture in different parts of the catchment (Fig. 11).

Near the cropland and the partly forested area near the watershed, there is a strong correlation between soil moisture and backscatter. For example, Shajapur and Ujjain both have correlations of 75% with a coefficient of determination ( $R^2$ ) of 50%. On the other side, the correlation has been observed to be lower than 30% in built-up areas such as Hosangabad and Udaipur, as well as Bundi and Kota. This is because these areas have higher backscatter values, expecting high soil moisture there because of their built-in properties. Still, according to the SMAP algorithm, they have a low soil moisture value. So due to this inverse behavior of backscatter and soil moisture, there are low correlations near the built-in areas. Also, the correlations are excellent in the forest region with sparse vegetation and forest, with examples such as Sagar, Raisen, and Gwalior having correlations of 60, 75, and 75%, respectively. Additionally, certain built-up areas close to the river Chambal, such as Neemuch and Chittaurgarh, have a moderate correlation because they are at least 36 km apart. SMAP gives the soil moisture at a resolution of 36 km, so if the built-in areas are at least 36 km, then CYGNSS backscatter will expect the soil moisture there, which will be almost the same for the SMAP soil moisture. Therefore, these regions near the built-in area show high correlations. This is because these areas have soil moisture since they are located close to the river. Similarly, the built-in area near Banswara and Indore shows a 75% correlation. It has been shown that the backscatter and soil moisture have a nearly 1:1 correlation.

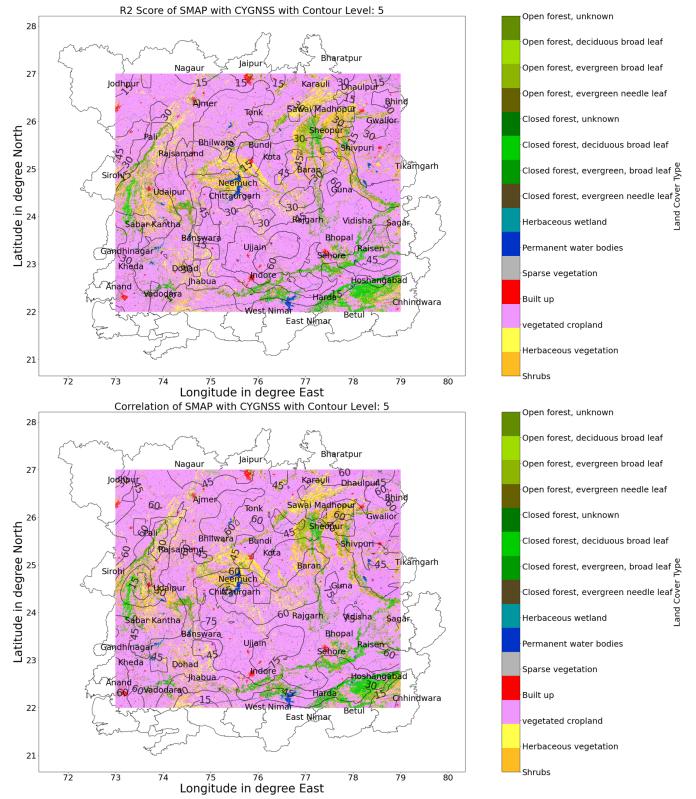


Fig. 11. Contour Plot of Coefficient of Determination ( $R^2$ ) and the Coefficient of Correlation between the CYGNSS surface reflectivity and SMAP soil moisture with the Land Cover

When a fixed area of ground with the same soil composition is probed, changes in surface reflectivity can stand in for changes in soil moisture near the surface.

#### 4.2. Adjustment of Ulaby's Water Cloud Model

$$P_{r,eff} = AL^{E_1} \cos\theta (1 - e^{-2BL^{E_2} \sec\theta}) + (CM_v + D)e^{-2BL^{E_2} \sec\theta} \quad (25)$$

The non-linear least squares regression analysis, which involves fitting the model to experimental data sets, is used to estimate the values of the parameters A, B, C, and D, as well as  $E_1$ ,  $E_2$ , and  $M_v$ , for a certain location in the field of interest. Within a single SMAP grid cell of  $36 \times 36 \text{ km}^2$ , the average daily count of CYGNSS observations was found to be more than ten. So the daily non-linear least squares adjustment for the soil moisture parameter can be performed in each grid cell having more than seven observations available daily. The biomass was derived from LAI using equation (10). The GLDAS soil moisture value is the initialized value of soil moisture. During the adjustment, initializing the model parameters A, B, C, D,  $E_1$ , and  $E_2$  is done using hit and trial. As can be seen, there are a total of 13 observations of CYGNSS data in a single day, each corresponding to a single pixel of SMAP soil moisture data. So, using these 13 observations, a non-linear least squares

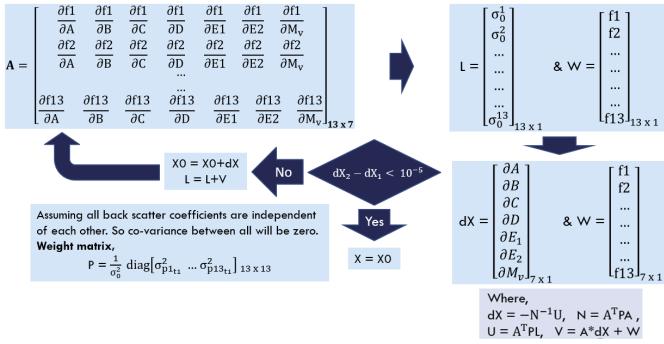


Fig. 12. Flowchart illustrating the process of performing non-linear least square adjustment with a single day of data

adjustment can be made for seven unknown parameters ( $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E_1$ ,  $E_2$ , and  $M_v$ ), making 13 observation equations with a redundancy of 6. First, the design matrix and residual matrix were calculated using the initial values of the parameters, the Jacobian, and the weight matrix for the observation; then, the initial parameters were updated using this. This process was repeated until there were no more updates to the parameters. The flowchart under Fig. 12 shows the entire adjustment process.

