Chapter 6

Fusion of optical and SAR data using three approaches for the estimation of LAI with modified Integral Equation model

6.1 Introduction

Remote sensing is a technology that uses sensors deployed on satellites or airplanes to acquire information about the Earth's surface. On a worldwide scale, the approach has been used to examine land usage, land cover, and environmental changes. There are several forms of remote sensing data accessible, such as SAR and optical sensors (Singh et al., 2022; Veloso et al., 2017). SAR data offers information about the structural qualities of vegetation and soil moisture content, while optical data provides information about the surface's spectral properties and vegetation cover (Mirsoleimani et al., 2019; Villarroya-Carpio et al., 2022; Vreugdenhil et al., 2018).

Several researchers have used Sentinel-1 SAR and Sentinel-2 optical datasets to determine vegetation (LAI, VWC, plant height) and surface attributes (soil moisture, dielectric constant, surface roughness) using suitable models (Gherboudj et al., 2011; Singh et al., 2023; Yadav et al., 2022b). These datasets face considerable hurdles. In inclement weather, optical sensors cannot see the surface, and studies have shown that TOA and BOA data include mistakes when processing biophysical parameters (Kganyago et al., 2020; Singh et al., 2022). SAR is unaffected by weather and day/night circumstances, but biophysical and surface property calculations are complicated. This requires SAR scattering models with different polarizations (VV, VH). WCM has successfully linked biophysical and surface parameters to total backscattering coefficients (Attema and Ulaby, 1978). The simplified semi-empirical WCM can readily incorporate vegetation and surface scattering backscattering components. This letter focuses on the surface backscattering model, specifically the physical Integral Equation Model(IEM) (Baghdadi and Zribi, 2006; Fung and Chen, 2004). The IEM stands out due to its high applicability and effectiveness. Notably, this letter highlights the novel approach of the IEM at VH polarization, which adds to its significance.

While WCM connects vegetation and surface backscattering, finding the crucial indicator of the land surface dynamics is essential. LAI is a fundamental parameter used to quantify the foliage in a given area, providing valuable insights into vegetation density and productivity. In this research, we estimate LAI from optical and SAR datasets. Two widely recognized RTM are employed to achieve this: the PROSAIL model for optical data and the WCM for SAR data (Mandal et al., 2019; Rivera et al., 2013; Yadav et al., 2022a). The LAI must be accurately estimated to understand vegetation dynamics better. Although the accuracy and temporal resolution of optical and SAR sensors are limited when used individually, combining the data from the two sensors may get beyond these constraints and provide a more precise and thorough estimate of LAI. In this study, we

explore a number of methods for combining optical and SAR data, including deep learning, Principal Component Analysis (PCA), and non-linear regression algorithm. Combining the complementary information from the two sensors, we may gain a more profound knowledge of vegetation dynamics, which is essential for efficient ecosystem management and environmental monitoring.

6.2 Study area and datasets

6.2.1 Study area

This study assesses the biophysical variable (LAI) using a proposed methodology in the Varanasi district of India, located at a mean altitude of 81m above sea level and covering 192 km^2. The soil moisture levels fluctuate throughout the sampling location because of the region's wet subtropical environment and seasonal temperature shifts. The location of the region is part of the Indo-Gangetic plain, with a latitude of $25^0 17' 51"N$ and a longitude of $82^0 56' 36"E$, known for its fertile land and suitable climate for agriculture. The study area has been divided into two distinct locations, namely Location 1 and Location 2, as illustrated in figure 6.1. Location 1 comprises two separate sites: Site 1 in red and Site 2 in yellow.

6.2.2 Satellite data

This research is carried out using Sentinel-1 SAR and Sentinel-2 L2A satellite images. Both sets of satellite data are obtained from the Copernicus open access hub at the dates close to the ground sampling dates of 2020 and 2023. Both Sentinel-1 and Sentinel-2 satellite data are pre-processed using the European Space Agency's (ESA) Sentinels Application Platform (SNAP) software toolkit.

