

RANDOM FOREST REGRESSION FOR THE ESTIMATION OF LEAF AREA INDEX OF OKRA CROP USING GROUND BASED BISTATIC SCATTEROMETER

Suraj A. Yadav^{1,*}, Rajendra Prasad¹, Ajeet Kumar Vishwakarma¹, Vijay Pratap Yadav¹

Department of Physics, Indian Institute of Technology (BHU), Varanasi – (yadavsuraja.rs.phy17, rprasad.app)@itbhu.ac.in,
(ajeetbhu87, victory327759)@gmail.com

Commission V, SS: Emerging Trends in Remote Sensing

KEYWORDS: Bistatic scatterometer system, Okra crop, LAI, RF

ABSTRACT:

The specular bistatic scattering mechanism of Okra's crop was analyzed using dual polarized ground based bistatic scatterometer system at X, C, and L bands in the specular direction with the azimuthal angle ($\theta = 0^\circ$). An outdoor Okra crop bed of area $10 \times 10 \text{ m}^2$ was specially prepared for the estimation of leaf area index (LAI) at HH and VV polarizations over the angular range of incidence angle 20° to 60° at steps of 10° . The regression analysis was done between bistatic specular scattering coefficients and crop biophysical parameter at X, C, and L bands for HH and VV polarization at different angle of incidence to determine the optimum parameters of bistatic scatterometer system. The linear regression analysis showed the high correlation at 40° angle of incidence for all bands and polarizations for the Okra crop. The computed scattering coefficients and measured LAI of Okra crop for the seven growth stages at 40° angle of incidence were interpolated into 61 data sets. The data sets were divided into input, validation and testing for the training and testing of the developed random forest regression (RF) model for the estimation of LAI for Okra crop. The estimated values of LAI of Okra crop, by the developed RF regression model, were found more closer to the observed values at X band for VV polarization with coefficient of determination ($R^2 = 0.928$) and low root mean square error (RMSE = $0.260 \text{ m}^2/\text{m}^2$) in comparison to C and L bands.

1. Introduction

Monitoring growth and health of vegetation is one of the key applications in remote sensing for agricultural planning. Okra is available almost throughout the year and widely spread over tropical and temperate climate region of the world. At the proper ripe period, the fruits are excellent source of calcium, potassium, vitamins, proteins and total minerals which are often deficient in diet of developing country. It has plentiful other economic uses also, as a result there is a strong need for an effective monitoring programme of okra crop to provide information about its growth and growing conditions. A traditional importance has been recognized to relate biophysical parameter to bistatic scattering coefficient for an operational monitoring of vegetation by means of remote sensing. The best correlation between bistatic scattering coefficient and biophysical parameter for particular angle is used to develop a model for the estimation of biophysical parameter of vegetation. The produced model for the estimation may be useful for extracting biophysical parameter of crop/vegetation at different frequencies and polarization configuration of scatterometer system. Although bistatic radar system is typically more complex and difficult to implement than monostatic, the potential to provide degree of stress in acrop, improvement in monitoring vegetation growth and ability of target detection motivates persistent studies. Various theoretical and experimental researchers have mainly

concentrated on the backscattering of microwave in soil and vegetation (Le Toan et al., 1984; Brisco and Brown, 1998). A number of authors illustrated the relationship of the backscattering coefficient with the Leaf Area Index, plant biomass, plant height, plant water content, and soil moisture content (Wooding et al., 1992; Daughtry et al., 1991; Ulaby et al., 1984; Dabrowska-Zielinska et al., 1994; Prasad, 2011). Microwave backscattering coefficient values depend on the incident angle, polarization of electromagnetic wave, soil moisture, plant dielectric and vegetation canopy structure hence the combination of microwave signatures at different frequencies and polarization useful for the estimation of soil moisture and vegetation health (Bouman and Uenk, 1992; Allen and Ulaby, 1984; Macelloni et al., 2001; Champion, 1996; Le Toan, 1982; Paloscia, 1998; Mo et al., 1984; Moran et al., 1998; Ulaby et al., 1986; Inoue et al., 2001; Attema and Ulaby, 1978). Very few researchers have been started theoretical modeling and experimental measurements on bistatic scattering responses of vegetation/soil at different microwave bands. Further, this model is simulated to find bistatic scattering responses for different incidence angle, azimuth angle, scattering angle, various frequency bands and vegetation parameters. For vegetation, a first-order microwave bistatic scattering model was developed for Wheat and Soybean by radiative transfer model at L band and C band based on the Michigan Microwave Canopy Scattering (MIMICS). The simulations of bistatic scattering responses by the developed MIMICS model reported that bistatic scattering response at C band was a good choice to retrieve the

