Chapter 3

Synergy of dual – polarimetric radar vegetation descriptor and Gaussian processes regression algorithm for estimation of leaf area index

3.1 Introduction

Biophysical variables are vital for many applications, especially in precision agriculture (Huete, 2012; Wenbo et al., 2006). Among various biophysical variables, the LAI was extensively studied by optical and microwave sensors mounted on Earth-observing satellites (Ulaby et al., 1984; Verrelst et al., 2012; Yadav et al., 2022c). It is a dimensionless variable calculated using the leaf area ratio per unit surface area (Bréda, 2008; Fang and Liang, 2008). The Global Climate Observing System recognized the LAI as one of its 50 essential climate variables. Traditional field observation (in-situ) of LAI or other biophysical parameters at a landscape scale can be spatially, temporally limited, and

expensive. Therefore, remote sensing techniques can solve this difficulty because they can map large landscapes at a fine spatio-temporal scale (Ling et al., 2016; Zhong and Zhang, 2012). After the launch of better resolution satellites, such as Sentinel-1 (S1) SAR and Sentinel-2 (S2), it is now possible to study crops with high spatial and temporal resolutions (Bauer-Marschallinger et al., 2021; Yadav et al., 2021).

S2, being an optical mission, cannot assess plant characteristics directly. Because the reflected radiation must first be converted into surface reflectance at the sensor that can easily be interpreted. In order to accomplish this goal, several required preprocessing steps were carried out. These steps include calibration, radiometric, geometric, and atmospheric adjustment. The atmospheric correction converts the Top Of Atmosphere (TOA) radiance to Bottom Of Atmosphere (BOA) reflectance, after which the retrieval of the biophysical variables was carried out (Estévez et al., 2020; Sola et al., 2018). Although atmospheric correction is a critical step, it is not a simple procedure. It entails the inversion of the atmospheric Radiative Transfer Model (RTM) through the lookup table approach. Along with the uncertainty associated with the inversion of the RTM, other sequential processes involved in the atmospheric correction procedure may introduce intrinsic errors in the BOA reflectance. To circumvent this issue, it was first proposed and then theoretically confirmed that biophysical variables can be obtained directly from TOA radiance, bypassing the atmospheric correction procedure (Bayat et al., 2020; Mousivand et al., 2015; Shi et al., 2017; Verrelst et al., 2019). Although the S2 was categorized in the TOA and BOA datasets, they belong to the optical region of the electromagnetic spectrum where weather conditions may cause some errors in the reflectance.

The limitations associated with the S2 optical datasets led to the adoption of microwave-based SAR sensors, which have a higher penetrating ability and are less affected by clouds and bad weather conditions. Many studies have evaluated the performance of S1 dual-polarized (VV-VH) SAR data in retrieving crop biophysical variables (Barbouchi et al.,

2016; Stankevich et al., 2017). Mandal(Mandal et al., 2020) proposed the novel Dual Polarimetric Radar Vegetation Index (DpRVI) derived from the S1 SAR C-band data. DpRVI is a bound quantity that ranges between 0 and 1 and helps assess the plant's growth.

The biophysical variables are derived from the remote sensing observations in mainly two ways: (1) the statistical approach, which works by establishing the relation between the biophysical variables and remote sensing data through linear (for example, vegetation indices) or nonlinear (for example, Artificial Neural Network (ANN)) regression methods; and (2) the physical retrieval approach, which works by the inversion of RTM against the remotely sensed data. The statistical approach delivers the estimation with high accuracy and lower computational cost. Nonlinear Machine Learning Algorithms (MLA) have shown significantly good performance in retrieving LAI (Verrelst et al., 2012; Yadav et al., 2020). For instance, ANN was based on the biophysical processor tool implemented in the Sentinel Application Platform (SNAP) to process the S2 images. However, in recent studies, the Gaussian Processes Regression algorithm (GPR), based on the Bayesian theory principle, has performed better than other well-established MLAs (Upreti et al., 2019). GPR, a kernel-based probabilistic model, uses a kernel function to establish a nonlinear relationship by studying the similarity between the input datasets (Liu et al., 2018; Schulz et al., 2018).