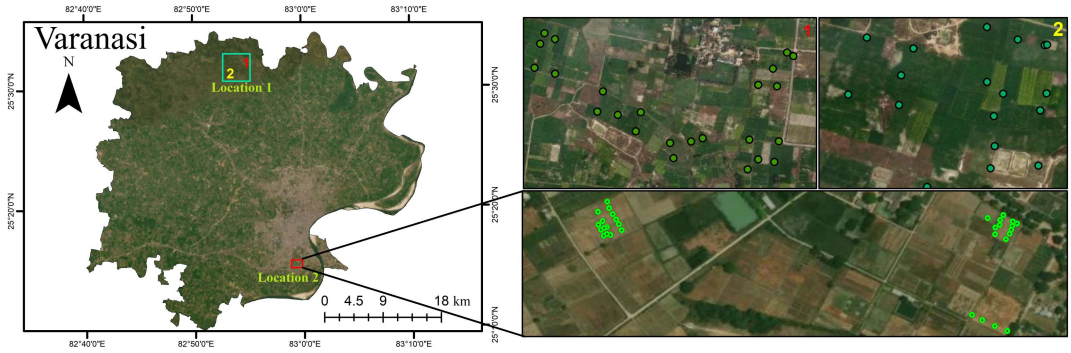

Figure 6.1 Study area showcasing Location 1 (comprising site 1 in red and site 2 in yellow) and Location 2 with sampling points, providing a comprehensive view of the entire study region.

6.2.3 In-situ measurement

The LAI measurements are taken in sampling areas using the LAI-2200 C plant canopy analyzer (LI-COR, Inc.). Soil moisture (m_v) is measured at a depth of an average of 5.0 cm in each location site using the Hydra Go device. The surface roughness parameter (s) is calculated using the one-meter-long metallic plate, which is painted with a 1 cm^2 spaced grids, and the plate is pushed into the surface until the grid line reaches the lowest point on the surface and the roughness variation is then recorded. Following this, the roughness profile is digitized to obtain the RMS height (h) and correlation length (l). The study location is divided into two parts, Location 1 and Location 2 and the division of the datasets ensures that it follows the 70:30 (training: validation) ratio. The field measurement statistics for Locations 1 and 2 are shown in Table 6.1 and 6.2, respectively. Three different RMS height and correlation length values are used corresponding to three different sites. The RMS height and correlation values for site 1 and site 2 of location 1 are (h = 1.93, l = 6.1) cm and (h = 1.53, l = 5.89) cm respectively. The h=2.85 cm and l=7.86 cm correspond to location 2 (validation data set). All the datasets of Location 1 are used to train the WCM to estimate the model parameters A and B. Further, for the validation of the WCM, 30

percent of datasets from Location 2 are used, the and rest data from Location 2 are used in the optical and SAR LAI fusion algorithm.

Table 6.1 Statistics of biophysical and surface parameters of Location 1 in Varanasi district.

Testing Data Points (Location 1)				
Sampling date	LAI (m^2m^{-2}) min-max	h (cm) min-max	SM (m^3m^{-3}) min-max	No. of points
10/01/20	0.8-1.8	18-52	0.12-0.33	42
19/02/20	1.2-3.8	54-86	0.13-0.42	42
05/03/20	2.8-5.8	74-102	0.11-0.39	42

Table 6.2 Statistics of biophysical and surface parameters of different Location 2 in Varanasi district.

Validation Data Points (Location 2)				
Sampling date	LAI (m^2m^{-2}) min-max	h (cm) min-max	SM (m^3m^{-3}) min-max	No. of points
15/02/23	2.1-4.2	22-78	0.09-0.34	31
27/02/23	2.5-5.9	55-103	0.15-0.37	31
23/03/23	1.8-3.5	66-93	0.12-0.29	31

6.3 Methodology

6.3.1 LAI calculation from optical data

The LAI retrieval using optical Sentinel-2 datasets was performed using a biophysical processor within the SNAP toolbox. Specifically, the biophysical processor employed a combination of the leaf Radiative Transfer Model (RTM) PROSPECT-4 and the canopy RTM SAIL, known as PROSAIL, to simulate the radiation transfer process within vegetation canopies and estimate vegetation biophysical parameters such as LAI and leaf

chlorophyll content. The biophysical processor utilized eight reflectance bands and three angle bands, along with PROSAIL and ANN, to produce the LAI estimates. The ANN algorithm is employed to optimize the PROSAIL model and improve the accuracy of LAI estimation.