$$\begin{aligned} f_1 &= AL_1^{E_1} \cos\theta_1 (1 - e^{-2BL_1^{E_2} \sec\theta_1}) + \\ &\quad (CM_v + D)e^{-2BL_1^{E_2} \sec\theta_1} - P_{r,eff}^1 \\ &\dots \\ f_{13} &= AL_{13}^{E_1} \cos\theta_{13} (1 - e^{-2BL_{13}^{E_2} \sec\theta_{13}}) + \\ &\quad (CM_v + D)e^{-2BL_{13}^{E_2} \sec\theta_{13}} - P_{r,eff}^{13} \end{aligned}$$

Every pixel in the catchment has had Ulaby's water cloud model adjusted for soil moisture every 366 days in 2020. The variations in model parameters  $A$ ,  $B$ ,  $C$ ,  $E_1$  and  $E_2$  over an entire year for a single SMAP grid cell are visualized; the values were found to be 0.009, 0.096, 13.9, -5, and -190, respectively. However, the value of  $D$  can range anywhere from -0.04 to 0.01, with a mean of -0.025 [Fig. 13]. Due to the high negative values of  $E_1$  and  $E_2$ , the contribution made by the vegetation fraction was found to be almost insignificant. The temporal correlation between adjusted and SMAP soil moisture was more significant than 84.6%, with an RMSE of 0.0053  $m^3/m^3$ , which was an improvement over the RMSE and correlation between GLDAS and SMAP, which was 0.12  $m^3/m^3$  and 67.1%, respectively [Fig. 14].

During daily adjustment, days with fewer than 7 CYGNSS observations were disregarded. To account for those days having at least one observation, batch adjustments of whole-year data were made. More than 3000 observations are available for every single SMAP grid cell. The adjustment was performed for six unknown model parameters ( $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E_1$ ,  $E_2$ ) and 272 unknown soil moisture parameters ( $M_{vt6}$ ,  $M_{vt8}$ ,  $M_{vt16}$ , ...,  $M_{vt362}$ ), resulting in 4203 observation equations

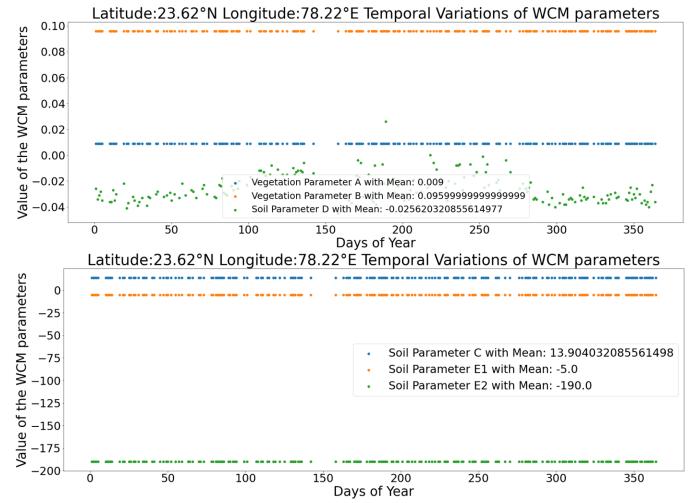


Fig. 13. Temporal Variations of the Ulaby's Water Cloud Model Parameters

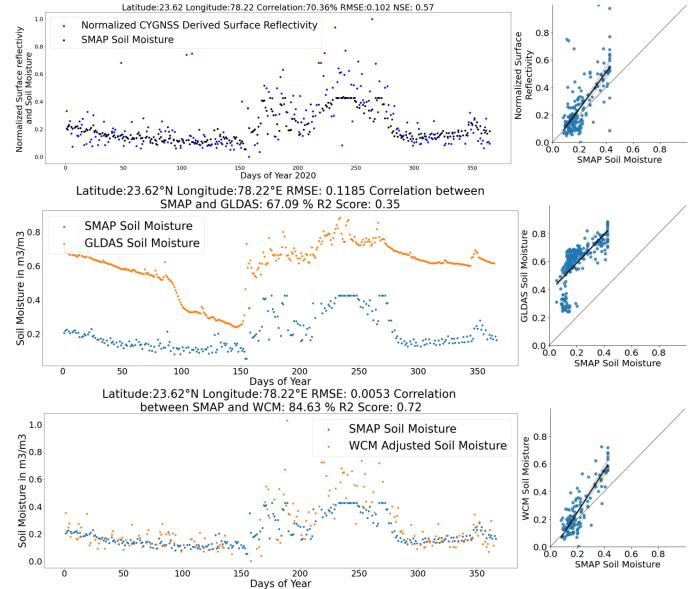


Fig. 14. Temporal correlation of the CYGNSS backscatter and SMAP soil moisture  
GLDAS soil moisture and daily adjusted soil moisture with SMAP soil moisture

with a redundancy of 3925. The observation equations for day six can be written as follows:

$$\begin{aligned} f_{1t6} &= AL_{1t6}^{E_1} \cos\theta_{1t6} (1 - e^{-2BL_{1t6}^{E_2} \sec\theta_{1t6}}) + \\ &\quad (CM_{vt6} + D)e^{-2BL_{1t6}^{E_2} \sec\theta_{1t6}} - (P_{t6r,eff}^1) \\ &\dots \\ f_{13t6} &= AL_{13t6}^{E_1} \cos\theta_{13t6} (1 - e^{-2BL_{13t6}^{E_2} \sec\theta_{13t6}}) + \\ &\quad (CM_{vt6} + D)e^{-2BL_{13t6}^{E_2} \sec\theta_{13t6}} - (P_{t6r,eff}^{13}) \end{aligned}$$

