Comparison of parametric and non-parametric methods for chlorophyll estimation based on high-resolution UAV imagery

Gaurav Singhal1,2, Babankumar Bansod1,2,*
Lini Mathew3, Jonali Goswami2, B. U. Choudhury1 and P. L. N. Raju4
1CSIR-Central Scientific Instruments Organization, Chandigarh 160 030, India
2Academy of Scientific and Innovative Research, Ghaziabad 201 002, India
3National Institute of Technical Teachers, Training and Research, Chandigarh 160 019, India
4North East Space Application Centre, Barapani 793 103, India
5ICAR Research Complex for NEH Region, Umiam 793 103, India

The present study provides a systematic comparison of parametric and non-parametric retrieval methods using high-resolution data provided by the unmanned aerial vehicle (UAV). We used turmeric crop reflectance data to evaluate the vegetation index (VI)-based parametric methods and compared them with linear and nonlinear non-parametric methods to build a rigorous LCC estimation model. The study demonstrates that the best-performing VI was the normalized green red difference index (GNRDI), with \( R^2 = 0.68, \text{RMSE} = 0.13 \) and high processing speed of 0.08 s. With regard to non-parametric methods, almost all methods outperformed their parametric counterparts. Particularly, methods such as random forest (RF) and kernel ridge regression (KRR) showed the best performance characterized by \( R^2 \geq 0.72 \) and \( \text{RMSE} \leq 0.12 \) mg/g of fresh leaf weight. These non-parametric methods possessed the benefit of total spectral information utilization and enabled robust, non-linear relationship between the predictor and target variables, but computational complexity is a major drawback.

Keywords: Chlorophyll, machine learning, unmanned aerial vehicle, vegetation index.

BEING the most important pigment of all photosynthetic cells, variations in leaf chlorophyll concentration (LCC) are an indicator of crop growth and stress status and help in estimating biomass and yield\(^1\text{-}^3\). Chlorophyll is best described by its absorption in the red band (600–720 nm) and major reflection from green and NIR wavebands\(^4\).

Remote sensing sensors play an important role in crop health monitoring in terms of large coverage area and fast estimates of crop biophysical and biochemical parameters such as LCC. Remote sensing satellites are capable of providing a vast coverage, but their potential is limited by poor resolution and cloud cover. Flying at low altitude...
and beneath the clouds, UAV overcomes these limitations and provides more flexibility\textsuperscript{5}. It can provide high spatial and temporal resolution data. The high resolution allows precise estimation of plant biophysical parameters such as LCC and leaf nitrogen content from multispectral (MS) images of a field.

Several non-destructive methods, including parametric and non-parametric approaches to estimate the LCC have been proposed. VI-based parametric approach has been fascinating in the remote-sensing community to estimate the crop biophysical parameters such as LCC\textsuperscript{6–8}. VI can extract more than 90% of spectral information of the vegetation. Index normalization typically upgrades the interpretability of plant properties by diminishing the effect of non-photosynthetic plant components, for example, soil background\textsuperscript{9}, but they are prone to show poor model performance because of some limitations\textsuperscript{10} such as atmospheric effects and canopy structural characteristics\textsuperscript{9,11}. In addition, VI regression models are likely to show unstable performance when imposed to MS images differing from the simulated ones.

Another set of methods to retrieve the LCC comprises non-parametric methods including multiple linear regression (MLR), principal component regression (PCR), partial least square regression (PLSR), support vector machine (SVM), relevance vector machine (RVM), random forest (RF) and kernel ridge regression (KRR). Literature review unfolds that these methods utilize full spectral data, avoids collinearity and provides a robust relationship between predictor variables and the variables of interest\textsuperscript{12}. So far, only a few studies have been carried out to estimate the LCC of turmeric based on high-resolution UAV imagery\textsuperscript{13–15}. Therefore, the explicit objective of this study includes the assessment of these retrieval methods for LCC estimation of turmeric crop using UAV based high-resolution multispectral imagery.

The UAV used for this field study was a Hex-copter from DJI (www.DJI.com). It was equipped with multi-spectral sensor (parrot sequoia), consisting of four bands, i.e. green (spectral bandwidth of 530–570 nm, centred at 550 nm), (640–680 nm, 660 nm), Red-Edge (730–740 nm, 735 nm) and NIR (770–810 nm, 790 nm). The spatial resolution achieved was 2 cm at an elevation of 30 m. The image capture rate was one frame per second. UAV flights were carried out in ambient light conditions within the time frame of 11:30 a.m. and 1:00 p.m. (IST). Pix-4D image processing software was used to perform radiometric calibrations, generating orthomosaics and reflectance maps from UAV images. A white reference (calibration) panel was used to calibrate and correct the acquired UAV data into surface reflectance based on the illumination condition and properties of sensors. For more information about the data collection, please refer to the previous study\textsuperscript{16}.