* Corresponding author

vegetation parameters, whereas, the response at L band was preferred to retrieve the soil parameters (Zhang and Wu, 2016). A radiative transfer model was developed for microwave bistatic scattering from forest canopies and the simulation results shows the bistatic scattering mechanisms and the potential application of bistatic measurements (Pan Liang, 2005). McLaughlin et al., (2002) studied the full polarization bistatic scattering from forest hills at various incident angle were studied by (McLaughlin et al., 2002). Gupta et al., (2015) estimated crop growth parameter of rice crop at X band by using soft computing technique. Soft computing techniques like Artificial Neural Network(ANN), Support Vector Machine(SVM), Random Forest (RF), Genetic Algorithm(GA), and Fuzzy Logic do not need complex mathematics and large number of input parameter and easy to estimate biophysical parameters. Vishwakarma et al. (2018) estimated biophysical parameters of ladyfinger crop by using fuzzy inference technique at X- band. This paper describes a set of ground based bistatic scattering measurement of Okra crop conducted in specular directions with angular range 20° to 60° at the step of 10° . The measurements were conducted using X, L and C bands. Bistatic scatterometer system composed dual polarized transmitter and receiver. Bistatic scatterometer measured scattered power from vegetation at pq - polarization, where, q and p corresponds to polarized incident plane wave and polarized scattered wave. Polarized incident or scattered wave can be horizontal (H) and vertical (V). The optimization technique namely random forest model was used for the estimation of LAI. The performance of the developed model was evaluated by the statistical parameter such as coefficient of determination (R^2) and root mean square error (RMSE).

2. Methodology

Okra crop bed of an area $10\text{m} \times 10\text{m}$ was specially prepared for the bistatic scattering measurements besides the department of Physics, IIT (BHU), Varanasi, India. Average leaf shape, height and age of okra crop were found $\sim 15\text{-}25$ cm broad, $\sim 90\text{-}100$ cm and $\sim 90\text{-}95$ days, respectively. Average leaf shape and height were measured by taking five different plants from outside the region of interest. The spacing between each okra plant was 20cm (row) and 20cm (column). The bistatic scattering and the crop growth variable (LAI) measurements were done for seven growth stages of okra crop from the date for sowing at the interval of 10 days. The measurements were started from 20th days after sowing. LAI is a dimensionless quantity that characterizes plant canopies and it is mathematically expressed as

$$LAI = \frac{\text{leaf area}}{\text{ground area}} \left(\frac{\text{m}^2}{\text{m}^2} \right) \quad (1)$$

3. Procedure and performance indices of the developed model

Systems used throughout the experiments is shown in Table (1). The pair of moving pyramidal horn antenna were mounted on horizontal track. The ends of horizontal track is fixed on two stand of height 3m both situated at distance of 12m for bistatic scatterometer measurement at X, C and L bands. Pyramidal horn antenna is calibrated with sliding circular scale having capability of incident angular movement ranging from 0° to 90° . The bistatic scatterometer measurements were performed for incident angular range of 20° to 60° at the interval of 10° in specular direction. Each day of measurement require calibration of bistatic scattering response of vegetation with the bistatic scattering response of horizontal bed of aluminum to ensure system reliability. Hence, the horizontal bed of aluminum sheet is prepared beside the okra crop and experiments were performed with system having same specification.

3.1 Root Mean Square Error (RMSE)

The RMSE gives the measure of difference between the value predicted or estimated by model and the value observed and mathematically expressed as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Where,
 x_i and y_i are observed and estimated value respectively.