Similarly, the observation equations for Day 8 observations

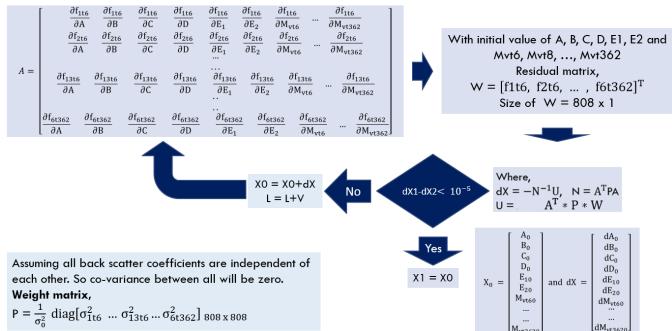


Fig. 15. Flowchart depicting the process of carrying out non-linear least square adjustment for a particular 36-kilometer grid cell

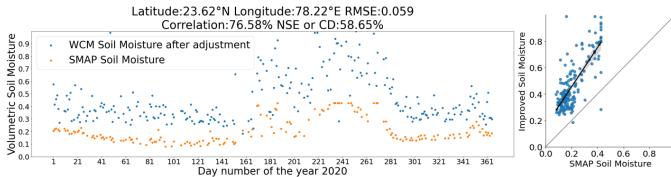


Fig. 16. Temporal correlation of the batch adjusted soil moisture with the SMAP soil moisture

can be written as follows:

$$\begin{aligned}
 f_{1t8} &= AL_{1t8}^{E_1} \cos\theta_{1t8} (1 - e^{-2BL_{1t8}^{E_2} \sec\theta_{1t8}}) + \\
 &\quad (CM_{vt8} + D)e^{-2BL_{1t8}^{E_2} \sec\theta_{1t8}} - (P^1_{t8r,eff}) \\
 &\dots \\
 f_{4t8} &= AL_{4t8}^{E_1} \cos\theta_{4t8} (1 - e^{-2BL_{4t8}^{E_2} \sec\theta_{4t8}}) + \\
 &\quad (CM_{vt8} + D)e^{-2BL_{4t8}^{E_2} \sec\theta_{4t8}} - (P^4_{t8r,eff})
 \end{aligned}$$

The observation equations for days 16, 17, 19,... 361, 362 can be written in the same way.

After the adjustment, the temporal correlation between adjusted and SMAP soil moisture was more than 76.6% with an RMSE of  $0.06 \text{ m}^3/\text{m}^3$ , which improved the temporal correlation between GLDAS and SMAP, which was found to be more than 71.1% with an RMSE of  $0.12 \text{ m}^3/\text{m}^3$  (Fig. 16). Even though the daily adjustment gave a high correlation, the batch adjustment needs to be done because it takes into account all the observations. The adjustment was performed individually at each of the 238 grid cells, and the respective correlations, coefficients of determination (R squared), RMSE, and relative errors were stored. The values were interpolated based on the grid size of the array, which is 3 kilometers. The results were summarized to visualize the Ulaby Model's performance across the entire catchment. The RMSE of adjusted soil moisture has significantly decreased compared to the RMSE obtained from GLDAS with SMAP soil moisture (Fig. 17).

A good correlation can be seen near the cropland area and the partially vegetated area close to the catchment. For

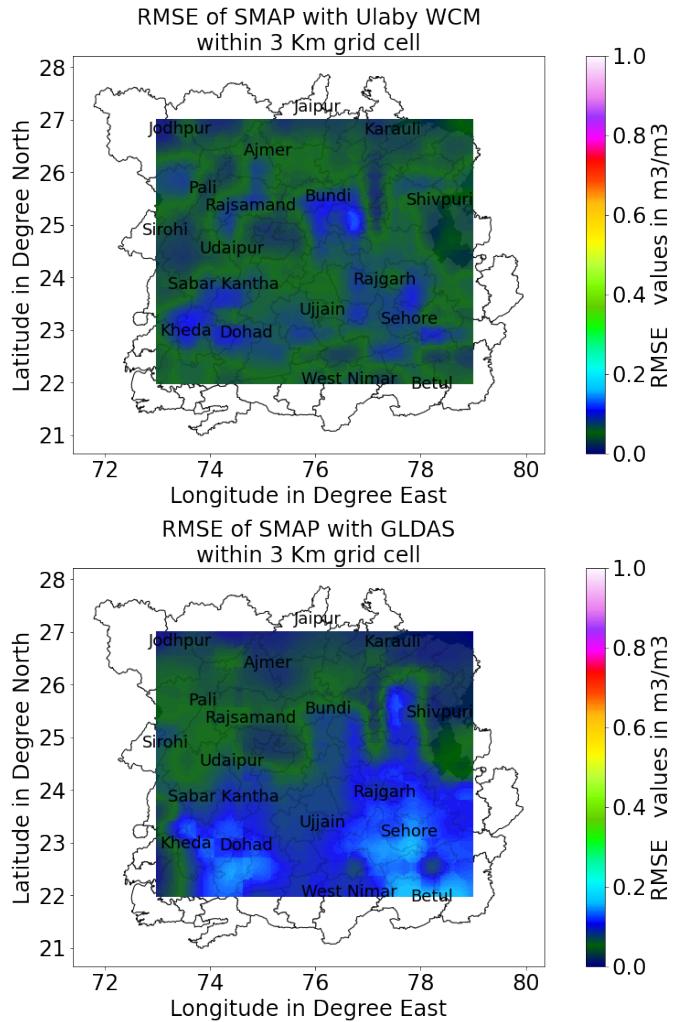


Fig. 17. RMSE  $\text{m}^3/\text{m}^3$  between the adjusted and Initialized GLDAS Soil Moisture with SMAP soil moisture

example, Sagar and Ujjain both correlate 75%. On the other hand, the model could be doing better in built-up areas like Hosangabad, Udaipur, Pali, Chhindwara, and Kota, which have less than 30% correlations. These areas have high backscatter, expecting high soil moisture because of built-in properties. Still, according to the SMAP algorithm, they have a low soil moisture value. So due to this inverse behavior of surface reflectivity and soil moisture, there are low correlations near the built-in areas. Also, the model functions well in forested areas such as Sagar, Raisen, and Gwalior, with correlations of 60, 75, and 75%, respectively. Additionally, certain built-up areas close to the river Chambal, such as Neemuch and Chittaurgarh, have a moderate correlation because they are at least 36 km apart. SMAP gives the soil moisture at a resolution of 36 km, so if the built-in areas are at least 36 km, then Ulaby's WCM will expect the soil moisture there, which will be almost the same as the SMAP soil moisture. Therefore, these regions near the built-in area show high correlations. This is because these areas have soil moisture since they are located

