2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS) | 979-8-3503-4542-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/MIGARS57353.2023.10064591

Enhancing Plant Area Index Retrieval Using Gaussian Process Regression from Dual-Polarimetric SAR Data

Swarnendu Sekhar Ghosh
Microwave Remote Sensing Lab (MRSLab)
Indian Institute of Technology, Bombay
Mumbai, India
swarnendughosh@iitb.ac.in

Narayanarao Bhogapurapu
Microwave Remote Sensing Lab (MRSLab)
Indian Institute of Technology, Bombay
Mumbai, India
bnarayanarao@iitb.ac.in

Avik Bhattacharya
Microwave Remote Sensing Lab (MRSLab)
Indian Institute of Technology, Bombay
Mumbai, India
avikb@csre.iitb.ac.in

Saeid Homayouni
Centre Eau Terre Environnement
Institut national de la recherche scientifique (INRS)
Quebec, Canada
saeid.homayouni@inrs.ca

***Abstract*—In this paper, a Gaussian Process Regression (GPR) model is implemented to retrieve the Plant Area Index (PAI) of wheat and canola. Backscatter information from Sentinel-1 dual-pol GRD SAR data and in-situ measurements collected during the Soil Moisture Active Passive Validation Experiment 2016 (SMAPVEX16-MB) Manitoba campaign were used to calibrate and validate the proposed GPR model. A recently proposed pseudo scattering entropy, H_c derived from dual-pol GRD SAR data has been used along with backscatter information to investigate the improvement in retrieval accuracy. Including the pseudo entropy parameter in the feature, space showed an improvement of $4.28\,\%$ and $3.66\,\%$ in the correlation coefficient (ρ) for wheat and canola respectively. Similarly, a decrease in nRMSE by $4\,\%$ for wheat and $4.76\,\%$ for canola was observed during PAI retrieval.**

***Index Terms*—Gaussian Process Regression (GPR), Plant Area Index (PAI), Sentinel-1, Pseudo Entropy Parameter (H_c), SMAPVEX16-MB**

I. Introduction

Plant Area Index (PAI) is an essential biophysical parameter linked to crop productivity and is a crucial variable in crop-growth models. Knowledge of variation in PAI during the entire crop phenology is critical for crop growth monitoring and yield forecasting. Utilizing Earth Observation (EO) data to estimate biophysical parameters has gained significant importance over the decades because of wide-area coverage, high temporal revisit, and spatial and spectral diversity. Synthetic Aperture Radar (SAR) data has drawn appreciable attention toward various agricultural applications because of its all-weather imaging capability and sensitivity toward geometric and dielectric properties of crops [1].

Authors acknowledge the GEO-Planetary Computer Earth Observation Cloud Credits Program supported the computation with Sentinel-1 on the cloud platform through the project: "Azure4GEO-Deep learning based crop characterization with synergistic use of SAR and optical data on cloud computing platform" and formed the testbed for processing pipelines.

Previous studies have demonstrated how several biophysical parameters can be modeled from SAR backscatter [2]. Semi-empirical estimation of bio-physical parameters has received considerable attention [3]. Although well-suited for the operational monitoring of crops, the inversion of semi-empirical models is limited because of their ill-posed nature [4]. In this regard, several machine-learning algorithms have been implemented to retrieve biophysical parameters from SAR data. The continuous nature of biophysical parameters motivates utilizing regression-based algorithms.

Kernel-based methods have shown promising results in retrieving different biophysical parameters [5] [6]. These methods utilize a kernel to quantify the similarity between input features. One such kernel-based method is Gaussian Process Regression (GPR) which works on a Bayesian framework. In addition, unlike other parametric and non-parametric methods, GPR provides an uncertainty estimate of the target variable, which helps analyze the error in the model's predictions. In earlier studies, GPR has shown impressive results in estimating biophysical parameters from optical [7] and full-polarimetric SAR data [8].

Compared to full-pol mode, dual-pol modes have several advantages, which include larger swath widths and lower data volume, but at the expense of polarimetric information. Sentinel-1 SAR sensor acquires data in dual-polarization VV-VH for land observation. Earlier studies have shown that polarimetric parameters can be related to the physical properties of crop canopy and thus helps in monitoring crop phenology [9]. However, the polarimetric parameters reported in these studies were not critically applicable in the case of dual-pol GRD SAR data. In a recent study [10], three polarimetric descriptors derived from Sentinel-1 GRD SAR data were proposed. The study revealed the sensitivity of the pseudo scattering entropy parameter (H_c) across different phenological stages of various crops across different test sites [11].