For this study, flight’s chlorophyll and reflectance data were collected from the research farms of Indian Council of Agricultural Research (ICAR) for North East hilly region at Umiam (Shillong, Meghalaya) located at 25°41’N lat, 91°55’E long. Turmeric was grown over 20 equal-sized plots with 35–40 plants in each plot. Simultaneously with UAV flight, the SPAD-502 chlorophyll metre readings were taken from plant samples. Six hundred reflectance and SPAD values from 20 plots (30 samples per plots) were available to develop the model. The SPAD-502 chlorophyll metre measurements exhibited an exemplary agreement with laboratory-measured leaf chlorophyll concentration (LCC) values in terms of regression coefficient ($R^2 = 0.92$), which was reported earlier\textsuperscript{16}. These spectral reflectance values of all bands were combined with LCC measurements to calibrate and validate the LCC estimation model.

The parametric methods used in this study were based on the calculation of VIs from reflectance maps. A total of 11 two and three bands indices (Table 1) were formulated to regress them against the LCC using ordinary least square method.

Non-parametric methods were subdivided into linear and non-linear methods; former include MLR, PCR and PLSR. These methods are fascinating because of their fast speed and develop a linear relationship between the predictor and target variables. Methods like PCR and PLSR often remove the collinearity with dimensionality reduction approach\textsuperscript{27}. Non-linear non-parametric methods are also known as machine learning methods. These include SVR, RVR, RF and KRR, which are comprehensively described in the following section.

Support vector regression (SVR) combined with a kernel function, maps the training samples into a higher dimensional feature space, which is nonlinearly related to the original space\textsuperscript{25}. It is used for classification purpose when the number of features is larger compared to the number of data points in the dataset. SVR using radial basis function (RBF) kernel requires the tuning of two hyper-parameters, i.e. the cost-penalty parameter C and the standard deviation ($\sigma$). C and $\sigma$ parameters directly control the model’s performance. For higher values of C, the model becomes intolerable for large errors, and is more prone to overfitting, whereas, for smaller values of C, the model shows rigidity and shows underfitting.

The relevance vector machine (RVM) is a well-recognized machine learning algorithm that uses a Bayesian framework and follows a probabilistic sparse kernel model for classification and regression\textsuperscript{26}. This approach delivers a statistical relationship based on only a smaller number of training samples, called relevance vectors. Sparsity is achieved because the posterior distributions of weights are sharply peaked around zero.

The kernel ridge regression (KRR) is essentially regularized SVM. A regularization parameter ($\lambda$) is added to the cost function to control the values of the anticipated parameters, forms the closed from solutions\textsuperscript{27} and removes overfitting. The KRR method\textsuperscript{28} solves the cost
Table 1. Vegetation index used in this study

<table>
<thead>
<tr>
<th>Vegetation index (Acronym)</th>
<th>Formula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized difference vegetation index (NDVI)</td>
<td>[ \frac{NIR - RED}{NIR + RED} ]</td>
<td>17</td>
</tr>
<tr>
<td>Normalized green red difference index (NGRDI)</td>
<td>[ \frac{GREEN - RED}{GREEN + RED} ]</td>
<td>18</td>
</tr>
<tr>
<td>Normalized difference red-edge index (NDRE)</td>
<td>[ \frac{NIR - RED _{EDGE}}{NIR + RED _{EDGE}} ]</td>
<td>19</td>
</tr>
<tr>
<td>Green normalized difference vegetation index (GNDVI)</td>
<td>[ \frac{NIR - GREEN}{NIR + GREEN} ]</td>
<td>13</td>
</tr>
<tr>
<td>Chlorophyll vegetation index (CVI)</td>
<td>[ \frac{NIR * RED}{GREEN^2} ]</td>
<td>20</td>
</tr>
<tr>
<td>Soil adjusted vegetation index (SAVI)</td>
<td>[ \frac{NIR - RED}{NIR + RED} - 0.5 ]</td>
<td>20</td>
</tr>
<tr>
<td>Simple ratio vegetation index (SRVI)</td>
<td>[ \frac{NIR}{GREEN} ]</td>
<td>2</td>
</tr>
<tr>
<td>Chlorophyll green index (Clgreen)</td>
<td>[ \frac{NIR}{GREEN} - 1 ]</td>
<td>22</td>
</tr>
<tr>
<td>Chlorophyll red-edge index (Clrededge)</td>
<td>[ \frac{NIR}{RED _{EDGE} - 1} ]</td>
<td>20</td>
</tr>
<tr>
<td>Modified chlorophyll absorption ratio index (MCARI)</td>
<td>[ \frac{(RED _{EDGE} - RED) - 0.2(RED _{EDGE} - GREEN)}{(RED _{EDGE}/RED)} ]</td>
<td>23</td>
</tr>
<tr>
<td>Transformed chlorophyll absorption ratio index (TCARI)</td>
<td>[ \frac{3(RED _{EDGE} - RED) - 0.2(RED _{EDGE} - GREEN)}{(RED _{EDGE}/RED)} ]</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics of individual bands and spectral indexes derived from UAV images