6.3.2 Simulation of LAI from SAR datasets through WCM

In this study, the authors consider the soil surface and vegetation cover using the Water-Cloud model, a basic form of the first-order radiative transfer solution. The model computes the backscatter coefficient (σ^0) by assuming a uniform distribution of water particles within a homogeneous dielectric slab, representing the simplified vegetation canopy. The Water-Cloud model is adapted to its traditional form and improved by including plant height as a parameter to enhance the estimation of vegetation cover. The model is formulated as follows:

$$\sigma^0_{pq} = \sigma^0_{veg} + \tau^2 \sigma^0_{soil} \tag{6.1}$$

where

$$\sigma^0_{veg} = \frac{A cos\theta h}{2BV}(1 - e^{-2VBsec\theta}) \tag{6.2}$$

The constants A and B are empirical in nature. A is a parameter that represents the scattering of vegetation, while B represents the combined effect of vegetation scattering and sensor configuration. V is the vegetation descriptor and in this study, $V = LAI$ is used. The plant height is represented by h and θ is the look angle.

6.3.3 Modified integral equation model

The backscattering coefficient of soil over bare soil surfaces with random roughness can be described using the physical model of IEM, which is expressed as follows:

$$\sigma_{soil}^{pq} = \frac{k^2}{4\pi} e^{-2k^2 s^2 cos^2\theta} \sum_{n=1}^{\infty} |I_{pq}^n| \frac{W^{(n)}(2ksin\theta, 0)}{n!} \tag{6.3}$$

where

$$I_{pq}^n = 2ks\ cos\theta f_{pq} e^{-k^2 s^2 cos^2\theta} + (kscos\theta)^n F_{pq} \tag{6.4}$$

the constant ($k = \frac{2\pi}{\lambda}$ is wave-number, s is rms surface roughness, μ, ε_r are dielectric and permeability respectively. $f_{vv} = \frac{-2R_v}{cos\theta}$ in which $R_p(p = v\ or\ h)$ is polarised Fresnel coefficient. F_{vv} is defined in Equation (5)

$$F_{vv} = 2\frac{sin^2\theta}{cos\theta}\left[\left(1 - \frac{\varepsilon_r cos^2\theta}{\mu_r\varepsilon_r - sin^2\theta}\right)(1-R_v)^2 + \left(1 - \frac{1}{\varepsilon_r}\right)(1+R_v)^2\right] \tag{6.5}$$

To derive σ_{vh}^0 we need F_{vh} and f_{vh} in Equation 6.4. For F_{vh} we need the incident term which is taken from the F_{vv} and scattered term which taken from the F_{hh} and the following derivation was performed

$$F_{vh} = (ax + b'y')^2 = (a^2x^2 + 2ab'xy' + b'^2 + b'^2y'^2)$$

$$F_{vh} = \left[a^2(1+R_v)^2 + 2ab'(1+R_v)(1-R_h) + b'^2(1-R_h)^2\right] \tag{6.6}$$

$$\begin{aligned} F_{vh} = &\left(\frac{\sin^2\theta}{\cos\theta} - \frac{sq}{\varepsilon_r}\right)(1+R_v)^2 - 2\sin^2\theta\left(\frac{1}{\cos\theta} - \frac{1}{\sqrt{\mu_r\varepsilon_r - \sin\theta}}\right)(1+R_v)(1-R_h) \\ &+ \left(\frac{\sin^2\theta}{\cos\theta} - \frac{\mu_r(1+\sin^2\theta)}{\sqrt{\mu_r\varepsilon_r - \sin^2\theta}}\right)(1-R_h)^2 \end{aligned} \tag{6.7}$$