In the purview of the above, this study utilizes the GPR algorithm to develop the LAI retrieval model by using the freely available S2 TOA reflectance and S1 SAR backscattering datasets, respectively. In addition, LAI was also retrieved from the biophysical processor (based on the ANN method) in SNAP using S2 BOA reflectance for comparative analysis. The workflow of this paper will follow the sequence (1) evaluate the performance of the S1 SAR and S2 optical imagery, especially the performance of TOA reflectance with the atmospherically corrected BOA reflectance; and (2) compare the performance of the GPR with the ANN algorithm and validate the estimation performance against the ground

measurements. Finally, the spatio-temporal LAI maps from the constructed GPR and ANN LAI retrieval models were generated for the wheat crop.

3.2 Study area and data used

3.2.1 Study area

The study area is located in the Varanasi district of the Indian state of Uttar Pradesh, and Figure 3.1 provides a general overview of the region. The center longitude and latitude of

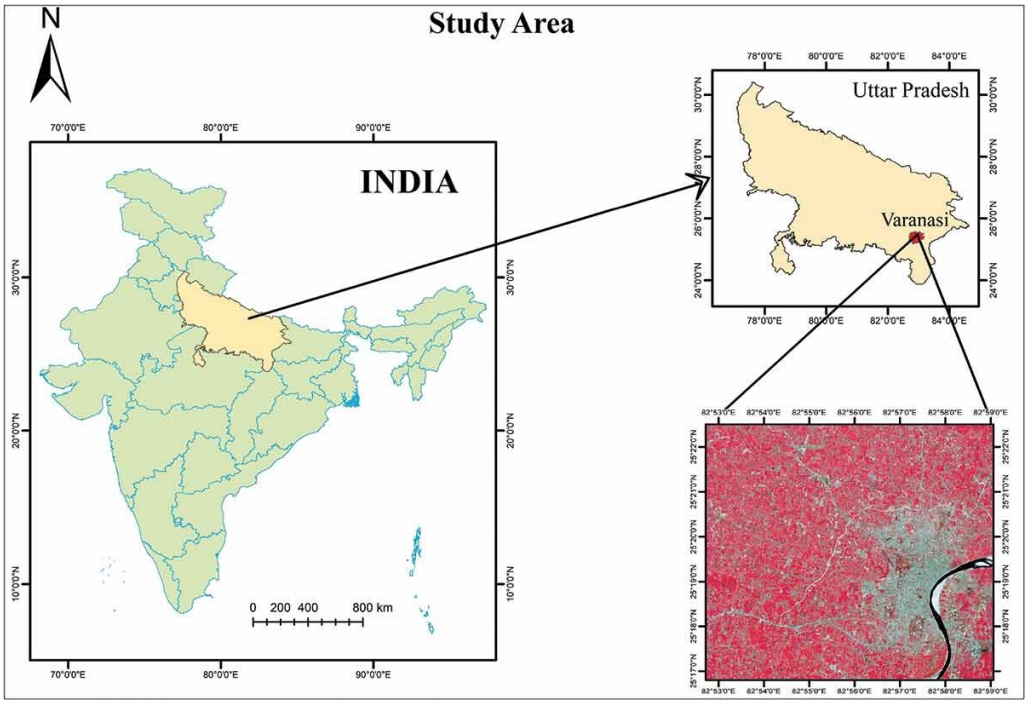

Figure 3.1 The location of the district of Varanasi and the study areas in the false-color composite picture created by the S2 image.

the region are 82°58′56.92″ E and 25°15′55.27″ N, respectively. This site was chosen due to its geographical position, having a total district area of about 1,535 km^2. This location is part of the Indo-Gangetic Plain, which consists of large alluvial deposits, making this land the most fertile and suitable for agriculture. The region's annual rainfall is approximately 1110 mm, with almost 80% of it coming from the monsoon season, typically from July

to September. The region's temperature varies from 9°C to 42°C over the year. The soil capability and reasonable temperature variation make the Gangetic plains capable enough to grow any crop from tropical or temperate regions. Wheat, rice, sugarcane, and maize are the dominant agricultural crops in this region, accounting for more than 90% of cropland.

3.2.2 In-situ measurements

LAI ground measurements were performed in the Varanasi region between January 2020 and March 2020. At the study sites, the plots were chosen at random, and it was made sure that they were inside the image strips. The LAI data is acquired using the Li-COR plant canopy analyzer (Li-Cor Inc.). An average of three measurements under the canopy is taken to get the final LAI data. It was ensured that Li-COR was working correctly, and data acquisition was made under a clear sky to avoid contamination in LAI data. The statistics of observed LAI are presented in Table 3.1.