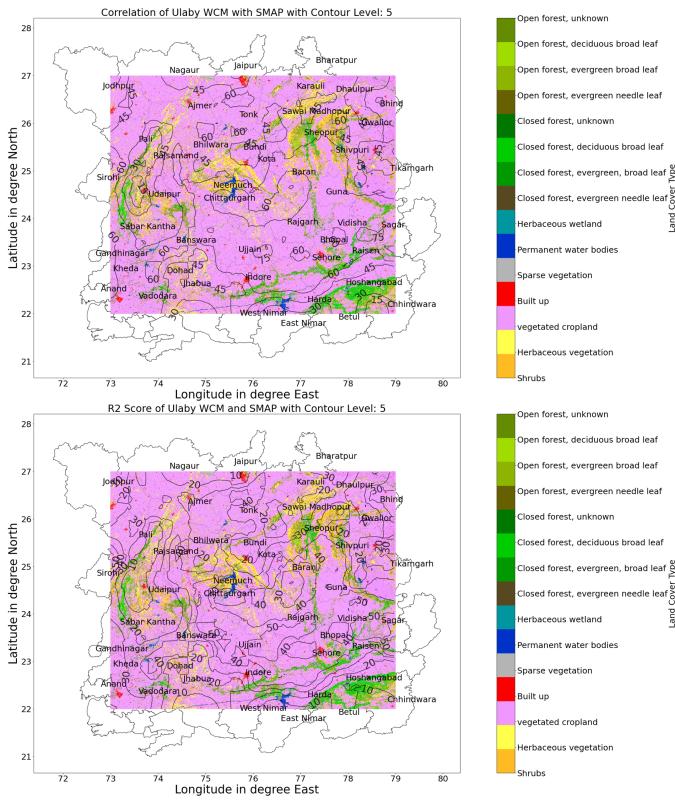


Fig. 18. Contour Plot of Coefficient of Correlation and Coefficient of Determination between the adjusted and SMAP soil moisture with the Land Cover

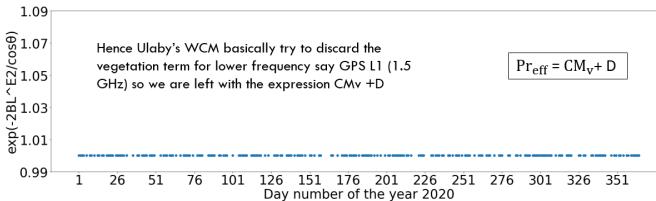


Fig. 19. Temporal variation of the exponential term is almost constant, 1 neglecting the vegetation term from the Ulaby's WCM for GPS L1 frequency, Which is Lower than the threshold set for vegetation attenuation during FT Ulaby's Experiment.

close to the river; similarly, built-in areas near Banswara and Indore show a 75% correlation. However, areas farther than 36 km from the catchment show only a moderate level of correlation [Fig. 18].

The Water Cloud Model was thought to use a microwave signal with a frequency of more than 8.6 GHz to track the growth of crops. However, the GPS-transmitted signal gives off backscatter with an L1 frequency of 1.5 GHz and a wavelength of 19 centimeters. The canopy geometry parameters  $E_1$  and  $E_2$ , -5 and -190, respectively [Fig. 19], are responsible for vanishing the vegetation backscatter term containing the vegetation attenuation parameters A and B after adjustment due to their higher negative values.

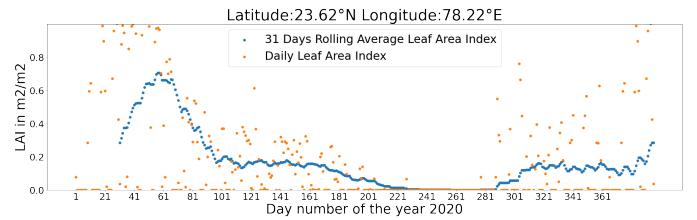


Fig. 20. Temporal variation of Leaf Area Index

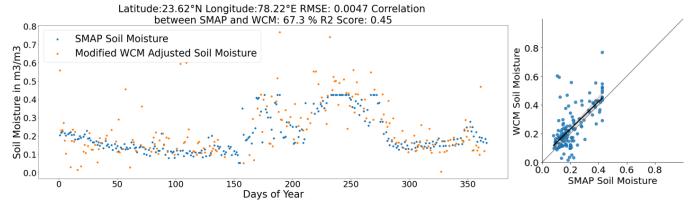


Fig. 21. Temporal correlation of adjusted soil moisture with the SMAP soil moisture

So, the model did not consider the attenuation caused by vegetation. Instead, it acted like a linear model for microwave frequencies lower than 9 GHz and like a nonlinear model for microwave frequencies higher than 9 GHz, taking attenuation into account [28].

