



Sensitivity analysis of artificial neural network for chlorophyll prediction using hyperspectral data

Prashant K. Srivastava^{1,2} · Manika Gupta³ · Ujjwal Singh¹ · Rajendra Prasad⁴ · Prem Chandra Pandey⁵ · A. S. Raghubanshi¹ · George P. Petropoulos⁶

Received: 8 May 2019 / Accepted: 12 June 2020
© Springer Nature B.V. 2020

Abstract

Hyperspectral acquisition provides the spectral response in narrow and continuous spectral channel. The high number of contiguous bands in hyperspectral remote sensing provides significant improvements in assessing subtle changes as compared to the multispectral satellite datasets in context of spectral resolution. Therefore, the main goal of the present research is to evaluate the sensitivity of the artificial neural networks (ANNs) for chlorophyll prediction in the winter wheat crop using different hyperspectral spectral indices. For evaluating relative variable significance in the study, the Olden's function has been applied. The Lek's profile method is used for sensitivity analysis of ANNs for chlorophyll prediction using the vegetation indices such as Red Edge Inflection Point (REIP), Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), and Structure-Insensitive Pigment Index (SIPI) derived from hyperspectral radiometer. The analysis indicates a high sensitivity of SAVI followed by NDVI, REIP and SIPI for chlorophyll retrieval using ANNs. The statistical performance indices obtained from calibration (RMSE=0.27; index of agreement=0.96) and validation (RMSE=0.66; index of agreement=0.83) suggested that the ANN model is appropriate for chlorophyll prediction with good efficiency. The outcome of this work can be used by upcoming hyperspectral missions such as Airborne Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) and Hyperspectral Infrared Imager (HyspIRI) for large-scale estimation of chlorophyll and could help in the real-time monitoring of crop health status.

Keywords Hyperspectral Radiometry · Vegetation indices · Sensitivity analysis · Neural network · Chlorophyll

1 Introduction

Hyperspectral remote sensing (HRS) technologies provide detailed spatio-spectral information simultaneously about the feature objects in a single pixel of an image (Ben-Dor et al. 2013). The narrow spectral bands of HRS has advantages as it can provide desired

✉ Prashant K. Srivastava
prashant.just@gmail.com; prashant.iesd@bhu.ac.in

Extended author information available on the last page of the article

wavelengths of the spectral range, which is not possible with multispectral bands. (Lamine et al. 2018; Singh et al. 2017). It has provided the techniques to capture information on a spatial rather than the pixel basis to signify the spectral information of the target features in narrow and continuous bands (Pandey et al. 2018). Thus, HRS has potential to enhance similar target features discrimination as compared to multispectral images with significant improvements in the results (Blackburn 1998; Petropoulos et al. 2015). HRS technologies can provide promising outcomes in crop types identification, stage growth and associated phenological changes (Cho et al. 2007). Due to narrow band and spatio-spectral methods of data acquisition, HRS has profound applications in several research domains (Malhi et al. 2020; Lamine et al. 2019; Anand et al. 2020). In agricultural domains, HRS is being used to estimate irrigational water demand (Smith 1992), biophysical parameters (Goel et al. 2003), yield prediction (Quarmby et al. 1993), Evapotranspiration (Asrar et al. 1984), as well as for disease and pest management (CÁrcamo and Spence 1994). The use of hyperspectral indices can further enhance the accuracy of the vegetation parameter estimation (Ortenberg et al. 2011; Thenkabail et al. 2000). This characteristics of HRS have advantages in developing indices to measure variable nutrient rate and crop growth analysis over other sensors (Blackburn 2006; Wei et al. 2008; Gupta et al. 2014).