Property	**Details**
Date of in–situ data Sampling	3-Jan-2020, 25-Jan-2020, 07-Feb-2020 23-Feb-2020, 05-March-2020
Biophysical Paramete	LAI
Instrument	Li-COR plant canopy analyser
Max LAI	5.8 (m^2/m^2)
Average LAI	2.8 (m^2/m^2)
Min LAI	1.1 (m^2/m^2)

Table 3.1 The detailed specification of in-situ measurement for wheat crop.

3.2.3 Satellite data

In this investigation, S1 SAR and S2 Multispectral Imager (MSI) images are downloaded from (https://scihub.copernicus.eu/dhus/home) for five different dates between January and March 2020. We used the S1 Single Look Complex (SLC) product in Interferometric Wide

(IW) swath mode, its primary operational mode, to acquire land data. Using dual-polarized (VV+VH) SAR data, the polarimetric indicator DpRVI was derived. The Copernicus S2, the polar-orbiting satellite, consists of a constellation of two satellites in the same sun-synchronous orbit, phased at 180° to each other. Both S2 satellites mounted MSI with 13 spectral bands ranging from visible, near-infrared to short-wave infrared. Depending upon the wavelength of the spectral bands of MSI, which have center values from 400 to 2400 nm, they have resolutions of 10 m, 20 m, and 60 m. Three of the 13 spectral bands provided images in the red-edge (B5, B6, B7), most important for vegetation-related studies (Delegido et al. 2011). The preprocessing of both S1 and S2 images is performed in the open-source software provided by the European Space Agency (ESA) Sentinels Application Platform (SNAP) version 8.0 toolbox (http://step.esa.int/main/download/). The image tiles for both satellites were downloaded from the Copernicus open-access platform. The detailed characteristics of both satellite images are presented in Table 2.1 and 2.2 for Optical and SAR, respectively.

3.3 Methodology

3.3.1 LAI estimation based on radiative transfer model within SNAP biophysical processor

S2 TOA reflectance image is converted into the BOA reflectance image by means of the sen2cor (based on lookup tables generated through libRadtran RTM) procedure (Louis et al., 2019). The TOA reflectance represents the reflectance of the atmosphere as well as of the Earth's surface. In contrast, in BOA reflectance, an attempt to minimize the reflectance of the atmosphere was made so that the final contribution in BOA reflectance comes from the Earth's surface only. The BOA reflectance image at 20 m spatial resolution is used to derive the LAI from the built-in SNAP Biophysical processor. The Biophysical

processor using PROSAIL and ANN applied on the various bands (B3, B4, B5, B6, B7, B8A, B11, and B12) along with viewing zenith, solar zenith, and relative azimuth angles of BOA reflectance image to produce LAI (Weiss et al.; Xie et al., 2021). The produced LAI was compared to LAI generated through the GPR model and validated against ground measurements.

3.3.2 Computation of dual-polarized radar vegetation index (DpRVI)

S1 SAR SLC data is obtained from the Copernicus Open Access Hub and processed in the SNAP toolbox. This toolbox performed the radiometric calibration to convert the digital image values into the calibrated backscattering coefficients. This dual polarized complex output data product is used to generate a 2×2 covariance matrix. Then the speckle noise is reduced by using the redefined Lee speckle filter that averages the images while keeping the edges. The formulation of the covariance matrix is done using the scattering matrix. For the dual-polarization case, the scattering matrix can be used to get the target vector Equation 3.1, which can be written as

$$\mathbf{k_T} = \begin{bmatrix} S_{VV} \\ S_{HH} \end{bmatrix} \tag{3.1}$$

From the target vector, the covariance matrix is obtained using the relation $\langle C \rangle = \langle k_T * k_T^{\dagger} \rangle$

$$\langle \mathbf{C} \rangle = \begin{bmatrix} \langle |S_{VV}|^2 \rangle & \langle S_{VV} S_{VH}^* \rangle \\ \langle S_{VH} S_{VV}^* \rangle & \langle |S_{VH}|^2 \rangle \end{bmatrix} \tag{3.2}$$

The degree of polarization m is a ratio of all components of the polarized electromagnetic wave and total incident energy. In the study, Chang (Chang et al., 2018) reported the direct relationship between the m and vegetation canopy. Barakat (Barakat, 1977), in his research,

obtained the expression of m for the $N \times N$ covariance matrix. In this study, the covariance matrix is reduced to 2×2, so the expression of m took the form as

$$m = \sqrt{1 - \frac{4|C|}{(Trace(C))^2}} \tag{3.3}$$

where C is the 2×2 covariance matrix and $|C|$ is the determinant of the covariance matrix. Further, the Eigendecomposition of the C matrix is done to extract the two non-zero eigenvalues (λ_1, λ_2) by which parameter β is defined, and this eigenvalue calculates the dominant scattering mechanism. By using the m and β, the dual-polarized radar vegetation index was defined as (Mandal et al., 2020)

$$DpRVI = 1 - m * \beta \qquad 0 \leq DpRVI \leq 1 \tag{3.4}$$

where $\beta = \frac{\lambda_1}{Trace(C)}$. The values of DpRVI lie between 0 and 1, where 0 indicates no vegetation presence and 1 indicates dense vegetation. It is helpful in data interpretation and crop assessment.