3.2 Correlation coefficient:

Linear regression analysis was done to measure a relationship between bistatic scattering coefficient and LAI for different angle of incidence, polarization at X, C and L bands which is presented in Table 2.

3.3 Coefficient of determination:

The coefficient of determination is the square of the correlation coefficient between the estimated and observed values.

| Parameters | Specifications | | |
|-----------------------------|--|----------------------|----------------------|
| | X-band(8.2~12.4 GHz) | C-band(4.9~7.05 GHz) | L-band(1.7~2.6 GHz) |
| Frequency bands | X-band(8.2~12.4 GHz) | C-band(4.9~7.05 GHz) | L-band(1.7~2.6 GHz) |
| Frequency used (GHz) | 10 | 6 | 2 |
| Band width | | | |
| E Plane (°) | 16.5 | 16.9 | 45.4 |
| H Plane (°) | 16.1 | 14.3 | 64.8 |
| RF generator (Agilent made) | E8257D, PSG high power signal generator, (10 MHz–20 GHz) | | |
| Power meter (Agilent made) | E4416A, EPM-P series power meter, 10 MHz–20 GHz | | |
| Power sensor (Agilent made) | Peak and average power sensor (E9327A, 50 MHz–18 GHz) | | |
| Polarization modes | Horizontal transmit–horizontal receive (HH) Vertical transmit–vertical receive (VV) | | |
| Antenna (Pasternack made) | Dual-polarized pyramidal horn | | |
| Calibration accuracy (dB) | 1 | | |
| Platform height (m) | 3 | | |
| Incident angle | 20° –60° | | |

Table 1. Specification of bistatic scatterometer system.

4. Random Forest (RF) model

Random forest is deep machine learning technique to predict the samples data under non-parametric approach. Generally, if each ϕ_h is a decision tree, then the ensemble is a random forest technique. RF model uses the samples of the training data to develop the multiple nodes of trees. These nodes are the set of conditions, which are applied from the root to the leaves of the tree. The RFR model averages the results from individual trees and generally predicts the better results (Breiman et al.,1996).Each tree was trained with an independent bootstrap sample selected randomly from the sample data. On average, two-third of the sample data were used for training the model and remaining one-third of data samples were not selected for growing the trees (Breiman et al.,2001). The samples not selected into a bootstrap sample are the out of the bag (OOB) samples. As these OOB points were used to validation of the model (Kumar et al., 2018). The RF regression model generally requires some input parameters used to decide the accuracy of the model predicted data. The parameters are number of nodes, number of trees and the number of variables (Barrett et al., 2014). RF model prediction can be expressed mathematically as

$$RF \text{ prediction} = \frac{1}{h} \sum_{h=1}^h \phi_h \quad (3)$$

Where, the index h runs over the individual trees and ϕ_h is tree response of h^{th} tree.

5.Results and Discussion

The linear regression analysis between experimentally calculated bistatic scattering response and crop growth variable (LAI) at all angle of incidence for HH polarization and VV polarization is shown in Table 2. The analysis shows higher correlation between bistatic scattering coefficient and leaf area index for VV polarization in comparison to HH polarization at 40°angle of incidence excluding at L band as shown in Table 2. The highest correlation was found for X band at VV polarization and hence used for estimation of crop growth variable using random forest regression model. Bistatic scattering coefficient of Okra crop at 40° angle of incidence for X band at VV polarization were interpolated in 61 data sets at the interval of 1 day (20-80 days after sowing) which was further divided into 40 and 21 data sets for training and testing of RF model, respectively.

| Incidence angle in degree (°) | Correlation coefficient between bistatic scattering coefficient and leaf area index(LAI) at X, C and L bands | | | | | |
|-------------------------------|--|-----------------|----------------|-----------------|-----------------|-----------------|
| | X-band | | C-band | | L-band | |
| | HH | VV | HH | VV | HH | VV |
| 20° | -0.65372 | -0.79306 | -0.71308 | -0.79906 | -0.73756 | -0.75567 |
| 30° | -0.67334 | -0.8185 | -0.6666 | -0.86862 | -0.8331 | -0.80777 |
| 40° | -0.8844 | -0.95628 | -0.8021 | -0.87682 | -0.84586 | -0.83347 |
| 50° | -0.84882 | -0.78226 | -0.77843 | -0.852 | -0.66769 | -0.77305 |
| 60° | -0.72603 | -0.82844 | -0.76844 | -0.79658 | -0.74401 | -0.81768 |

Table 2. Correlation coefficient between bistatic scattering coefficient and leaf area index (LAI).