In the present study, we implement a GPR model with two different feature spaces to estimate the biophysical parameter PAI of Wheat and Canola. The first feature space includes VH and VV backscatter coefficients from dual-pol SENTINEL-1 GRD SAR data. The second feature space includes the pseudo scattering entropy parameter (H_c) along with VH and VV backscatter coefficients. The efficacy of GPR models has been characterized by various statistical measures such as the normalized Root Mean Square Error (nRMSE), Mean Absolute Error (MAE), and the Pearson correlation coefficient (ρ).

II. Study area and data-set

The test site is located in the Red-river watershed of southern Manitoba, Canada. The Soil Moisture Active Passive Validation Experiment (SMAPVEX16-MB) campaign site covers a vast area of $26\,\mathrm{km} \times 48\,\mathrm{km}$. The test site has an annual crop mix of spring wheat, soybean, canola, corn, oats, beans, and other crop types. The current study focuses primarily on spring wheat and canola which covers $24.12\,\%$ and $18.24\,\%$ of the area respectively. Over 50 agricultural fields were selected for sampling which had a nominal size of $800\,\mathrm{m} \times 800\,\mathrm{m}$. During the campaign, 3 sampling points were selected in each field over two transects for vegetation sampling. A detailed description of vegetation ad soil sampling techniques can be found in the SMAPVEX16-MB field report [12].

This study uses in-situ measurements acquired over wheat and canola fields to train and validate the GPR model. Further, five scenes of dual polarimetric (VV+VH) C-band Sentinel-1A data in Interferometric Wide Swath (IW) mode have been selected and processed to derive backscatter intensities ($\sigma^\circ_{\mathrm{VV}}$ and $\sigma^\circ_{\mathrm{VH}}$) and the pseudo entropy [13]. The observation periods considered are 13$^{\text{th}}$June, 30$^{\text{th}}$June, 7$^{\text{th}}$July, 19$^{\text{th}}$July and 24$^{\text{th}}$July. For both crop types, the data has been divided by randomly splitting them into training ($70\,\%$) and validation ($30\,\%$) data sets.

III. Methodology

A. Gaussian Processes

A Gaussian process is a generalization of multi-variate Gaussian distribution and is defined by its mean function and covariance (kernel) function [14]. The aim is to find the function f from an infinite set of unknown latent functions, a finite number of which are jointly Gaussian distributed. The posterior distribution provides the probable functions that can retrieve the unknown target variable $\boldsymbol{y}$ from a set of input features, $\boldsymbol{x}$. An additive noise model $\boldsymbol{y} = f(\boldsymbol{x}) + \varepsilon$, is assumed, where the inherent noise is assumed to follow a standard normal distribution with 0 mean and variance $\sigma_n, \varepsilon \backsim \mathcal{N}(0, \sigma_n^2)$. Thus the prior joint distribution of the functional values $\boldsymbol{y}$ at known points and the unknown functional values denoted by $\boldsymbol{f_*}$ is given by,

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{pmatrix} \backsim \mathcal{N}\left(0, \begin{pmatrix} \boldsymbol{K} + \sigma_n^2 \boldsymbol{I} & \boldsymbol{K}_* \\ \boldsymbol{K}_*^T & \boldsymbol{K}_{**} \end{pmatrix} \right) \tag{1}$$

where the terms $\boldsymbol{K}$, $\boldsymbol{K}_*$ and $\boldsymbol{K}_{**}$ represent the covariance matrix between the observed (known) features, between observed and test features and between test features respectively. The noise variance gets added as a diagonal matrix $\sigma_n^2 \boldsymbol{I}$ along with the covariance matrix $\boldsymbol{K}$. The role of the kernel functions is to quantify the underlying non-linear relationships among the features. In this study, a non-linear Radial Basis Function (RBF) kernel along with a zero-mean Gaussian additive noise has been utilized, as shown below,

$$k(x_i, x_j) = \sigma_f^2 \exp\left(\frac{- \parallel (x_i - x_j) \parallel^2}{2\, l^2} \right) + \sigma_n^2 \delta_{ij} \tag{2}$$

here, δ_{ij} represents a Kronecker delta function. The hyper-parameters of the RBF kernel σ_f known as RBF variance, and l known as the length scale, are optimized by maximizing the log marginal likelihood of the Gaussian process. The log marginal likelihood is expressed as,