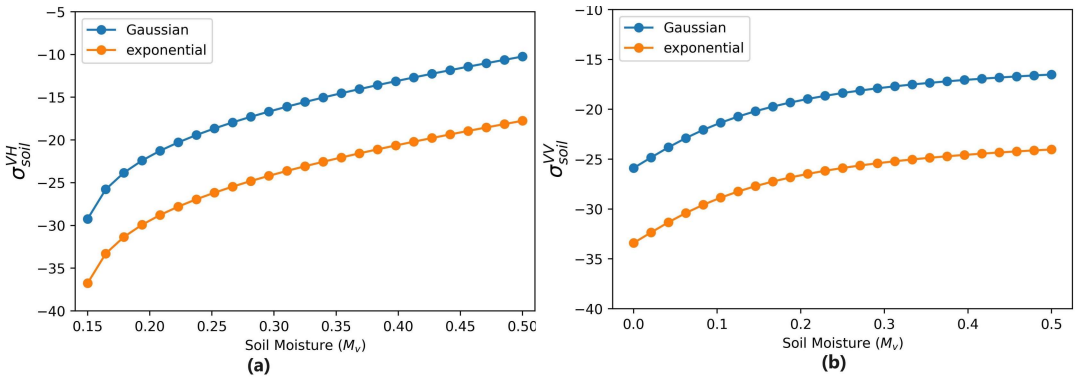

Figure 6.2 Evaluation of soil backscattering coefficient's (a) Sensitivity for σ^{VH}_{soil} and (b) Sensitivity for σ^{VV}_{soil} with Gaussian and exponential correlation functions.

where $x = (1+R_v)$ and $y' = (1+R_h)$, by substituting Equation 6.7 and $f_{vh} = \frac{-2\sqrt{R_h R_v}}{cos\theta}$ in Equation 6.4, the calculation of I^n_{vh} can be done. Figures 6.2(a) and 6.2(b) illustrate the sensitivity analysis of the soil backscattering coefficients, $\sigma^{VH}soil$ and $\sigma^{VV}soil$, respectively. $W^n(a,b)$ is the Fourier transform of the nth power of the surface correlation $\rho(x,y)$ function:

$$W^n(a,b) = \iint \rho^n(x,y)e^{iax+by}dxdy \tag{6.8}$$

The analysis is performed using the two different correlation functions for 1-d: Gaussian($\rho(x) = e^{-\frac{x^2}{L^2}}$) and exponential($\rho(x) = e^{-\frac{x}{L}}$) against the different values of soil moisture(m_v). The aim is to assess the impact of the correlation function on the sensitivity of the soil backscattering coefficients for both VH and VV polarizations. The analysis found that the Gaussian function exhibits better sensitivity for both VH and VV polarizations. Therefore, the Gaussian correlation function is adopted for the soil backscattering calculation.

6.3.4 WCM parametrization and inversion

The parameterization of the Water–Cloud model is done to optimize the unknown parameters A and B for the wheat crop. The procedure followed iterative optimization algorithms

such as the Levenberg-Marquardt algorithm (McNairn et al. (2012)). In this algorithm, an initial guess for the parameter's values is made at the start and refined through iteration. The algorithm terminates when minimization conditions are met. The computed model parameters for VV and VH polarizations at different dates are shown in Tables 6.3 and 6.4, respectively. After the parameterization of the model look-up table (LUT) search approach for the inversion of WCM is adopted (Lera and Pinzolas (2002)). In the Python 3.2 software, a lookup table (LUT) is generated by conducting forward simulations of the WCM model. This involved utilizing fixed values for parameters A and B while systematically changing the input variables such as LAI, h, and soil moisture. The adjustments in this analysis were guided by prior knowledge of the field, with soil moisture values ranging from 0.05 to 0.40 m^3m^{-3} and LAI values falling within the practical range of 0.5 to 6.2 m^2m^{-2}. The increments step for both soil moisture and LAI are set at 0.2. Plant height (h) ranging from 10 to 110 cm at an interval of 1.5 cm is taken. The combination resulted in more than 3000 unique cases for both σ^0_{VV} and σ^0_{VH} (Das et al. (2023).)

Table 6.3 The model parameters estimated after training the WCM model for VV polarisation.

	VV		
Parameters	January 10	February 19	March 05
A	0.276	0.211	0.168
B	0.081	0.086	0.087

Table 6.4 The model parameters estimated after training the WCM model for VH polarisation.