#### 4.3. Adjustment of modified Water Cloud Model

$$\sigma = (CM_v + D)(1 - LAI) + \sigma_v LAI \quad (26)$$

Here, the soil moisture and the backscatter from the vegetation were of particular interest. The hit-and-try method was used to figure out the initial values for C and D, which were found to be 20 and -1, respectively. In order to set up the backscatter from the vegetated area, the CYGNSS backscatter in each grid cell was used to find the lowest backscatter. To obtain more than 600 parameters, more than 2,000 observation equations were considered. After batch adjustments, there was an improvement in the correlations (up to 67%) and the RMSE (up to 0.0047) between the adjusted soil moisture on the GLDAS and the SMAP [Fig. 14, 21]. After performing the batch adjustment at each of the 238 grid cells individually, the respective correlations, coefficients of determination ( $R^2$  squared), RMSE, and relative errors were stored within a grid with a total of 238 grid cells. After that, those values were interpolated from 36 kilometers to a three-kilometer grid cell. The results were summed up so that it could be seen how well the Extended Ulaby's Model compared to Ulaby's original WCM across the whole catchment [Figs. 22,23 and 24].

After comparing the performances of Ulaby's and Modified Ulaby's Models in the Chambal catchment, it can be seen that the modified WCM does a worse job in terms of correlation and coefficient of determination in most of the Chambal catchment (near Pali, Shivpuri, Sirohi, Ujjain, and Udaipur) [Figs. 24 and 18]. However, the modified WCM

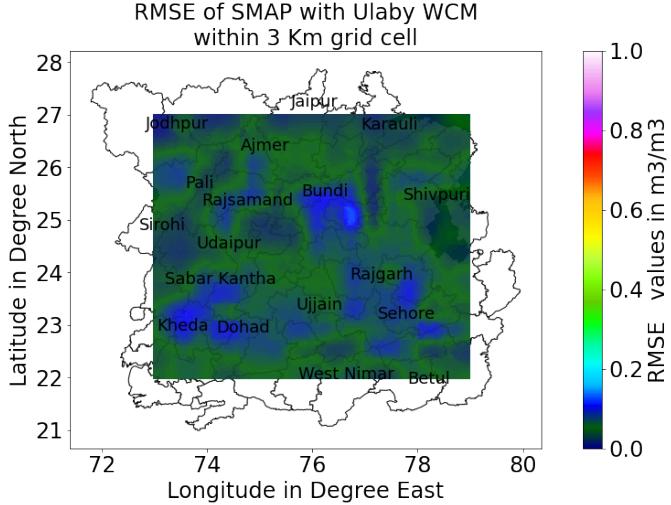


Fig. 22. RMSE of Adjusted Soil Moisture with respect to SMAP

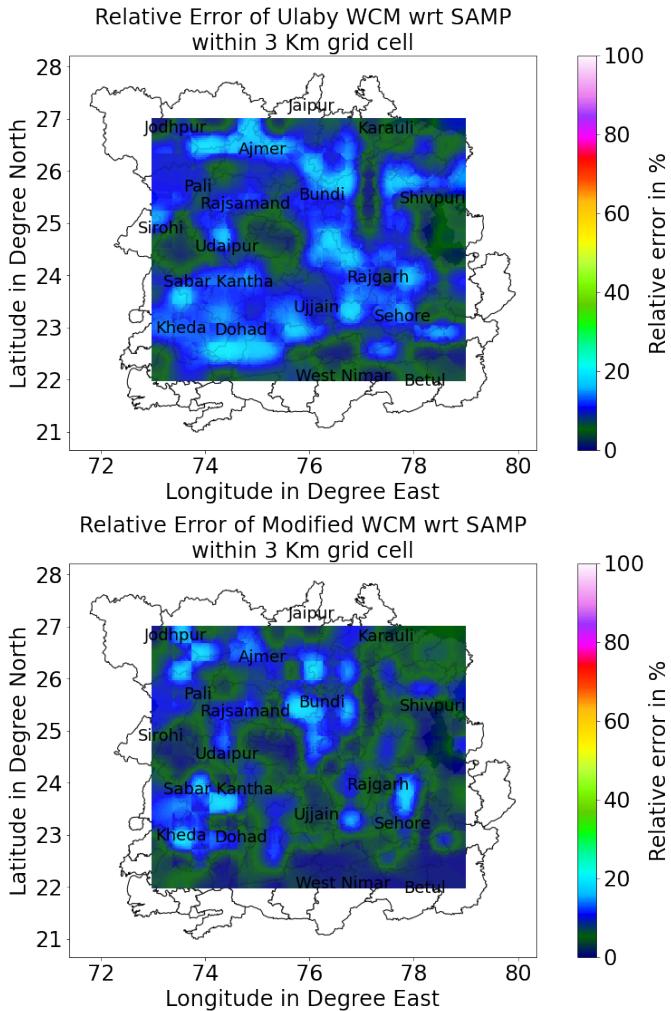


Fig. 23. Relative error in Adjusted Soil Moisture with respect to SMAP using Modified WCM comparing with the Ulaby's WCM

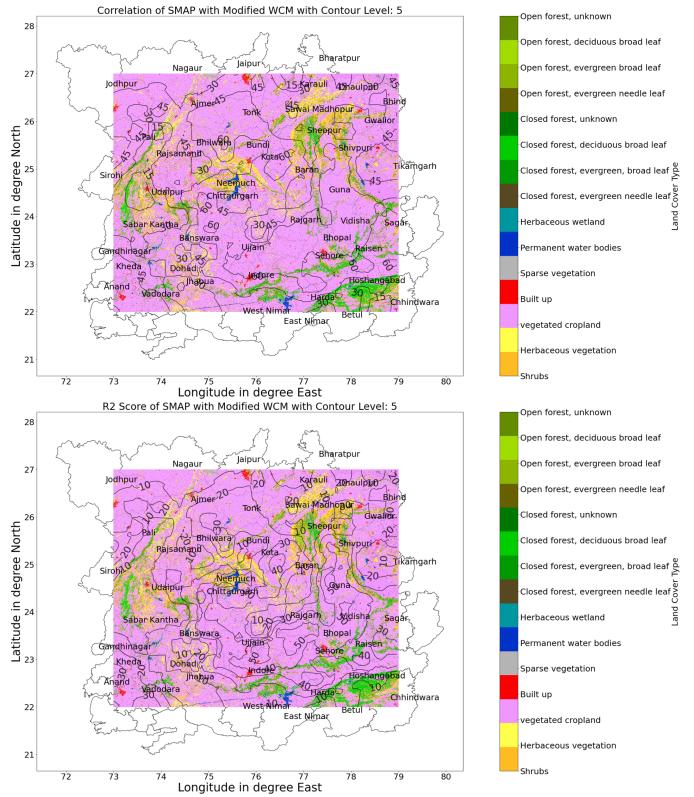


Fig. 24. A contour plot of the coefficient of correlation and the coefficient of determination ( $R^2$  square) of Adjusted Soil Moisture with respect to SMAP using Modified WCM with the land cover