3.4 Gaussian processes regression (GPR)

GPR was implemented in this study to retrieve the LAI. GPR has shown its potential as a powerful and efficient regression tool in remote sensing communities (Gewali et al., 2019; Pasolli et al., 2008). GPR is a probabilistic Bayesian methodology that associates input variables (bands) with output parameters

$$y = f(x) = \sum_{k=1}^{n} c_i K(x_i, x) \tag{3.5}$$

where x_i is the input spectra for the training purpose, and x is the test data. Assessing the similarity between the input test spectrum x and training spectrum x_i (where $i =$

$1,2,3,4,\ldots n$), K, the kernel or covariance function is calculated, ci is weightage assigned to each kernel function. In regression studies, kernel functions play a vital role in extracting vegetation parameters (Ghosh et al., 2022; Montesinos-López et al., 2021). In this study, a scaled Gaussian kernel function was used as:

$$K(x_i,x_j) = vexp\left(-\sum_{b=1}^{B}\frac{\left(x_i^{(b)}-x_j^{(b)}\right)^2}{2\sigma_b^2}\right)+\delta_{ij}.\sigma_n^2 \tag{3.6}$$

where v is the factor scaling the kernel, B is the total number of bands, δ_{ij} is a knocker delta function, σ_n is the standard deviation associated with noise, σ_b is standard deviation associated with spectral bands. The parameters v, σ_n, and σ_b are collectively called hyper-parameters and denoted as (v, σ_n, σ_b). In the Gaussian process output function, y is assumed to be a combination of actual value and Gaussian noise defined as $f(x)+\varepsilon$. This Gaussian noise ε is assumed to be additive with zero mean and σ_n variance. This is called prior distribution. The predictive posterior distribution (predictive mean and variance) is calculated by conditioning the data and test observation. The hyper-parameters (v, σ_n, σ_b) were automatically optimized by maximizing the negative log marginal likelihood. For this study, GPR is implemented with nine Bands of S2. The analysis is carried out by the GPML package available freely for MATLAB (http://www.gaussianprocess.org/gpml)

3.4.1 Performance metrics

To assess the performance of the estimated and measured LAI, the statistical parameters used are as follows

- Coefficient of determination (R^2) as

$$R^2 = \left(\frac{\sum_{i=1}^{N} \left(Y_{obs}^i - \overline{Y_{obs}^i} \right) \left(Y_{est}^i - \overline{Y_{est}^i} \right)}{\sqrt{\sum_{i=1}^{N} \left(Y_{obs}^i - \overline{Y_{obs}^i} \right)^2} \sqrt{\sum_{i=1}^{N} \left(Y_{est}^i - \overline{Y_{est}^i} \right)^2}} \right) \tag{3.7}$$

- Bias is defined as

$$Bias = \sum_{i=1}^{N} \frac{Y_{obs}^i - Y_{est}^i}{N} \tag{3.8}$$

- Nash – Sutcliffe model efficiency coefficient (NSE) is defined as

$$NSE = 1 - \frac{\sum \left(Y_{obs}^i - Y_{est}^i \right)^2}{\sum \left(Y_{obs}^i - \overline{Y_{obs}^i} \right)^2} \tag{3.9}$$

- Quantification of retrieval accuracy is done by root mean square error (RMSE), defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_{obs}^i - Y_{est}^i \right)^2}{N}} \tag{3.10}$$

where Y_{obs} and Y_{est} are observed and estimated variables, N is the total number of datasets. $\overline{Y_{obs}}$ is average of Y_{obs}^i defined as $\overline{Y_{obs}} = \frac{1}{m} \sum_{i=1}^{m} Y_i$.