Figure 1 shows temporal variation of LAI measured at the interval of 10 days from the date of sowing. LAI was found to increase with the age of the okra crop till day 70 and then after LAI decreased slightly. Figures 3, 4 and 5 show the temporal variation of specular bistatic scattering response with respect to days after sowing of okra crop for X, C and L band at HH and VV polarization, respectively. Scatterer component of vegetation like crown (represented as branch and leaves), trunk (represented as stem) were small at early growth stage (Ulaby et al. 1990). Hence, attenuation caused by these components was low and the main contribution in specular bistatic scattering coefficient was mainly due to ground surface i.e. soil beneath vegetation. At maturity stage (20-70 days), the attenuation caused by increased crop growth variable caused to decrease in specular bistatic scattering coefficient. After the maturity stage (70-80 days), the crop growth variable started decreasing led to increase specular bistatic scattering coefficient slightly as shown in figures 3, 4 and 5.

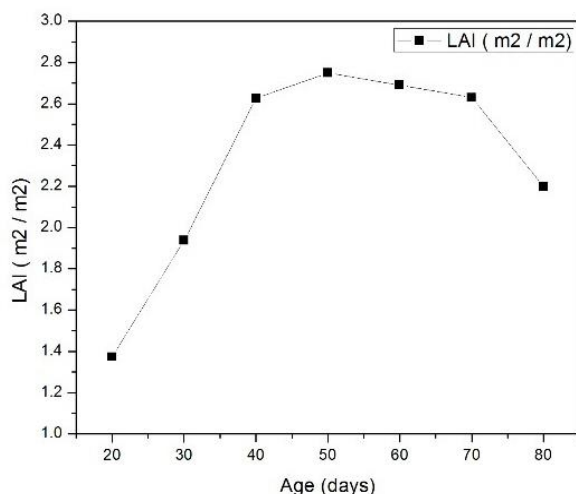


Figure 1: Temporal variation of LAI after sowing.

5.1 Retrieval of Okra crop Leaf Area Index using RF regression model

Determining optimal number of tree and regularizing stopping parameters were used to find accurate estimation of crop growth variable using RF regression model (Shataee et al. 2012). Determining optimal number of tree is based on the tree number that produces stable error. 150 initial trees were used to produce graph between average squared error and number of trees for X-band at VV polarization and 40° angle of incidence as shown in Figure 2. In addition, the stopping conditions include minimum 5 child node and maximum 100 child node to stop growing tree in 10 cycles to calculate mean error. The graph shows that improvement in accuracy was not found significant after 40 trees where as 60-80trees showed lowest error and stable response and hence good to estimate optimum parameter. Sub-sample proportion (i.e. 50 %) of training samples was used in bagging bootstrap sampling. Unused test sample was used to evaluate the validity of RF