<table>
<thead>
<tr>
<th>Spectral index</th>
<th>Range</th>
<th>Mean, standard deviation</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td>8.57–20.79</td>
<td>13.28, 2.47</td>
<td>0.64</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.028–0.891</td>
<td>0.658, 0.151</td>
<td>0.78</td>
</tr>
<tr>
<td>NDRE</td>
<td>0.07–0.32</td>
<td>0.15, 0.03</td>
<td>0.28</td>
</tr>
<tr>
<td>NGRDI</td>
<td>0.290–0.515</td>
<td>0.47, 0.04</td>
<td>0.83</td>
</tr>
<tr>
<td>GNDVI</td>
<td>0.52–0.74</td>
<td>0.64, 0.04</td>
<td>0.68</td>
</tr>
<tr>
<td>SAVI</td>
<td>0.009–0.469</td>
<td>0.302, 0.086</td>
<td>0.734</td>
</tr>
<tr>
<td>CVI</td>
<td>1.13–2.67</td>
<td>1.69, 0.33</td>
<td>0.48</td>
</tr>
<tr>
<td>Cl-Green</td>
<td>2.23–5.88</td>
<td>3.71, 0.73</td>
<td>0.64</td>
</tr>
<tr>
<td>Cl-RedEdge</td>
<td>0.154–0.94</td>
<td>0.35, 0.09</td>
<td>0.203</td>
</tr>
<tr>
<td>MCARI</td>
<td>−0.498–0.097</td>
<td>−0.183, 0.112</td>
<td>0.26</td>
</tr>
<tr>
<td>TCARI</td>
<td>0.22–0.90</td>
<td>0.51, 0.123</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Radial basis function (RBF) = \( e^{\frac{-(x-x')^2}{2\sigma^2}} \), \( (1) \)

where \( x \) and \( x' \) are two samples representing feature vectors. \( \sigma \) is the tunable parameter for RBF kernel. It directly affects model performance.

It ought to be noticed that the whole data set containing the predictor variables and LCC was divided into training and testing dataset in a ratio of 70 : 30 for model development and validation. \( K \)-fold cross-validation (\( K = 10 \)) was used for hyper-parameters tuning, where the training data was further divided into training and validation dataset at 90 : 10 ratios for \( K \) times.

We first concentrated on assessing the connection between LCC and multispectral VIs based on the coefficient of determination \( R^2 \) and RMSE. Later, we evaluated non-parametric methods and contrasted the outcomes with optimal VIs based on \( R^2 \) and RMSE.

Relationship between LCC and 11 different two-band vegetation indices was recognized using the ordinary least square method. Table 2 represents the average correlation coefficient (\( R \)) values between multispectral VIs and LCC using \( K \)-fold cross-validation of 100 iterations. Results elucidate that both NGRDI and NDVI are the best performing vegetation indices (\( R = 0.83 \) and 0.78) respectively, outperforming other indices. These results agree with earlier studies\(^{25,31,32} \), where NGRDI and NDVI were found as efficient methods for LCC estimation in crop health monitoring. It is noteworthy that both optimal indices consisted of the red band (660 nm). Despite the fact that outcomes of adequate accuracy were acquired in this study utilizing this basic model, various disadvantages
The large processing time of RVR (16.35 sec) limits this approach to integrate with real-time crop health monitoring systems.

The KRR method proved to be the best performing nonlinear regression method among all methods yielding, an RMSE of 0.10 mg/g, $R^2$ of 0.7452 and extremely fast processing speed of 1.434 seconds as the most robust outcomes. This approach clearly distributed the LCC values spatially throughout the turmeric plots. This approach set a nonlinear relationship between LCC and spectral data and had the advantage of handling a large number of datasets with high speed and accuracy. The results of KRR achieved in this study corroborate with previous studies.$^{12,34}$

Table 3 shows the $R^2$ and RMSE values for all machine learning methods with respect to training and testing dataset. A larger difference between training and testing dataset represents the high-performance difference between them. KRR showed the least difference between the two sets of data as this method is less sensitive to the outliers. Although the RF is also not sensitive to outliers, it could not perform well with the test data with the smaller number of samples. Overall, the KRR model performed best with both the datasets and therefore, can be used for generating the chlorophyll maps.$^{16}$

Figure 1a presents scatter plots of estimated versus measured LCC values of all samples derived from the best performing non-parametric KRR algorithm. From the scatter plot, it can be deduced that KRR is superior to NGRDI in terms of model performance as the estimated LCC values are closer to 1 : 1 line than the optimal VI scatter plot. Notwithstanding its better execution, KRR is substantially more computationally complex than NGRDI due to optimization steps.