3.5 Results and discussion

3.5.1 Validation of LAI retrieved from SNAP biophysical processor

The SNAP-derived (using the Biophysical processor) LAI from the S2 BOA reflectance level at 20 m spatial resolution is validated using the in-situ LAI data. The accuracy statis-

tics (R^2 = 0.696, RMSE = 0.760 m^2m^{-2}, and NSE = 0.578) of the estimated LAI against the measured LAI are shown in Figure 3.2. It is evident from the scatter plot (figure 3.2) that the SNAP biophysical processor has overestimated the results (Bochenek et al., 2017). This peculiar behavior can be attributed to the prior assumptions made into the retrieval algorithms of SNAP Biophysical processor, that is, ANN and PROSAIL RTM, which may not have been calibrated with the local conditions and led to the overestimation of the retrieved parameter (Kganyago et al., 2020). Moreover, the extent of inaccuracies in SNAP-

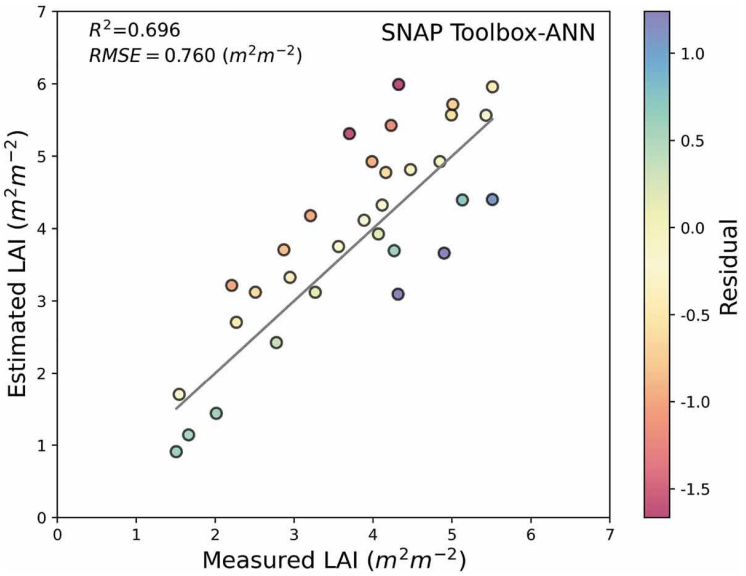

Figure 3.2 Validation of SNAP-derived LAI against ground measured LAI.

derived LAI shown for BOA reflectance data can be related to the residual uncertainty arising from the atmospheric correction procedure using the sen2cor algorithm. This is because of the dependence of atmospheric correction techniques on regions, environments, and spectral bands (Doxani et al., 2018; Martins et al., 2017). In the upcoming section, it has been shown how an alternative machine learning-based retrieval model has produced a better result in the estimation of LAI.

3.5.2 Determination of DpRVI using S1 SAR images

C_2 and m are used to compute the DpRVI as discussed in section 3.2 for the acquired images of S1 dual polarized. The spatio-temporal changes of DpRVI for the Varanasi region are illustrated in Figure 3.3. The temporal analysis of the crop using DpRVI helps us to assess the wheat's different phenological stages during its growth period. In the early days of the crop, the vegetation is relatively less; therefore, the value of DpRVI is found to be less, and as the crop density increased, the DpRVI values are found to be increased (Mandal et al., 2020).