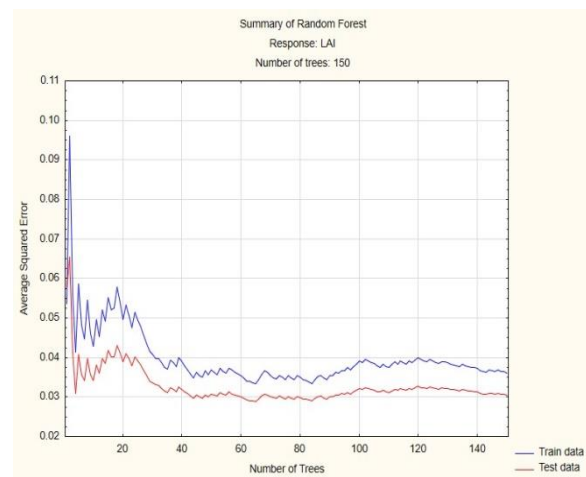
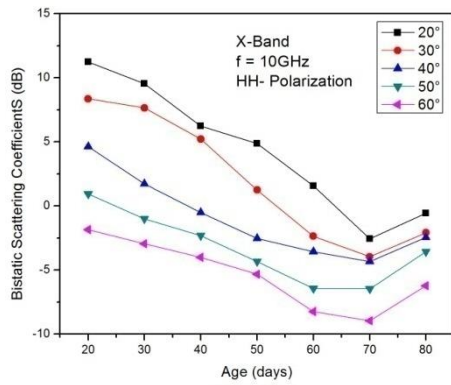
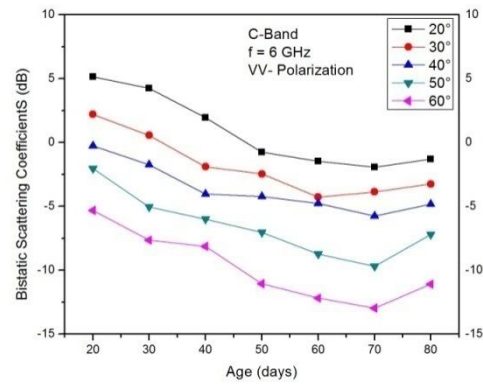


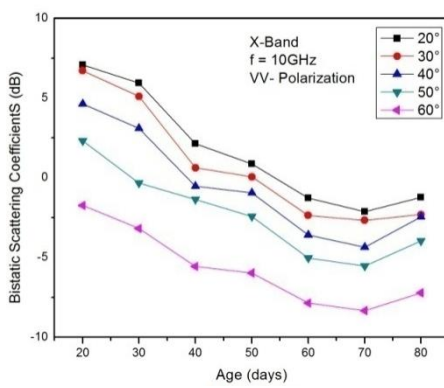
Figure 2: Average square error against number of trees in training and testing data for LAI retrieval of Okra crop for X band at VV polarization.



3.(a)

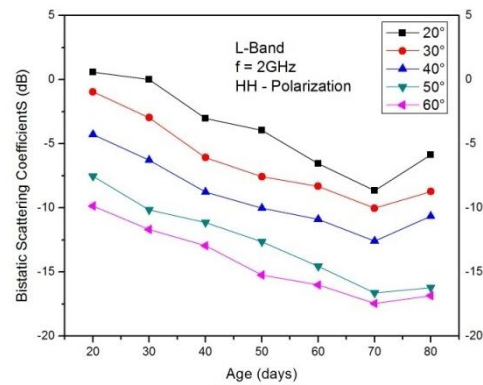


4.(b)



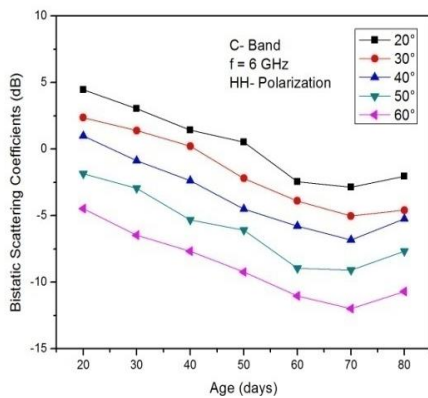
3.(b)

Figure 4: Temporal variation of bistatic scattering response of crop okra for C-band at (a) HH and (b) VV polarization, respectively.

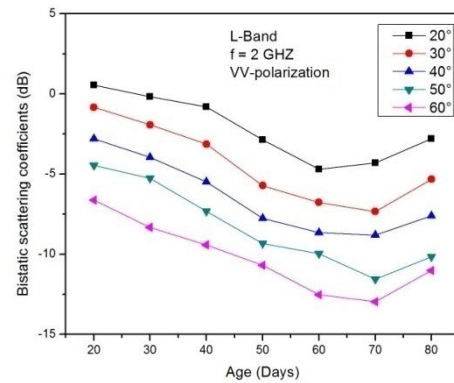


5.(a)

Figure 3: Temporal variation of bistatic scattering response of crop okra for X-band at (a) HH and (b) VV polarization, respectively.