After the development of sophisticated artificial intelligence (AI) techniques such as ANNs (Bharose et al. 2013; Srivastava et al. 2013), incorporation of hyperspectral remotely sensed datasets has enhanced the outcome probability for prediction of plant biochemical properties (Gupta et al. 2014; Mulla 2013). ANNs are considered as one of the precise and robust mathematical tools that were successfully incorporated in several remote sensing (RS) research domains for modelling problems analysis (Kisi and Shiri 2011; Wu and Chau 2011; Wu et al. 2010). ANNs have potential for accepting the challenges that are complex in nature such as pattern recognition (Petropoulos et al. 2012a, b) prediction and biological variables modelling for environmental aspects (Nagy et al. 2002; Schaap et al. 1998; Specht 1991). The principal drawback with the ANNs is that they are “black box” because of little explanatory perception of the independent variables used inside the prediction method (Olden and Jackson 2002; Tzeng and Ma 2005). In 2003, Gevrey et al. (2003) resolved this issue with a comprehensive correlation of different procedures for evaluating the variable significance in neural systems.

The model recognition step involves estimation of a suitable parameter set and the actual adjustment in the structure of the model (Gupta et al. 2006; Srivastava et al. 2014). Factors of each model and structure need adjustment until acceptable levels of agreement are obtained (Wagener et al. 2003). Sensitivity analysis (or SA) is to identify the most and/or least important parameters in explaining variances in the model output (Saltelli et al. 2004; Petropoulos and Srivastava 2016). It is evident that the method used to quantify the sensitivity of the model will also affect the outcome of the SA (Saltelli et al. 2008). Conventionally, in few RS studies, the local SA is commonly used to understand the input data and model structure (Saltelli et al. 2008). In local SA, a base case is taken for all model simulations in which all the parameters are set as predefined variable and then each parameter in turn perturbed by some specific level of variations (Gustafson et al. 1996; Saltelli et al. 2000). In general, in remote sensing-related SA, the calculated sum of squared difference between the base case and each perturbation is used as a measure to find the relative significance of each model parameters (Barton and North 2001). However, nowadays, researchers uses global SA or GSA to calibrate model parameters in remote sensing (Verrelst et al. 2015). The advantage of GSA is that in this all the model parameters vary simultaneously and therefore it is more useful for rigorous data analysis and model calibration (Ireland et al. 2015; Petropoulos et al. 2014).

Spectroradiometer is used in many applications, including in characterization of natural materials and vegetation analysis. During the last two decades, technological advances in hyperspectral Earth Observation (EO) technology in particular have been widely incorporated into vegetation studies (Goel et al. 2003; Thenkabail et al. 2000; Pandey et al. 2014, 2019; Anand et al. 2020), providing such estimates at different scales of observation in a rapid and cost-effective ways. Therefore, this research work has been carried out to study the sensitivity of ANNs for chlorophyll retrieval using hyperspectral datasets. The studies involve extensive fieldwork and subsequent laboratory analysis of crop samples, together with the advanced techniques integrated with sensitivity analysis to explore the detection of the spectral variability responses for total chlorophyll (T_Ch1). The paper has the following structure—Sect. 2 provides a quick overview of the study area, remotely sensed datasets, ANN structure and SA methods. Section 3 demonstrates research outcomes, results and discussion, while Sect. 4 focuses on final remarks and conclusions of the present study.

2 Materials and methods

2.1 Study area

The study site is located in the croplands of Saliyar village, Roorkee, (29° 51' 0" N latitude and 77° 53' 0" E longitude), categorized under humid subtropical climatic zone of India. Study region has three distinct crop seasons such as zaid (April–May) prevalent during summer season, kharif (June–September) during rainy season and rabi crop sown in winter season (October–March). The area receives an average annual rainfall of 1074 mm. Most of the rainfall received during kharif (84%) and rabi seasons (16%). Distribution is highly variable with high precipitation in northern part than the rest of the areas. Both high and low landforms are found in the surrounding, e.g. an elevation of 232 m was recorded in south, while high elevation of about 869 m was found in the north having the Shiwalik ranges. Main source of fresh water is through River Ganga, which is perennial river. It is most important district from the agriculture point of view and having all the fertile land for cultivation of crops. Soil is highly weathered with reddish-yellow colour, acidic in nature, having a clay-rich B horizon. In soil taxonomy, it comes under the category Ultisols. The location of the study site is shown in Fig. 1.