$$\log p(\boldsymbol{y} \mid \boldsymbol{\theta}, \sigma_n) = \log \mathcal{N}(\, \boldsymbol{y} \mid 0, \boldsymbol{K} + \sigma_n^2\, \boldsymbol{I}) \tag{3}$$

where $\boldsymbol{\theta}$ is the set of all the model hyper-parameters. A gradient-based approach is used while optimizing the log marginal likelihood. Finally, the posterior predictive distribution of the Gaussian process gives us the mean (point) and the variance (uncertainty) estimate of the target variables, which in our case is the biophysical parameter. The posterior predictive distribution of the Gaussian process is given by,

$$\boldsymbol{f}_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_* \backsim \mathcal{N}(\overline{\boldsymbol{f}_*}, \boldsymbol{\Sigma}_*) \tag{4}$$

The mean ($\overline{\boldsymbol{f}_*}$) and covariance ($\boldsymbol{\Sigma}_*$) of the predictive distribution are expressed as,

$$\overline{\boldsymbol{f}_*} = \boldsymbol{K}_*^T [\boldsymbol{K} + \sigma_n^2\, \boldsymbol{I}]^{-1} \boldsymbol{y} \tag{5}$$

$$\boldsymbol{\Sigma}_* = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T (\boldsymbol{K} + \sigma_n^2\, \boldsymbol{I})^{-1} \boldsymbol{K}_* \tag{6}$$

IV. RESULTS AND DISCUSSION

A. Sensitivity of H_c with Plant Area Index (PAI)

The correlation between the pseudo entropy parameter and the ground-measured PAI over the entire phenology of wheat and canola has been analyzed.

1) Wheat: According to in-situ measurements, on 13 June, the wheat crops were in their tillering stage, while some were advanced from their tillering stage towards stem elongation and booting stage. During this period, the scattering entropy increases, possibly because of a change in crop morphology in the vertical direction with an increase in the size of the main stem and side tillers [10]. It is evident from Figure 1 that the spread of a majority of points in the medium to high entropy region. As the crop progresses toward its flowering stage, the complex structure of the wheat canopy increases randomness in scattering. This increase in scattering randomness is probably due to the presence of flowers on the upper canopy layer of the crop. During this period, high entropy values are noticeable. The crop reaches its early dough and maturity stages towards late July. The randomly oriented wheat heads significantly contribute to the total scattering during this period. Most of the points cluster around the high entropy region during this phenological period of the wheat crop. An overall linear correlation of 0.37 between H_c and in-situ measured PAI was observed.

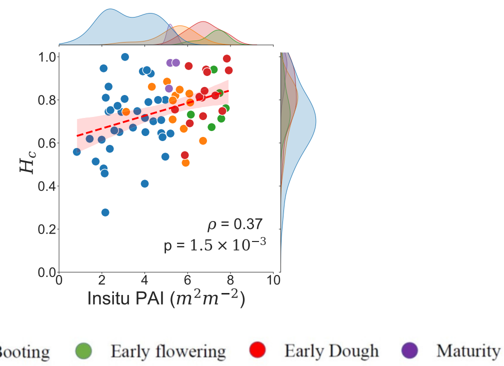


Fig. 1: Correlation analysis between pseudo scattering entropy parameter (H_c) and in-situ measured PAI for Wheat.

2) Canola: During the leaf development stage of canola, there is an increase in PAI. An increase in leaf density and the formation of branches increase scattering randomness. At this stage, attenuation of vertically polarized waves occurs. As evident in Figure 2, most points during this stage are around the medium entropy region. As the crop progresses towards its flowering stage, the formation of branches increases scattering entropy. We can observe that majority of the points have shifted towards the higher entropy region. This increase can be attributed to the complex canopy geometry, which occurs as buds develop into flowers and the branches start to grow. As the crop progresses toward pod development and maturity, there is an increase in randomness in the scattering mechanism. The majority of points during this phenological stage of canola seem to be toward higher entropy region. An overall linear correlation of 0.73 was observed between the pseudo entropy parameter and in-situ measured PAI in the case of canola.

B. PAI retrieval results

The Gaussian process regression model discussed in Section III-A has been utilized to retrieve the PAI of two different crop types, namely Wheat and Canola. The retrieval results are presented in Figures 3 and 4. The red dashed line represents the best fit line with the shaded region showing the 95 % confidence interval.

1) Wheat: During the campaign period, the in-situ measured PAI for wheat varied between $0.83\,\mathrm{m^2\,m^{-2}}$ to $7.92\,\mathrm{m^2\,m^{-2}}$.