4.(a)



5.(b)

Figure 5: Temporal variation of bistatic scattering response of crop okra for L-band at (a) HH and (b) VV polarization respectively.

regression model by calculating performance indices (RMSE) on the unused test sample. Figure 6 shows scatter plot with 1:1 line between observed LAI and estimated LAI of Okra crop. The values of observed LAI and estimated LAI by random forest regression model were found closer with high coefficient of determination ($R^2 = 0.928$) and low root mean square error (RMSE= $0.26 \text{ m}^2/\text{m}^2$). The performance of RF regression model for estimation of LAI was found well for X band at VV polarization.

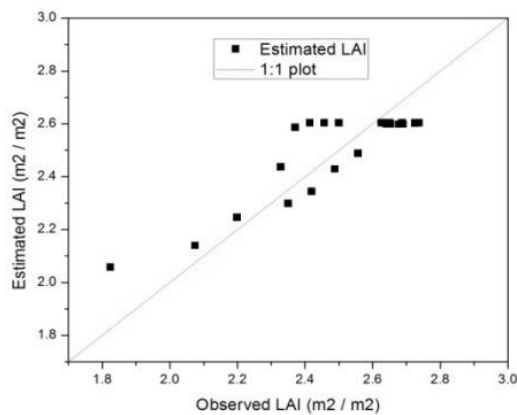


Figure 6: Scatter plot with 1:1 line between observed and estimated crop growth variables by RF model and.

6. Conclusions

In the present study, X band was found highly sensitive to Okra crop since bistatic scattering response was higher for X band at HH polarization and VV polarizations. However, bistatic scattering coefficient at HH polarization was found more sensitive to VV polarization at X and C bands. However, the L band at VV polarization was found more sensitive to Okra crop growth variables was found higher for L band at VV polarization in comparison to HH polarization. The correlation coefficient was found maximum for X band at VV polarization and 40° angle of incidence. The performance indices and estimation of leaf area index by developed RF regression model was found good and close to observed values, respectively. Hence, we can use RF regression model as a powerful tool to estimate crop growth variables.

References

Allen, C.T., Ulaby, F.T. Modelling the polarization dependence of the attenuation in vegetation canopies. Proceedings of the IGARSS'84 Symposium, 119–124, 1984.

Attema, E.P., Ulaby, F.T. Vegetation modeled as a water cloud. *Radio Sci.* 13 (2), 357–364, 1978.

Bouman, B.A.M., Uenk, D. Crop classification possibilities with radar in ERS-1 and JERS-1 configuration. *Remote Sens. Environ.* 8 (40), 1–13, 1992.

Breiman L. 1996. Bagging predictors. *Mach Learn.* 24:123–140.

Breiman L. 2001. Random forests. *Mach Learn.* 45:5–32.

Brisco, B., Brown, R.J. Agricultural applications with radar. Principles and applications of imaging radar – Manual of Remote Sensing, vol. 2, Third Edition, F.M Henderson and A.J Lewis Ed., pp. 381–406, 1998.

Champion, I. Simple modelling of radar backscattering coefficient over a bare soil: variation with incidence angle, frequency and polarization. *Int. J. Remote Sens.* 17 (4), 783–800, 1996.

Champion, I., Prevot, L., Guyot, G. Generalized semi-empirical modelling of wheat radar response. *Int. J. Remote Sens.* 21 (9), 1945–1951, 2000.

Dabrowska-Zielinska, K., Gruszczynska, M., Janowska, M., Stankiewicz, K., Bochenek, Z. Use of ERS-1 SAR data for soil moisture assessment. Proceedings of the First Workshop on ERS-1 Pilot Projects, pp. 79–84, 1994.

Daughtry, C.S.T., Ranson, K.J., Biehl, L.L. C-band backscattering from corn canopies. *Int. J. Remote Sens.* 12 (5), 1097–1109, 1991.

Gupta, D. K., Kumar, P., Mishra, V. N., Prasad, R., Dikshit, P. K. S., Dwivedi, S. B., Ohri, A., Singh, R. S., Srivastava, V. (2015). Bistatic measurements for the estimation of rice crop variables using artificial neural network. *Advances in Space Research*, 55, 1613–1623.