We also measured the importance of predictor variables using the Gaussian process regression (GPR)$^{15}$. GPR band assessment technique was performed in the Artmo toolbox in MATLAB and the best contributing band was detected on the basis of sigma value as shown in Figure 1b. Since lower sigma value leads to higher information, the red (660 nm) and green (550 nm) bands hold the biggest quality impact on LCC estimation, regardless of distinctive methodologies utilized for LCC estimation. These results justify the selected vegetation indices, which show the dominance of red band at least for this study site and specific crop.

Although several studies have endeavoured to utilize UAV-based MS images to estimate the LCC status, the datasets were not sufficient enough to get the satisfactory performance of the models for generating LCC maps at the field level.
LCC estimation accuracy. Conversely, the outcomes of this analysis demonstrate that GNDRI created higher assessments, while the non-parametric KRR technique prompted a higher level of accuracy under various conditions. These outcomes propose that UAV-based MS images give a promising methodology that can be integrated with crop health status monitoring. Moreover, even after getting higher order estimation accuracy, the built-up LCC estimation model in any case should be tried with another crop and ecological site with distinct conditions.

In this study, parametric and non-parametric modelling techniques were compared to estimate the LCC using UAV-based high-resolution images. Results demonstrated that NGRDI and NDVI are highly correlated with LCC, but show some limitations like saturation effect at higher LCC values. The non-parametric techniques outperformed parametric techniques in terms of model performance. Furthermore, among nonparametric schemes, techniques like RF and KRR showed the best performance. Linear non-parametric approach was outperformed by its nonlinear partners in this study. The extremely fast speed recommends them to be integrated with the crop health status monitoring systems.

Disclosure statement: Authors declare no conflict of interest to disclose.

35. Verrelst, J., Rivera, J. P., Gitelson, A., Delegido, J., Moreno, J. and Camps-Valls, G., Spectral band selection for vegetation properties retrieval using Gaussian processes regression. Int. J. Appl. Earth Observ. Geoinform., 2016, 52, 554–567.\* ACKNOWLEDGEMENTS. The authors thank the UAV team of NESAC (Shillong) and ICAR (Shillong) for their help in carrying out field data collection and analysis. The experimental research was funded by Department of Science and Technology (DST), New Delhi under the project ‘Characterization of acid soils under different land uses and its impact on crop growth: a hyperspectral approach’ (Grant Number: DST-BBDID/01/23/2014 – HSRS dated 18.03.2016).
Received 2 June 2019; revised accepted 24 July 2019
doi: 10.18520/cs/v117/i11/1874-1879

Fate and transport of microplastics from water sources

Monnisha Ganesan¹, Gobi Nallathamby¹∗ and Seshachalam Srinivasulu²

¹Department of Textile Technology, and
²Department of Geology, Institute of Ocean Management, Anna University, Chennai 600 025, India

Microplastics as environmental pollutants affect surface water and groundwater. Surface water, groundwater and branded drinking water bottles were analysed in and around Chennai, Tamil Nadu, India. The total count of microplastics was found to be 66 particles with fibrous and fragmented shape, colours such as white, blue, green, yellow, pink and black under optical microscope. SEM-EDX used to study morphology and elemental analysis of microplastics confirmed the presence of heavy metals such as Cr, Ti, Mo, Ba and Ru adhered to their surface. Polyethylene terephthalate and polyamide were confirmed by the presence of functional groups of the polymers by FTIR equipped with attenuated total reflectance.

Keywords: Microplastics, heavy metals, pollution, polyamide, water sources.

PLASTICS are synthetic organic polymers formed by the process of polymerization. The use of plastics has increased worldwide and the annual production is around 322 million metric tonnes. In India, approximately 5.6 million tonnes (mt) of plastic waste is generated annually. Plastic debris which is less than 5 mm is referred to as microplastic, and categorized as primary and secondary. The main sources of microplastics in the marine environment are land and sea-based litter. Microplastics are a big threat to marine organisms as they are ingested by them. A study was done on the distribution, weathering and chemical characteristics of microplastics on the beaches of Goa, India during the southwest and northeast monsoon seasons. The distribution and characteristics of microplastic pollution along the coast of Chennai, Tamil Nadu (TN), India during pre- and post-flood were also studied. Presence of microplastics and their distribution and characteristics were reported from Rameswaram Coral Island, TN, India’s marine water from Kuala Nerus and Kuantan Port in Malaysia, drinking water treatment plants in Germany, freshwater resources like Vembanad lake in Kerala, Huron Lake in Canada, and three urban estuaries in China. Microplastics were also identified in bottled drinking water, because of the packaging materials (polyethylene terephthalate), which are consumed by humans. In India, only a few studies are available on microplastics contamination in sediment samples of...