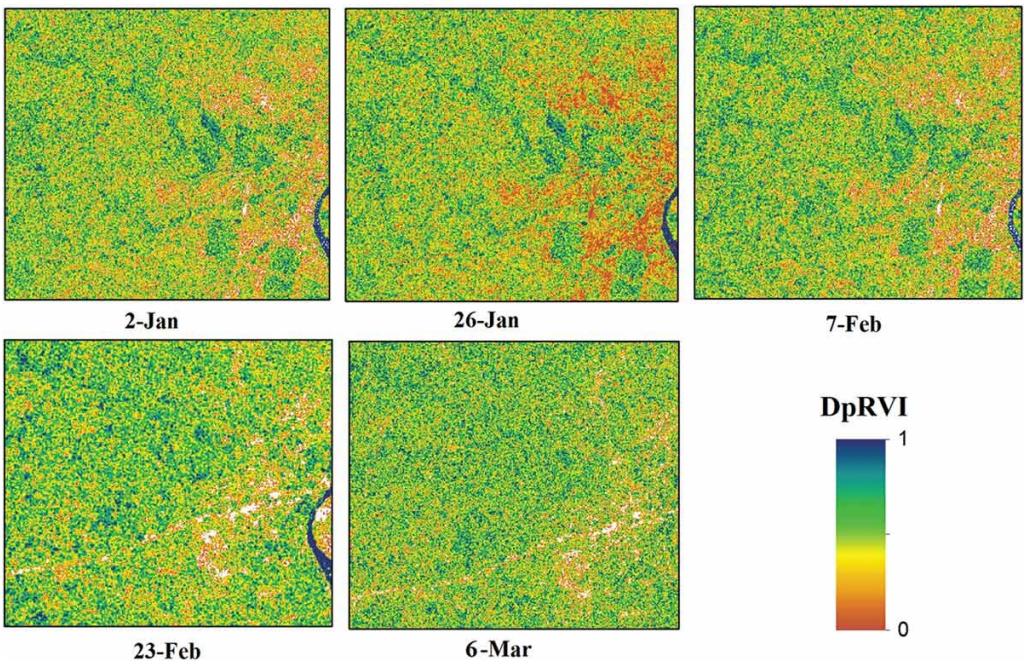

Figure 3.3 Spatio-temporal maps of DpRVI simulated from the dual-polarised S1 SAR datasets for VV-VH polarisation mode.

3.5.3 LAI estimation using trained GPR-LAI model and validation

To assess the capability of GPR in estimating LAI, the GPR model is trained using the lookup table generated from S1-derived DpRVI and the lookup table developed from the multispectral bands of S2 TOA image at 20 m spatial resolution. The trained GPR

models are applied to respective S2 TOA and DpRVI images to retrieve LAI and generate corresponding LAI maps. The retrieved LAI is validated using the measured LAI for the Varanasi region.

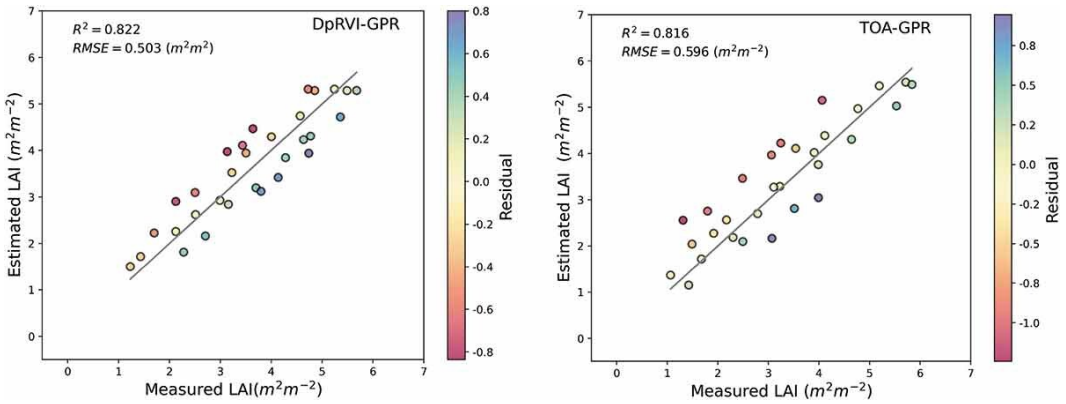

Figure 3.4 Validation of GPR-derived LAI from DpRVI and S2 TOA images against in-situ data.