	VH		
Parameters	January 10	February 19	March 05
A	0.091	0.064	0.071
B	0.111	0.159	0.187

6.3.5 Fusion approach for optical and SAR data

We utilized a fusion approach that combined optical and SAR-estimated LAI to obtain a more comprehensive assessment of vegetation dynamics. To ensure that the assessment is unbiased, the unused data of Location 2 is used in the fusion approach. We employed three methods in the fusion of optical and SAR LAI:

1. Non-linear regression: In this approach to account for potential non-linear relationships in the LAI data estimated from SAR and optical, a polynomial regression model with a degree of 2 is utilized to predict the LAI and compared against the observed LAI.

2. Principal Component Analysis: The basic idea is to use PCA to extract the most essential spectral and textural features from each optical and SAR LAI. Once the principal components have been extracted from each data set, they can be combined into a new, fused data set using a weighted sum of the principal component scores.

3. Deep learning approach: In the deep learning approach algorithm is structured using a sequential model from the TensorFlow library. It comprises three layers: an input layer with ReLU activation, followed by two hidden layers, and an output layer. The first hidden layer contains 64 units, and the second hidden layer contains 32 units, both employing the ReLU activation function. The output layer is designed with a linear activation function. To optimize the hyper-parameters Adam optimizer was employed. Training the model was carried out over 50 epochs, with a validation split of 30 percent. The input data consisted of optical and SAR LAI measurements, resulting in a dataset with two features. This deep learning model incorporated two hidden layers where the first hidden layer comprised 64 nodes, while the subsequent hidden layer contained 32 nodes.

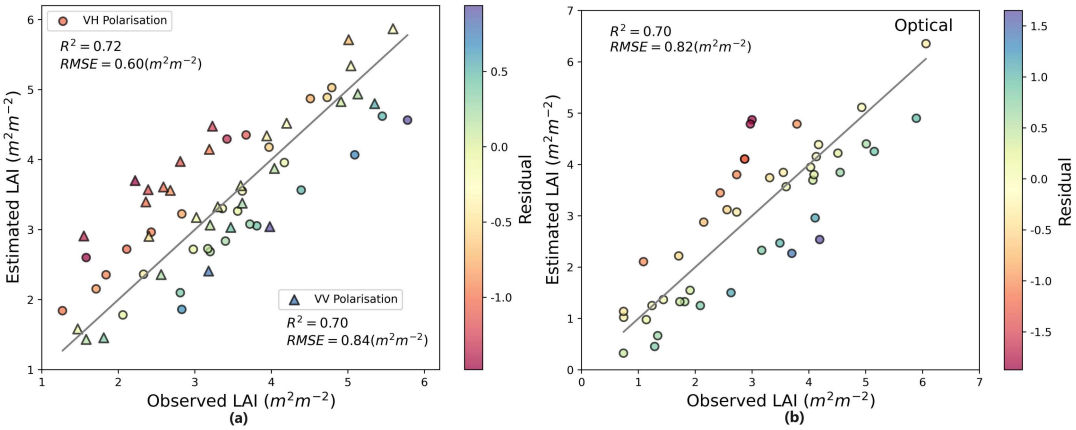

Figure 6.3 (a) Shows the estimated and observed LAI at VV and VH polarizations (b) Shows the estimated LAI calculated from PROSAIL RTM and plotted against the observed LAI.

6.4 Result and discussion

6.4.1 LAI estimation from SAR and Optical data

In this study, LAI estimation is performed using both SAR and optical data, and their performance is compared against the ground-measured LAI. Figure 6.3 compares the estimated LAI from SAR and optical data with the ground-measured LAI. Both the SAR and optical models produced good results, but the SAR model's performance is better. Specifically, in SAR estimation, the VH polarization showed a better correlation ($R^2 = 0.72$) and lower RMSE ($0.60 m^2m^{-2}$) than the VV polarization ($R^2 = 0.70$, $RMSE = 0.84 m^2m^{-2}$). The estimation accuracy of the LAI from L2A optical data is $R^2 = 0.70$, $RMSE = 0.82 m^2m^{-2}$. The better performance of the SAR model, especially in VH polarization, can be attributed to the modifications made to the IEM with the mapping of the surface roughness using the Gaussian spectrum. These modifications helped to improve the SAR model's accuracy by reducing the effect of noise and enhancing the model's sensitivity to the LAI variation.