Inoue, Y., Dabrowska-Zielinska, K., Kurosu, T., Maeno, H., Uratsuka, S., Koza, T. Analysis of multi-frequency and full – polarization backscatter signatures over paddy field – results of daily measurements of Ka, Ku, X, C and L bands for entire growing season. Proceedings of the 8th Symposium Int. Physical Measurements and signatures in Remote Sensing, pp. 427–437, 2001.

Kumar P, Prasad R, Choudhary A, Gupta D, Mishra V, Vishwakarma A, Singh A, Srivastava PJGI. 2018. Comprehensive evaluation of soil moisture retrieval models under different crop cover types using C-band synthetic aperture radar data. 1–20.

Le Toan, T. Active microwave signatures of soils and crops: significant results of three years of experiments. Proceedings International Geoscience and Remote Sensing Symposium, (New York: I.E.E.E.), pp. 25–32, 1982.

Le Toan, T., Lopes, A., Huet, M. On the relationships between radar backscattering Coefficient and Vegetation Canopy Characteristics. Proceedings of the International Geoscience and Remote Sensing Symposium, pp. 155–160, 1984.

Liang, P., M. Moghaddam, L. E. Pierce, and R. M. Lucas, "Radar backscattering backscattering model for multilayer mixed-species forests," *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, No. 11, 2612–2626, 2005.

- Macelloni, G., Paloscia, S., Pampaloni, P., Marliani, F., & Gai, M. (2001). The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 873–884
- McLaughlin, D. J., Y. Wu, W. G. Stevens, X. Zhang, M. J. Sowa, and B. Weijers, "Fully polarimetric bistatic radar scattering behavior of forested hills," *IEEE Transactions on Antennas and Propagation*, Vol. 50, No. 2, 101–110, 2002.
- Mo, T., Schmugge, T., Jackson, T. Calculation of radar backscattering coefficient off vegetation covered soils. *Remote Sens. Environ.* 15, 119–133, 1984.
- Moran, M.S., Vidal, A., Troufleau, D., Inoue, Y., Michell, T.A. Ku and C-band SAR for discriminating agricultural crop and soil conditions. *IEEE Trans. Geosci. Remote Sens.* 36 (1), 265–272, 1998.
- Paloscia, S. An empirical approach to estimating leaf area index from multifrequency SAR data *Int. J. Remote Sens.* 19 (2), 359–364, 1998.
- R. Prasad, (2011) Estimation of crop variables of kidney-bean using ground based scatterometer data at 9.89 GHz. *International Journal of Remote Sensing*, 32(1): 31–48.
- Shataee S, Kalbi S, Fallah A, Pelz D. 2012. Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *Int J Remote Sens.* 33:6254–6280.
- Ulaby, F.T., Allen, C.T., Eger, G. Relating the microwave backscattering coefficient to leaf area index. *Remote Sens. Environ.* 14, 113–133, 1984.
- Ulaby, F.T., Moore, R.K., Fung, A.K. *Microwave remote sensing, active and passive, vol. III: From Theory to Applications*, M.A. Dedham Ed. Artech House, 1986.
- Ulaby, F.T., Sarabandi, K., McDonald, K., Whitt, M., Dobson, M.C. Michigan microwave canopy scattering model. *Int. J. Remote Sens.* 11 (7), 1223–1253, 1990.
- Vishwakarma, A.K., Prasad, R, Gupta DK, Kumar P, Mishra, 2018. Ground Based Bistatic Scatterometer Measurement for the Estimation of Growth Variables of Ladyfinger Crop at X-Band. *J Indian Soc Remote* 46:973-980.
- Wooding, M.G., Griffiths, G.H., Evans, R., Bird, P., Kenward, D., Keyte, G.E. Temporal monitoring of soil moisture using ERS-1 SAR data. *Proceedings of the First ERS-1 Symposium*, pp. 641-647, 1992.
- Zhang, Y.-Y., & Wu, Z.-S. (2016). Bistatic scattering characteristics of wheat and soybean by radiative transfer model in L band and C band. *Progress In Electromagnetics Research B*, 67, 121–136.