2.2 Canopy spectral and chlorophyll measurements

In this experiment, plots of size 4×4 square meter were designed and planted with the winter wheat crop. Vegetation radiance measurements were taken for each plot with ASD Field Spec Pro radio-spectrometer (Analytical Spectral instrument, Boulder, Co., USA). ASD has an in-built 25° instant discipline of view fibre optics, operated at 350–1050 nm spectral range having an interval of 1.5 nm for sampling.

The measurements comprises of 60 hyperspectral reflectance datasets with the estimation of Tot_Ch1 in the both manure-treated and controlled site of winter wheat crops. Before measurement of each reflectance, the radiance of a white standard board covered with BaSO₄ with prior known reflectivity was noted down for standardization purposes. The spectral properties of canopy were measured at a height around 1 m at nadir underneath cloudless/near cloudless conditions during 10:00–14:00 in the daytime. The finding

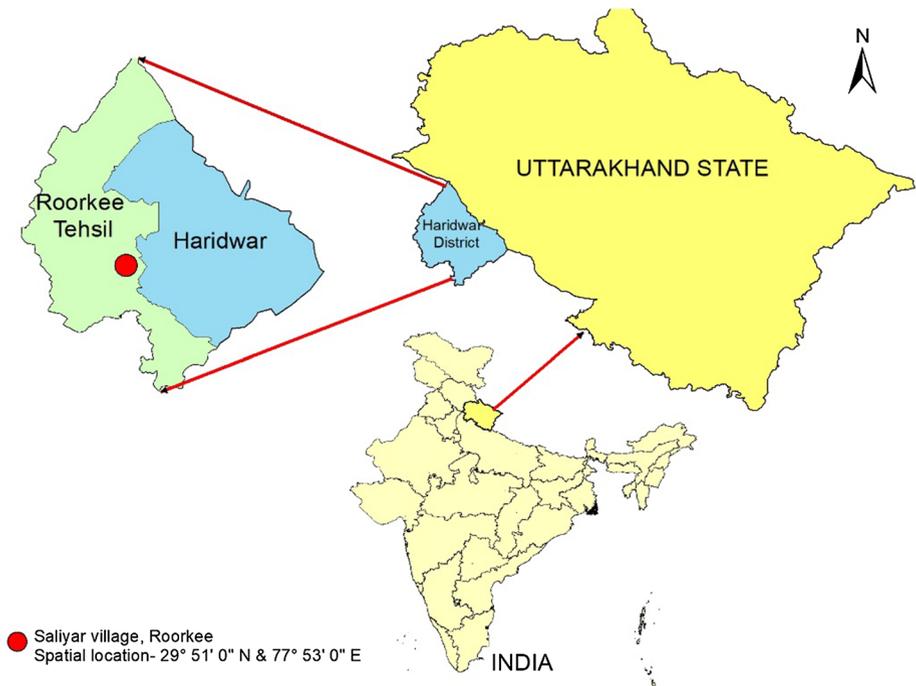


Fig. 1 Layout of the study area

set-up affirmed that the proportion of direct to diffuse solar radiation was steady during measurements.

Around 0.5 g of a crisp leaf with the aid of quartz sand was crushed until the point when no green shading was left in the crucible with acetone. Later, $(\text{CH}_3)_2\text{CO}$ (acetone) was blended in it to a volume of 50 ml and centrifuged at the speed for 30 min in order to separate the liquids from the solids. Every chlorophyll pigment has a particular absorption value. The trichromatic equation (calculation of different wavelengths at specific absorption point) has been utilized to locate the relative contribution of each pigment with reference to chlorophyll content. After centrifugation, a laboratory spectrophotometer (Hitachi Model) was used to measure the chlorophyll content at 663 nm for chlorophyll-a and 645 nm for chlorophyll-b (Arnon 1949). The following equation (Eq. 1) was used for the T_Ch1 (mg/l) estimation in the laboratory:

$$\text{T_Ch1 (mg/L)} = 20.2A_{645} + 8.02A_{663} \quad (1)$$

Afterwards, REIP for vegetation health, NDVI as an interpreter of vegetation greenness, SAVI as an indicative of soil-based indices and SIPI as a specific of chlorophyll-based indices were calculated from reflectance obtained from spectroradiometer. The calculation of REIP needs four-point interpolation technique. Studies by Clevers et al. (2002) indicated that the linear approach is computationally intense and powerful. It is the most useful method for the calculating REIP values from the hyperspectral data. This computation requires normally a simple interpolation, and only four channels of hyperspectral reflectance are needed. The specified four bands of wavelength are centred

at 670, 700, 740 and 780 nm. The reflectance estimations at 670 nm and 780 nm are used to find the inflection point (as illustrated in Eq. 2) and for the prediction of inflection point wavelength (Eq. 3). Thereafter, a linear interpolation procedure is used here as given by Kumar et al. (2002).

$$R_{\text{REP}} = (R_{670} + R_{780})/2 \quad (2)$$

$$\lambda_{\text{REP}} = \lambda_{700} + (\lambda_{740} - \lambda_{700}) \times \left[\frac{R_{\text{REP}} - R_{700}}{R_{740} - R_{700}} \right] \quad (3)$$

In 1988, Becker and Choudhury found spectral indices based on red and infra-red bands, sensitive to chlorophyll absorption (Becker and Choudhury 1988). Since then, NDVI has been broadly utilized among RS community for vegetation discrimination based on red and infra-red reflectance properties. It indicates that the health status of the crop and is measured by the following equation (Tucker 1979).

$$\text{NDVI} = (R_{\text{NIR}} - R_{\text{red}})/(R_{\text{NIR}} + R_{\text{red}}) \quad (4)$$

Ratio-based Vegetation Indices (VIs) are broadly preferred over soil-based indices (Broge and Mortensen 2002). Huete in (1988) proposed an index named as SAVI especially to minimize the soil background effect on the vegetation signature by including a constant soil adjustment factor (L) (Eq. 5). The authors demonstrated that NDVI is equal to SAVI when L corresponds to 0.

$$\text{SAVI} = ((L + 1)(R_{\text{NIR}} - R_{\text{red}}))/(R_{\text{NIR}} + R_{\text{red}} + L) \quad (5)$$

where R_{NIR} = near-infra-red reflectance, R_{red} = red reflectance, and L = soil adjustment factor (0–1).

In 1995, Penuelas et al. (1995) proposed new spectral indices named as SIPI to assess the leaf pigments using remotely sensed datasets. It utilizes the proportions of reflectance at 800, 445 and 680 nm, precisely to discover the proportions of carotenoids to Chl-a that generally reduces radiation effects at the leaf level and can be represented as follows (Eq. 6):

$$\text{SIPI} = (R_{800} - R_{445})/(R_{800} + R_{680}) \quad (6)$$

2.3 Artificial neural networks

To facilitate chlorophyll retrieval using ANNs, a traditional multilayered neural network was employed as shown in (Fig. 2). The connection among each input-to-node as well as node-to-node requires being modified by a weight in the network. This assemble contains an additional input in particular node. Each node is thought to have one as a constant value. Since one or more than one hidden layer is contained by a multilayered neural network, the biases and weights were initialized to the appropriately scaled values, before starting the training process. In 1990, Aleksander and Morton proposed that the output layer (O_a) can be calculated using linear activation function as shown in Eq. 7 (Aleksander and Morton 1990).