is doing an excellent job in some of the forest areas of the Chambal catchment. For example, near Tikamgarh and Sheopur, the correlation improved from 10 to 30; near Bundi, the correlation improved from 10 to 20; near Dohad, the correlation improved from 10 to 40. In most of the Chambal catchment (near west Nimar, Betul, Ujjain, Shivpuri, Jodhpur, and Ajmer), Modified WCM is doing a bad job based on the RMSE [Figs 24 and 18]. Also, the RMSE shows that the modified WCM is doing an excellent job in some of the forest areas of the Chambal catchment, such as Sehore, Bundi, Kheda, Rajgarh, and Dohad [Figs 17 and 22]. So, in terms of correlations and coefficients of determination with SMAP soil moisture, the modified model does a better job in the forest area and a worse job near the built-up area than Ulaby's WCM. Modified WCM is doing a comparatively good job in terms of the relative error of adjusted soil moisture to SMAP soil moisture in most of the area of the Chambal catchment (near Ajmer, Ujjain, Rajsamand, Rajgarh, and Dohad). Also, it can be seen that the modified WCM could be doing a better job in some of the forest areas of the Chambal catchment [Sabar, Bundi, Kheda, and Jodhpur] [Fig 23].

#### 4.4. XGBoost Regression Model

After running XGBoost regression on each of the 238 grid cells individually, the correlations, coefficients of

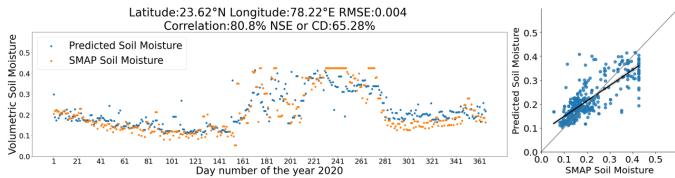


Fig. 25. Temporal correlation, RMSE and Coefficient of determination of the XGBoost predicted with SMAP soil moisture

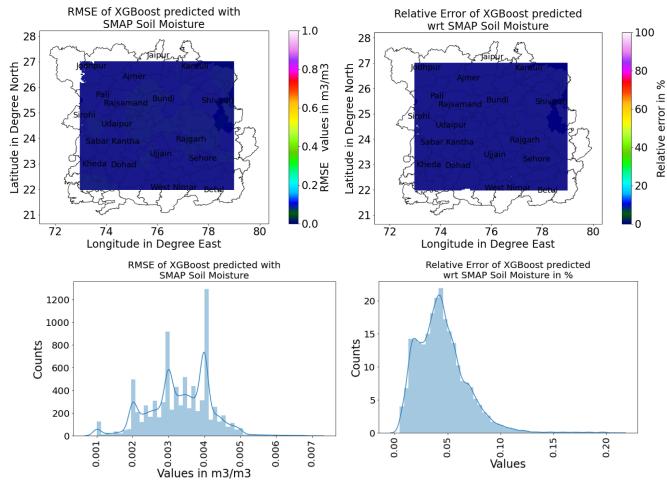


Fig. 26. RMSE and Relative error of the XGBoost predicted with SMAP soil moisture (comparing with Figs 17, 22 and 23)

determination ( $R^2$ ), RMSE, and relative errors were saved in a grid with a total of 238 grid cells. Then those values were interpolated from 36 kilometers to a 3-kilometer grid cell. The results showed how well the XGBoost model performed compared to Ulaby's original WCM and Extended Ulaby's Model across the whole catchment. The results showed much improvement in the correlations compared to Ulaby's and the modified WCM. Furthermore, the RMSE improved to 0.005 and the relative error to 0.2%, which are significantly lower than the previous models. The XGBoost did a better job in most of the Chambal catchment (near Pali, Shivpuri, Sirohi, Ujjain, and Udaipur) by improving the correlations, RMSE, and relative error [Figs 26, 27, 23, 24, 18 and 17]. For example, the correlation improved from 77% [from Water Cloud Model] to more than 80% and the RMSE to 0.003 in the cropland area. In the vegetated region, correlations improved up to only 75%. Here, one-year data has been taken for training purposes. To improve the correlations in areas with vegetation and forests, more data from CYGNSS must be used for training.

To improve the accuracy of the XGBoost regression model by adding extra features to the data, combining the biomass or canopy data as an ancillary feature with the CYGNSS features while training could improve the correlations in the cropland, vegetated, and forest regions with more than 90% accuracy.