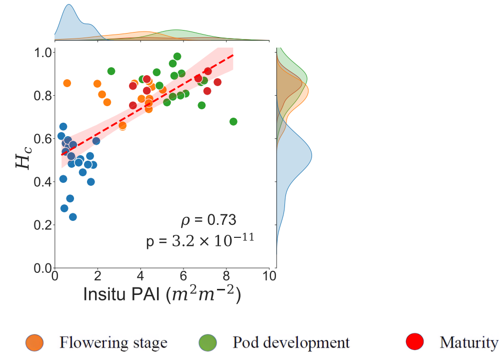
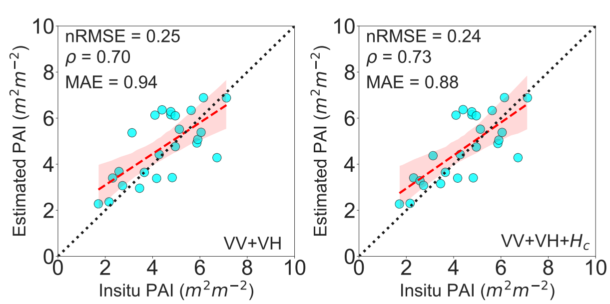


Fig. 2: Correlation analysis between pseudo scattering entropy parameter (H_c) and in-situ measured PAI for Canola.

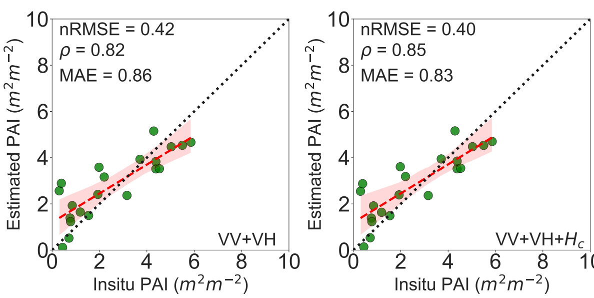
An overestimation for PAI ($< 2.0\,\mathrm{m^2\,m^{-2}}$) can be seen, which is most likely due to the major contribution from bare soil in the initial phenological stage of the crop. As the crop progresses towards the booting stage, there is an increase in the scattering entropy. Subsequently, the crop reaches its head stage, changing its canopy structure. During this period, the major scattering contribution occurs from the upper layer of the canopy. Subsequently, during the flowering stage, a significant increase in pseudo entropy can be seen from Figure 3. On the other hand, an underestimation of PAI ($> 6.0\,\mathrm{m^2\,m^{-2}}$) can be seen in both Figure 3a and Figure 3b. During this period, the plant reaches its dough stage. As the crop advances towards its higher phenological stages with an increase in biomass, a saturation in C-band occurs. This saturation is common among crops with erectophile geometry like wheat. The depolarization of the incident radar signal is mostly captured by the cross-pol component (VH). It is evident from Figure 3b with the inclusion of the entropy descriptor H_c in the feature space, the GPR model has been able to better estimate PAI for wheat. A 4 % and 6.38 % decrease in error estimates nRMSE and MAE can be seen respectively. In addition, there is an increase of 4.28 % in the correlation between in-situ and estimated PAI of wheat.



(a) (b)

Fig. 3: Comparison of estimated and in-situ PAI for wheat utilizing (a) VV+VH and (b) VV+VH+H_c. The red dashed line represents the best fit line with the shaded region showing the 95 % confidence interval.

2) Canola: Canola is a broadleaf plant with a unique canopy structure. The in-situ measured PAI varied between $0.3\,\mathrm{m}^2\,\mathrm{m}^{-2}$ to $8.33\,\mathrm{m}^2\,\mathrm{m}^{-2}$. In case of canola an overestimation of PAI ($< 1.2\,\mathrm{m}^2\,\mathrm{m}^{-2}$) can be seen in Figures 4a and 4b. During this period, the crop is in its leaf-development stage. So the major backscatter contribution that reaches the radar comes from the bare soil due to sparse vegetation cover. As the crop progresses towards a higher phenological stage, there is an increase in PAI with increased leaf density. With the formation of branches, there is an increase in scattering randomness. During the inflorescence emergence, flower buds develop, and the density of leaves increases significantly. Due to this increase in density, the scattering entropy increased during this period. Following inflorescence emergence, as the crop reaches its flowering stage, with the development of a complex canopy structure, there is a further increase in entropy. Subsequently, as the crop reaches its maturity, there is a significant decrease in the overall canopy moisture content. Therefore a significant underestimation can be seen during maturity stage for PAI ($> 5.5\,\mathrm{m}^2\,\mathrm{m}^{-2}$). Similar to wheat, it is also evident in the case of canola that utilizing pseudo entropy descriptor H_c along with the backscatter intensities as a predictor has shown better results. There is a $4.7\,\%$ and $3.4\,\%$ decrease in error estimates nRMSE and MAE, respectively. Further, the linear correlation between in-situ measured and estimated PAI has shown an increase of $3.65\,\%$ when we utilize pseudo entropy descriptor H_c as a predictor along with the dual-polarimetric backscatter coefficients.