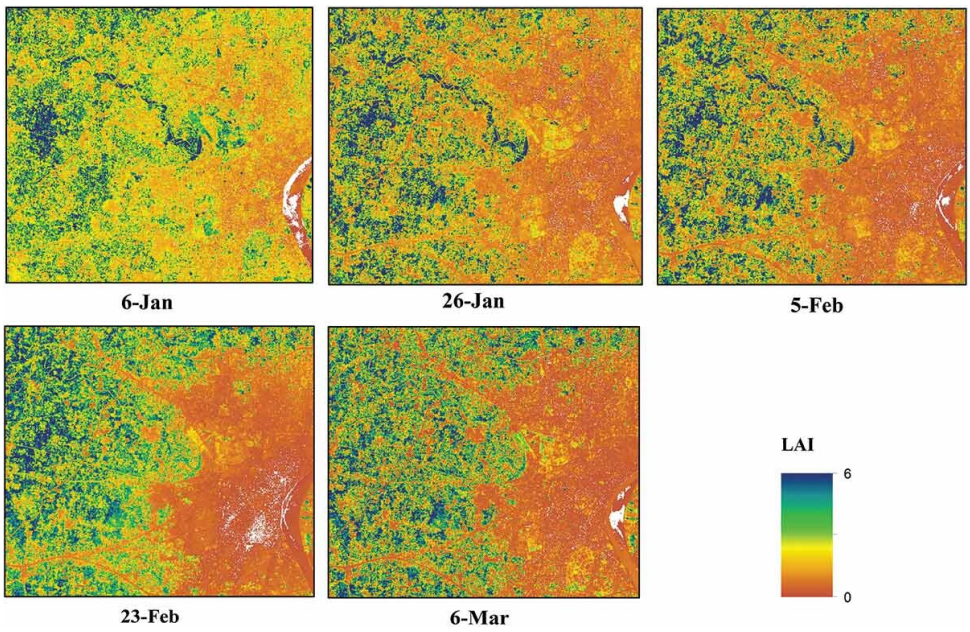

Figure 3.5 The spatio-temporal maps of LAI from S2 TOA spectral data using trained TOA-GPR. model.

Figure 3.4 shows the scattered plot of measured vs. estimated values of LAI. These GPR-based LAI models have shown better performance than the SNAP-derived LAI shown

in Section 4.1. For TOA, GPR achieved better accuracy (R^2 = 0.816, NSE = 0.803) than the SNAP-derived LAI for BOA (R^2 = 0.696, NSE = 0.578). Finally, for DpRVI, GPR has an estimated marginally greater accuracy than the other LAI retrieval models with an R^2 = 0.822 and NSE = 0.831. The error associated with the retrieval is also lower in the DpRVI-trained LAI retrieval model (DpRVI-GPR) than in the other two (RMSE for SNAP-derived LAI retrieval ias found as 0.760 m^2 m^{-2}, whereas the TOA trained LAI retrieval model (TOA-GPR) provided a value of 0.596 m^2 m^{-2}). The performance of the

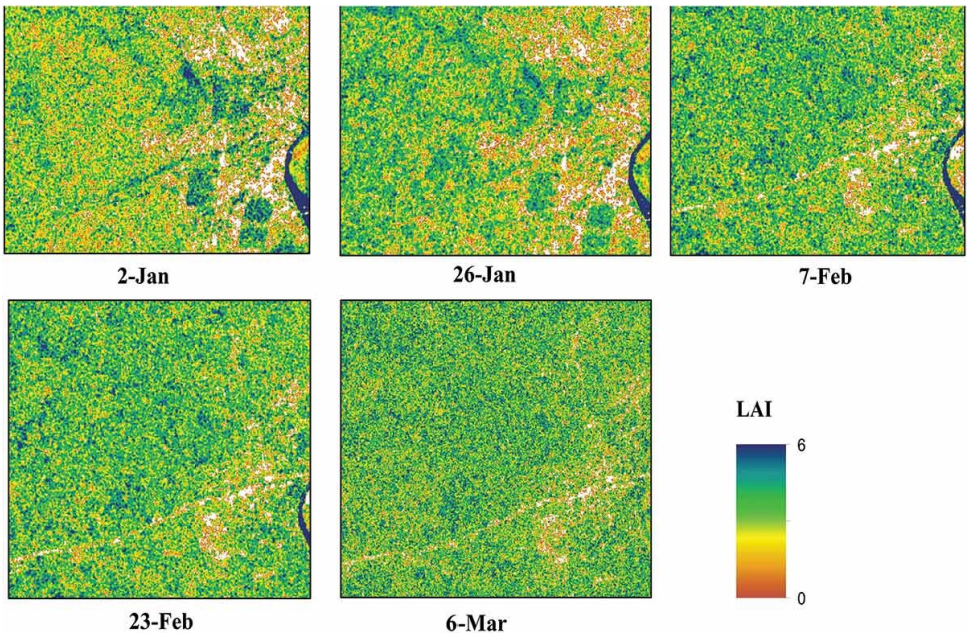

Figure 3.6 LAI's spatio-temporal maps were obtained from DpRVI datasets using the trained DpRVI-GPR model.

SNAP-derived model in the retrieval of LAI is found to be lowered than the DpRVI-GPR model and TOA-GPR because of the error that the atmospheric correction procedure produces during the application of the sen2cor processing tool implemented in the SNAP processing toolbox (Warren et al., 2019). The TOA-GPR and DpRVI-GPR LAI models are used in order to construct the spatiotemporal maps of LAI, and the results are shown in Figures 3.5 and 3.6, respectively. The maps can distinguish the wheat crop phenological

stages well. The temporal variation in the maps demonstrated that the average value of LAI increased from the early stage of vegetation from 2nd January to its peak on 6th March.