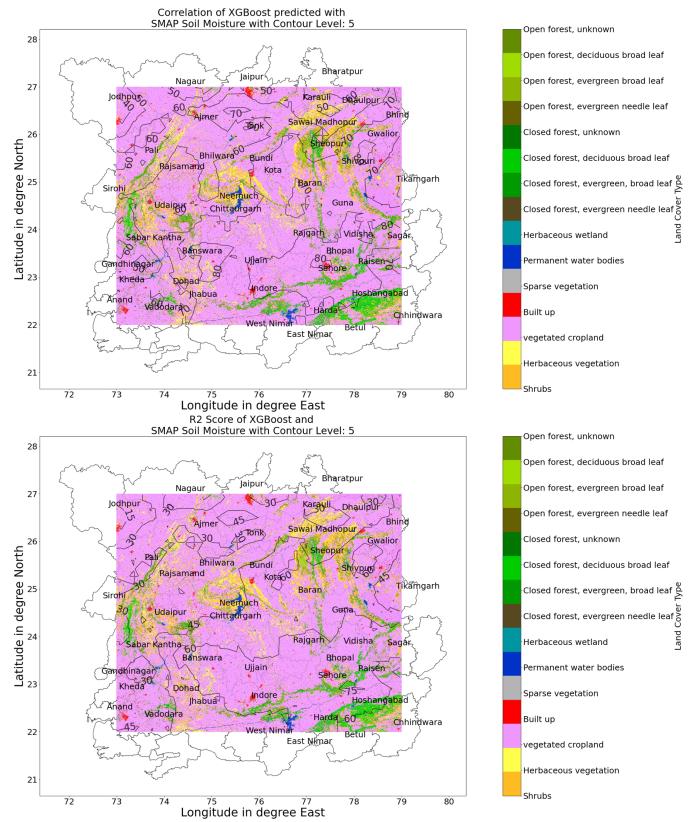


Fig. 27. A contour plot of the coefficient of correlation and coefficient of determination of predicted Soil Moisture with respect to SMAP using XGBoost with the land cover (comparing with Figs 18 and 24)

## 5. CONCLUSION

Python software was made to solve the problem of downloading and sub-setting CYGNSS data while reducing the amount of work that had to be done by hand and the amount of space needed. The software lets users choose the CYGNSS variables they want to see, the time range they want to look at, and the size of the area for which they want to download the data.

More than 80% temporal correlation of CYGNSS surface reflectivity with SMAP soil moisture data was reported by C. C. Chew in 2018 [10]. A temporal correlation was found between the backscatter and the SMAP soil moisture data of more than 81%. Furthermore, it was discovered that the temporal correlation for arid land devoid of vegetation or a canopy is somewhat high (81.6%, RMSE = 0.28  $m^3/m^3$ ). The correlation was lower on the vegetated land surface (81.2%, RMSE = 0.24  $m^3/m^3$ ). Also, the standard deviations of surface reflectivity were almost the same until July, but they were high in November and December. The relationship between the surface reflectivity of CYGNSS and SMAP soil moisture data was nearly 1:1.

CYGNSS observations were used to adjust the water-cloud model to estimate soil moisture. It was found that the average

values of model parameters A, B, C, E1, and E2 remained constant for a single grid cell. On the other hand, D was found to vary slightly. Ulaby's Model was found to be frequency sensitive and useful for microwave signals with a frequency greater than 9 GHz [28]. This model showed a linear behavior for microwave frequencies less than 9 GHz and would not consider vegetation attenuation, but for frequencies greater than 9 GHz, it would consider vegetation attenuation.

The modifications were done to the original Model Ulaby to make it worthwhile for crop field monitoring. It was found that the extended WCM improved the RMSE value more than Ulaby's WCM. The extended WCM performs poorly in the built-in area but excels in the forest area, which is the cause of the poor overall performance of the model.

The XGBoost regression model improved the correlation from 77% [from the Water Cloud Model] to more than 87% in the cropland area. The RMSE was reduced to 0.003. Also, the relative error found was not more than 0.2%. In the vegetated region, correlations improved up to only 75%. Here, one-year data has been taken for training purposes. It is necessary to add more and more data to CYGNSS for training to improve the correlations in the vegetated and forest regions. Also, extra features must be added to the data to achieve a higher correlation than 90%.

There is a need to improve Ulaby's empirical water cloud model by including the vegetation extinction coefficient depth of water and canopy geometry for vegetation attenuation to capture the surface reflectivity pattern in the forest and built-up areas.