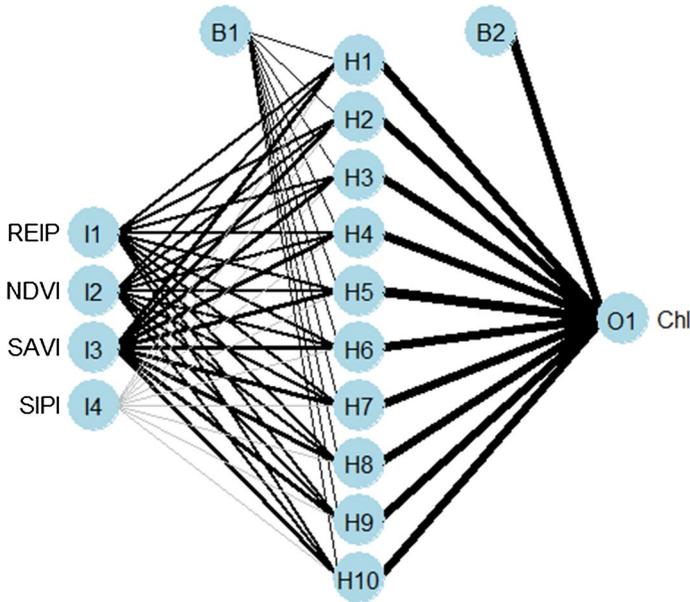


Fig. 2 Schematic representation neural network employed in this study, where *REIP* Red Edge Inflection Point, *NDVI* Normalized Difference Vegetation Index, *SAVI* Soil-Adjusted Vegetation Index, *SIPI* Structure-Insensitive Pigment Index, and *Chl* total chlorophyll concentration. *H* represents hidden layer, *B1* and *B2* represent bias neurons, *I* represent input payers, and *O* represents the output layers

$$O_a = h_{\text{hidden}} \left(\sum_{p=1}^P i_{a,p} w_{a,p} + b_a \right) \tag{7}$$

where $h_{\text{hidden}}(x) = \frac{1}{1+e^{-x}}$, P =node numbers, $w_{a,p}$ = the weight, $i_{a,p}$ = an input to node a from network input p , O_a = the current hidden layer node a output, and b_a is the bias (Aleksander and Morton 1990).

The sigmoid activation function is used with ANNs. In Eq. 7, $h_{\text{hidden}}(x)$ is the sigmoid activation function. Normalization of the training data is essential to avoid saturating the activation function, which was performed mainly to limit the ranges at -1 to 1 . The shape of this function plays an important role in ANNs learning. The following normalization equation was used (Zhang et al. 1998):

$$z_{\text{norm}} = \frac{z_o - \bar{z}}{z_{\text{max}} - z_{\text{min}}} \tag{8}$$

where Z_{norm} = normalized value; z_0 = original value; \bar{z} = mean; Z_{max} = maximum value; and Z_{min} = minimum value.

This network architecture consists of hidden, input and output layers, one or more neurons containing with each layer also the hidden and output layers connected with bias neurons. The present study illustrates a method sufficient to describe the classic family of 10-hidden-layer, four-input-layer and one-output-layer (4-10-1) neural network trained by the back-propagation algorithm (Rumelhart et al. 1986). These neural

networks are proposed to be a global predictor of every unremitting function (Hornik et al. 1989) and therefore can be incorporated in the research related to ecological studies.

2.4 Sensitivity analysis framework

SA is executed to analyse the relative significance of informative variables (Petrooulos and Srivastava 2016). It helps in explaining the relationship between informative and response or reactive variable by the model. Conversely, sensitivity analysis does not provide the categorical description of a relationship between variable but preferably allows its users to predict the nature of the association between variables. We may assume that the linkage between a reaction and illustrative variable may vary, provided the setting of another informative factors (i.e. a cooperation might be available). More or less sensitivity examination can provide these data.

In this study, the Lek's profile method is employed for sensitivity analysis. In 1996, the "Lek-profile method" Sovan Lek et al. (1996) was formulated, which was later explained in detail by Gevrey et al. (2003). This method produces profile (or contribution) plots of each output variable with respect to a range of one input variable wherein the rest of input variables are kept constant at their 0th, 20th, 40th, 60th, 80th and 100th percentiles. Subsequently, this method repeats the above-mentioned processes for each input variable. As a result, it generates response curves according to the change of the input variables (Gevrey et al. 2003). Figure 3 illustrates information regarding range of input and output variables along with the five different percentile values. For statistical model, utilizing various response factors identified with different explanatory factors, the user picks one of the two explanatory variables to simulate the responses of factors over the scope of qualities set for the explanatory variable.