(a) (b)

Fig. 4: Comparison of estimated and in-situ PAI for canola utilizing (a) VV+VH and (b) VV+VH+H_c.The red dashed line represents the best fit line with the shaded region showing the $95\,\%$ confidence interval.

V. CONCLUSION

In this study, retrieval of the biophysical parameter PAI for wheat, and canola, has been proposed by utilizing a Gaussian Process Regression (GPR) model. In this regard, C-band Sentinel-1 dual-pol GRD SAR data and in-situ measurements obtained during the SMAPVEX16-MB campaign have been used to calibrate and validate the GPR model. The model calibration and validation results show that high correlation and lower error estimates are observed when the pseudo-scattering entropy parameter, H_c was utilized as a feature in addition to the dual-pol (VV+VH) backscatter coefficients. A plausible explanation for this outcome can be the pseudo entropy parameter H_c, being a better representer of scattering randomness helps enhance the biophysical parameter estimation. Thus, utilizing the pseudo entropy parameter derived from dual-pol GRD SAR data and the backscatter intensities helps in utilizing the available polarimetric information and consequently enhances the performance of the GPR model. The proposed approach can be of particular interest to existing and upcoming dual-pol SAR missions for global scale estimation of LAI using GRD data only.

REFERENCES

[1] Susan C Steele-Dunne et al., "Radar remote sensing of agricultural canopies: A review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2249–2273, 2017.

[2] Grant Wiseman et al., "Radarsat-2 polarimetric sar response to crop biomass for agricultural production monitoring," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 11, pp. 4461–4471, 2014.

[3] Dipankar Mandal et al., "Joint estimation of Plant Area Index (PAI) and wet biomass in wheat and soybean from C-band polarimetric SAR data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 79, pp. 24–34, 2019.

[4] Emilie Bériaux et al., "Maize leaf area index retrieval from synthetic quad pol SAR time series using the water cloud model," *Remote Sensing*, vol. 7, no. 12, pp. 16204–16225, 2015.

[5] Surya S Durbha, Roger L King, and Nicolas H Younan, "Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer," *Remote sensing of environment*, vol. 107, no. 1-2, pp. 348–361, 2007.

[6] Gustau Camps-Valls et al., "A survey on Gaussian processes for earth-observation data analysis: A comprehensive investigation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 58–78, 2016.

[7] Jochem Verrelst et al., "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3," *Remote Sensing of Environment*, vol. 118, pp. 127–139, 2012.

[8] Swarnendu Sekhar Ghosh et al., "Gaussian Process Regression Model for Crop Biophysical Parameter Retrieval from Multi-Polarized C-Band SAR Data," *Remote Sensing*, vol. 14, no. 4, pp. 934, 2022.

[9] Subhadip Dey et al., "Novel clustering schemes for full and compact polarimetric SAR data: An application for rice phenology characterization," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 135–151, 2020.

[10] Narayanarao Bhogapurapu et al., "Dual-polarimetric descriptors from Sentinel-1 GRD SAR data for crop growth assessment," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 178, pp. 20–35, 2021.

[11] Narayanarao Bhogapurapu et al., "Crop Growth Assessment Using Sentinel-1 GRD SAR Descriptors," in *2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*. IEEE, 2021, pp. 545–548.

[12] Heather McNairn et al., "Experimental plan SMAP validation experiment 2016 in Manitoba, Canada (SMAPVEX16-MB)," 2016.

[13] Narayanarao Bhogapurapu et al., "PolSAR tools: A QGIS plugin for generating SAR descriptors," *Journal of Open Source Software*, vol. 6, no. 60, pp. 2970, 2021.

[14] Christopher KI Williams and Carl Edward Rasmussen, *Gaussian processes for machine learning*, vol. 2, MIT press Cambridge, MA, 2006.