## REFERENCES

- [1] Dong,Z.;Jin,S. Evaluation of the Land GNSS-Reflected DDM Coherence on Soil Moisture Estimation from CYGNSS Data. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China; zndong@shao.ac.cn 2 School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing, 100049, China. *Remote Sens.* 2021, 13, 570.
- [2] A. Halimi, C. Mailhes, J. -Y. Tourneret, P. Thibaut and F. Boy, "A Semi-Analytical Model for Delay/Doppler Altimetry and Its Estimation Algorithm," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4248-4258, July 2014.
- [3] Woodhouse, I.H. (2006). *Introduction to Microwave Remote Sensing* (1st ed.). CRC Press.
- [4] Edokossi, K.; Calabia, A.; Jin, S.; Molina, I. GNSS-Reflectometry and Remote Sensing of Soil Moisture: A Review of Measurement Techniques, Methods, and Applications. *Remote Sens.* 2020, 12, 614.
- [5] Shivani Tyagi; Dharmendra Kumar Pandey; Deepak Putrevu; Arundhati Misra, 'Sensitivity Analysis of CYGNSS derived Radar Reflectivity for Soil Moisture Retrieval over India: Initial Results', 2019 IEEE 16th India Council International Conference (INDICON).
- [6] Linqi Zhang, Yi Liu, Liliang Ren, Adriaan J. Teuling, Xiaoxiang Zhang, Shanhui Jiang, Xiaoli Yang, Linyong Wei, Feng Zhong, Lihong Zheng, Reconstruction of ESA CCI satellite-derived soil moisture using an artificial neural network technology, *Science of The Total Environment*, Volume 782, 2021, 146602, ISSN 0048-9697.
- [7] R. Fernandez-Moran, J.-P. Wigneron, G. De Lannoy, E. Lopez-Baeza, M. Parrens, A. Mialon, A. Mahmoodi, A. Al-Yaari, S. Bircher, A. Al Bitar, P. Richaume, Y. Kerr, A new calibration of the effective scattering albedo and soil roughness parameters in the SMOS SM retrieval algorithm, *International Journal of Applied Earth Observation and Geoinformation*, Volume 62, 2017, Pages 27-38, ISSN 1569-8432.
- [8] Y. H. Kerr et al., "The SMOS Soil Moisture Retrieval Algorithm," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1384-1403, May 2012.
- [9] Y. Jia et al., "Temporal-Spatial Soil Moisture Estimation from CYGNSS Using Machine Learning Regression With a Preclassification Approach," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4879-4893, 2021.
- [10] Chew, C. C., Small, E. E. (2018). Soil Moisture Sensing Using Spaceborne GNSS Reflections: Comparison of CYGNSS Reflectivity to SMAP Soil Moisture. *Geophysical Research Letters*, 45(9), 4049-4057.
- [11] Zhang, S., Roussel, N., Boniface, K., Ha, M. C., Frappart, F., Darrozes, J., Baup, F., and Calvet, J.-C.: Use of reflected GNSS SNR data to retrieve either soil moisture or vegetation height from a wheat crop, *Hydrol. Earth Syst. Sci.*, 21, 4767-4784.
- [12] Hyunglok Kim, Venkat Lakshmi: Use of Cyclone Global Navigation Satellite System (CyGNSS) Observations for Estimation of Soil Moisture.
- [13] L. Karthikeyan, Ming Pan, Niko Wanders, D. Nagesh Kumar, Eric F. Wood, Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms, *Advances in Water Resources*, Volume 109, 2017, Pages 106-120, ISSN 0309-1708.
- [14] Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and Wood, E. F., "Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons", *Advances in Water Resources*, vol. 109, pp. 236–252, 2017.
- [15] Edokossi, K.; Calabia, A.; Jin, S.; Molina, I. GNSS-Reflectometry and Remote Sensing of Soil Moisture: A Review of Measurement Techniques, Methods, and Applications. *Remote Sens.* 2020, 12, 614.
- [16] Park, S.-E.; Jung, Y.T.; Cho, J.-H.; Moon, H.; Han, S.-h. Theoretical Evaluation of Water Cloud Model Vegetation Parameters. *Remote Sens.* 2019, 11, 894.
- [17] C. Liu and J. Shi, "Estimation of Vegetation Parameters of Water Cloud Model for Global Soil Moisture Retrieval Using Time-Series L-Band Aquarius Observations," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5621-5633, Dec. 2016.
- [18] Fawwaz T. Ulaby, Pascale C. Dubois, Jakob van Zyl, Radar mapping of surface soil moisture. *Journal of Hydrology*, Volume 184, Issues 1–2, 1996, Pages 57-84, ISSN 0022-1694.
- [19] Kumar, Kamal Prasad, K. Arora, M.. (2012). Estimation of water cloud model vegetation parameters using a genetic algorithm. *Hydrological Sciences Journal-journal Des Sciences Hydrologiques-HYDROLOG SCI* J.57.1-14.
- [20] Hosseini, M. and McNairn, H., "Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields", *International Journal of Applied Earth Observation and Geoinformation*, vol. 58, pp. 50–64, 2017.
- [21] Kweon, Soon-Koo Hwang, Ji-Hwan Oh, Yisok. (2012). COSMO SkyMed AO projects -soil moisture detection for vegetation fields based on a modified water-cloud model using COSMO-SkyMed SAR data. *International Geoscience and Remote Sensing Symposium (IGARSS)*.

- [22] L. Prévot, I. Champion, G. Guyot, Estimating surface soil moisture and leaf area index of a wheat canopy using a dual-frequency (C and X bands) scatterometer, *Remote Sensing of Environment*, Volume 46, Issue 3, 1993, Pages 331-339, ISSN 0034-4257.
- [23] Tsan Mo, Thomas J. Schmugge, Thomas J. Jackson, Calculations of radar backscattering coefficient of vegetation-covered soils, *Remote Sensing of Environment*, Volume 15, Issue 2, 1984, Pages 119-133, ISSN 0034-4257.
- [24] Rodell M, Chao BF, Au AY, Kimball JS, McDonald KC. Global biomass variation and its geodynamic effects: 1982-98 *Earth Interactions*.
- [25] M. Lavalle et al., "Bistatic Scattering Modeling for Dynamic Mapping of Tropical Wetlands with Cygnss," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 239-242.
- [26] Damien Sulla-Menashe, Josh M. Gray, S. Parker Abercrombie, Mark A. Friedl, Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product, *Remote Sensing of Environment*, Volume 222, 2019, Pages 183-194, ISSN 0034-4257.
- [27] F.T. Ulaby, C.T. Allen, G. Eger, E. Kanemasu, Relating the microwave backscattering coefficient to leaf area index, *Remote Sensing of Environment*, Volume 14, Issues 1–3, 1984, Pages 113-133, ISSN 0034-4257.
- [28] Y. Oh, K. Sarabandi and F. T. Ulaby, "An empirical model and an inversion technique for radar scattering from bare soil surfaces," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 2, pp. 370-381, March 1992.
- [29] Singh, U., Srivastava, P.K., Pandey, D.K., Chaurasia, S. (2020). Assessment of SCATSAT-1 Backscattering by Using the State-of-the-Art Water Cloud Model. In: Ghosh, J., da Silva, I. (eds) *Applications of Geomatics in Civil Engineering. Lecture Notes in Civil Engineering*, vol 33. Springer, Singapore.
- [30] M. Imhoff, M. Story, C. Vermillion, F. Khan and F. Polcyn, "Forest Canopy Characterization and Vegetation Penetration Assessment with Space-Borne Radar," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-24, no. 4, pp. 535-542, July 1986.
- [31] Jia, Yan Jin, Shuanggen Savi, Patrizia Gao, Yun Tang, Jing Chen, Yixiang Li, Wenmei. (2019). GNSS-R Soil Moisture Retrieval Based on a XGboost Machine Learning Aided Method: Performance and Validation. *Remote Sensing*. 11. 1655. 10.3390/rs11141655.