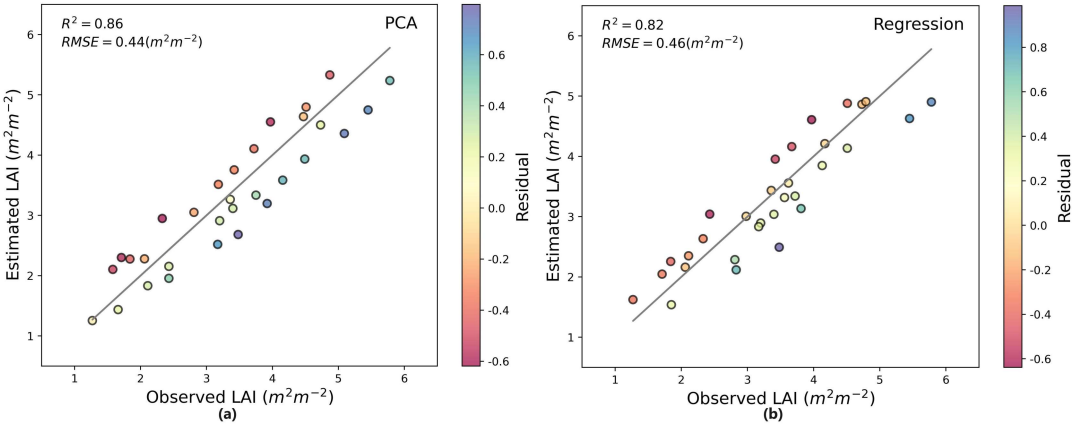

Figure 6.4 Comparison of LAI estimated from the (a) PCA (b) Regression plotted against the ground measurements.

6.4.2 Fusion of LAI estimated from the SAR and optical data

In this study, different algorithms have been developed to achieve accurate LAI estimation, such as PCA, non-linear regression, and deep learning techniques. Figures 6.4(a) and 6.4(b) compared the accuracy of PCA and non-linear regression methods. The results show that the PCA method outperformed the regression method in terms of accuracy. The correlation coefficient (R^2) was 0.86 for the PCA and 0.82 for the regression methods. The root mean square error (RMSE) was also lower for the PCA method, with a value of 0.44 m^2m^{-2} compared to 0.46 m^2m^{-2} for the regression method. There are several reasons why the PCA method performed better than the regression method in this study. First, the PCA method can extract the most significant features from the data and reduce the effect of noise, which can enhance the accuracy of the result. Second, the PCA method is more flexible and can handle complex relationships between the input features and the LAI estimation. Finally, the PCA method is less prone to overfitting compared to the regression method, which can lead to better generalization performance.

Figure 6.5(a) shows the fusion of LAI through the deep learning algorithm. The deep learning algorithm produced promising results for fused LAI prediction. The R^2 value of 0.91 indicates a strong correlation between the observed and predicted LAI