3.5.4 Sensitivity analysis of the GPR-LAI model

Taylor plot (Taylor, 2001) shows the graphical representation for assessing the model performance in an easier way. This study shows a brief performance analysis of all three models in terms of normalized standard deviation, root mean square deviation (RMSD), and correlation coefficient. The normalized standard deviation is represented on the y-axis, while the RMSD is presented on the x-axis. The correlation coefficient is drawn on the arc of the diagram, whereas RMSD was defined as the concentric circle with the center reference value. As the radius of the concentric circle increases, a higher error in the developed model can be reported. Figure 3.7 summarizes the performance of LAI

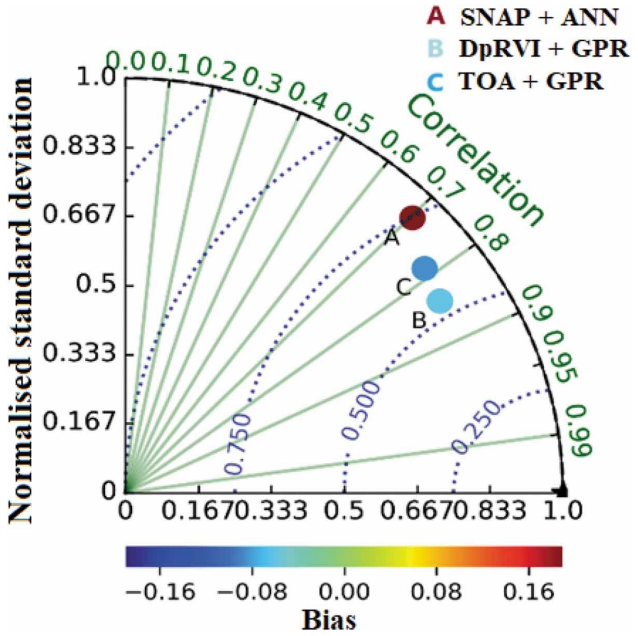

Figure 3.7 Comparative analysis of estimated LAI using Taylor plot.

estimation from different optical and active sensors satellite data using three different LAI retrieval models. Therefore, from the results, the following conclusions can be drawn: The

DpRVI-GPR model generated through the synergy of DpRVI and GPR indicated higher correlation, low bias, and comparatively less standard deviation (B) for the retrieval of LAI among the other two LAI estimation models. The S2 TOA + GPR (C) model has shown a weaker correlation, bias, and standard deviation than the DpRVI-GPR model, but its accuracy is better than the SNAP-based ANN (A) model, which indicates lower performance with the observed data. Therefore, this SAR and GPR algorithm synergy could approximate the estimated LAI values in the vegetative crop fields more accurately than the other existing machine-learning regression models.

3.6 Conclusion

This study developed the LAI retrieval model for wheat crops using the S2 optical TOA and S1 SAR data images. The Bayesian-based GPR retrieval algorithm is trained by using the S2 TOA bands and S1-derived DpRVI images. The generated DpRVI-GPR model and TOA-GPR model are used to estimate the LAI. For comparison purposes, LAI retrieval from the ANN-based SNAP biophysical processor is utilized for obtaining the S2 BOA radiance data. Estimated LAI from DpRVI-GPR, TOA-GPR, and SNAP-based models is validated against the ground-measured LAI. The SNAP-derived LAI correlates well (R^2 = 0.696, NSE = 0.578) with the ground data. However, the obtained result is slightly overestimated with a relatively higher error (RMSE = 0.760 m^2m^{-2}) compared to the field data, which could be due to the residual errors in the BOA data because of the atmospheric correction procedure. These possible errors may cause lower LAI retrieval accuracy through the SNAP-based model. The effect of the sen2Cor procedure on individual bands of S2 is yet to be seen because atmospheric correction procedure performance varies differently for different bands of S2. In contrast, the TOA-GPR model produced a relatively better result with better correlation (R^2 = 0.816, NSE = 0.803) and lower error than the SNAP-derived LAI. The overall conclusion drawn from this study

is that the S1-based DpRVI-GPR model produced the most accurate LAI (R^2 = 0.822, RMSE = 0.503 m^2m^{-2}, NSE = 0.831) in comparison to the other two models. At the same time, the GPR retrieval model from S2 is significantly better than the SNAP-derived LAI. This study aimed to develop a retrieval model to retrieve biophysical parameters using a new Gaussian processes regression model. Optical data can get contaminated because of atmospheric uncertainties. This study highlighted that an alternative machine learning approach could be opted to retrieve biophysical parameters using the SAR datasets.