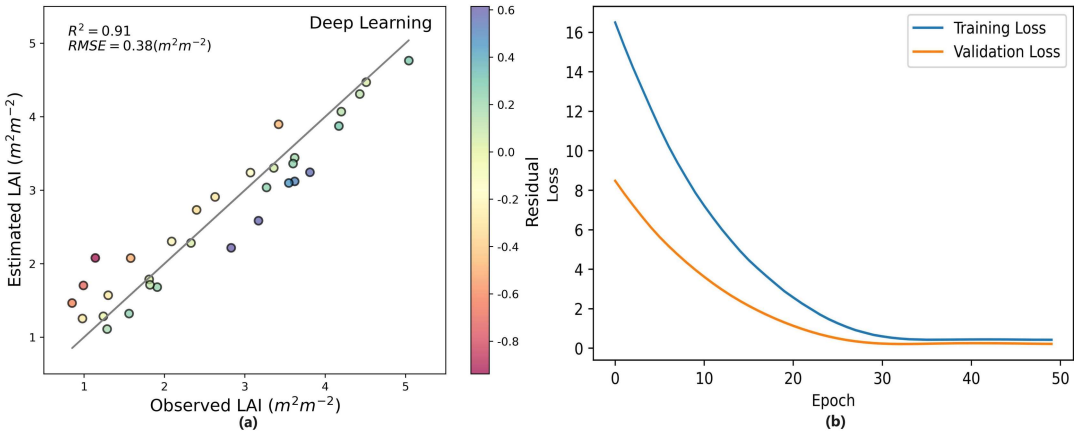

Figure 6.5 (a) Shows the accuracy of the estimated LAI from the deep learning against the ground LAI data. (b) Shows the training and validation loss over time.

values. Additionally, the low RMSE value of 0.38 m^2m^{-2} suggests that the model's predictions are accurate and consistent with the observed data. During training, the loss function decreased steadily with each epoch shown in Figure 6.5(b). This trend indicates that the model effectively learned from the data and optimized its parameters to minimize loss. The final training loss of 0.4258 and validation loss of 0.2181 further support the model's effectiveness. The results suggest that the deep learning algorithm successfully predicted fused LAI values from the given data compared to other techniques like non-linear regression and PCA. This approach can potentially improve LAI prediction accuracy and help to monitor vegetation growth and health. An empirical relation has been established between LAI values derived from a deep learning model and the corresponding backscattering values. This relationship has been employed to create an LAI map, as illustrated in Figure 6.6. These maps serve the purpose of visualizing the spatio-temporal changes in LAI. Notably, the map demonstrates a discernible upward trend in the average LAI values, indicating the progressive growth of the crop.

In this study, a significant advancement in the field by combining optical and SAR data through deep learning, principal component analysis, and non-linear regression techniques. In line with previous studies (Castro-Valdecantos et al., 2022; Estornell et al., 2013; Ma

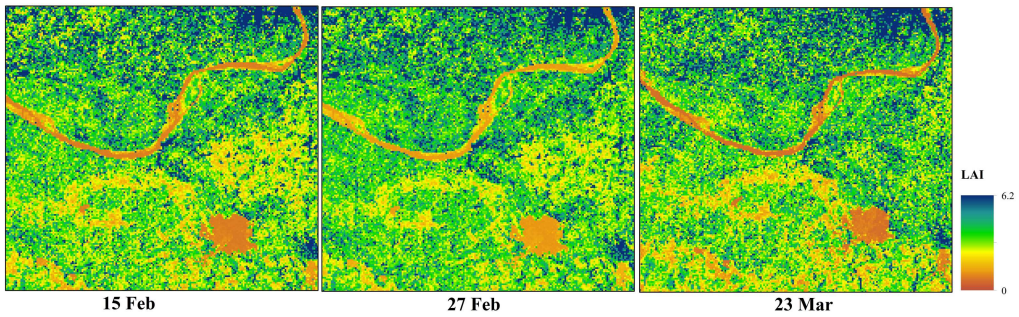

Figure 6.6 Spatio-temporal maps showing the variation of LAI at the large landscape.

et al., 2019), these results confirm that the fusion of multi-sensor data enhances the accuracy of LAI estimation, as it harnesses the complementary information provided by both optical and SAR modalities.

6.5 Conclusion

This study introduced a novel derivation of the backscattering coefficient for the IEM model at VH polarization. The goal is to estimate the LAI from SAR data at VV and VH polarization and from the optical dataset. In order to fuse the LAI data obtained from both sources, deep learning, PCA, and regression techniques were employed. The results show that the deep learning technique outperformed the other two methods regarding LAI estimation accuracy, with an R^2 value of 0.91 and an RMSE of 0.38 $m^2 m^{-2}$. The trend in the loss function during training indicated that the deep learning model was able to converge quickly to an optimal solution. Overall, the findings suggest that the proposed approach has the potential to improve LAI estimation accuracy from SAR and optical data, which could be useful in a variety of applications such as agriculture and forestry monitoring. Future research could focus on exploring the potential of using other deep learning models and incorporating additional SAR and optical data sources to further improve LAI estimation accuracy.