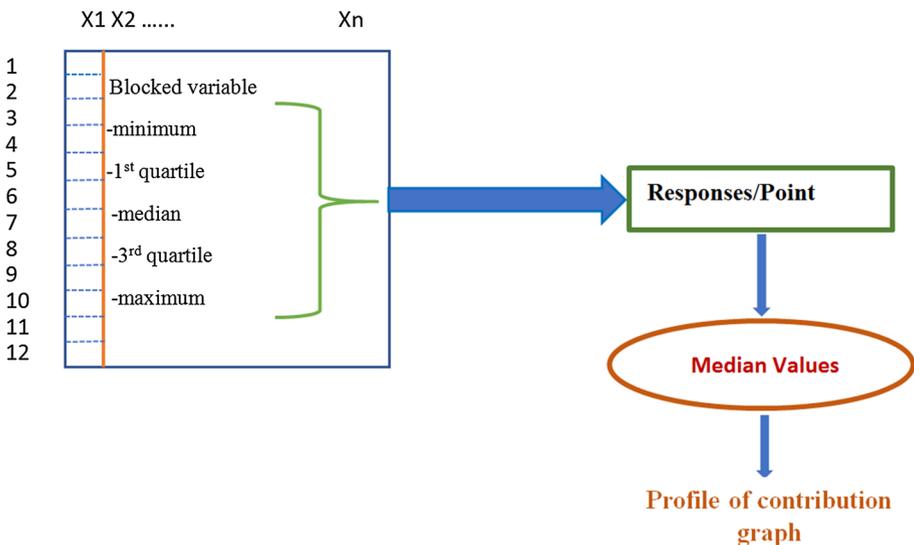


Fig. 3 Explanatory schema of the Lek's profile method employed in the present study

2.5 Performance evaluation

In this study, we compared the SA-ANNs simulated results with the Total Chlorophyll (T_Ch). The performances between the two methods are evaluated regarding the agreement index and RMSE. The ratio of the potential error and mean square error represents the index of agreement (Willmott et al. 1985) and is defined as mentioned below (Eq. 9):

$$d = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (|y_i - \bar{x}| + |x_i - \bar{x}|)^2} \quad (9)$$

where n = observation numbers; x_i = station measured data and y_i = modelled data; and \bar{x} = mean of station measured data (Willmott et al. 1985).

The RMSE can be expressed as shown below (Eq. 10):

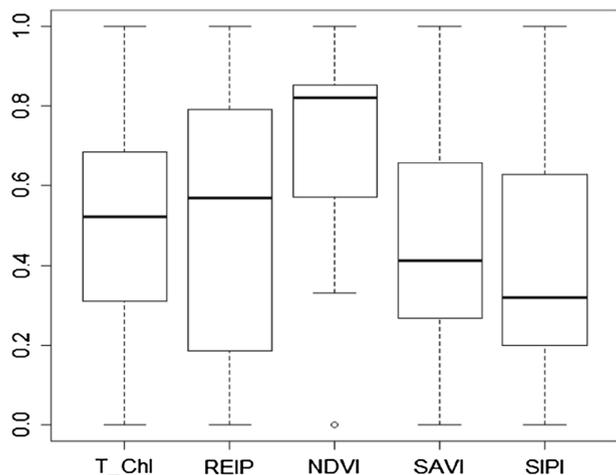
$$\text{RMSE} = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n [y_i - x_i]^2 \right)} \quad (10)$$

3 Results and discussion

3.1 Evaluation of input dataset

To estimate the distribution of REIP, NDVI, SAVI, SIPI and T_Ch, the box and whisker plot is utilized, which is clear, definite and simple to peruse the data as illustrated in Fig. 4. In box and whisker plot, the line indicates the median value across the box, the whiskers by the highest degree and the box is enclosed by the quartiles. The long whiskers within the container (box) plot suggest that the implicit distribution is skewed towards excessive estimates and vice versa for least whiskers. Box plots with a massive spread show high variations in the datasets. The plot illustrates lower inter-quartile range for NDVI and SAVI.

Fig. 4 Box and whisker plot of T_Ch, REIP, NDVI, SAVI and SIPI, showing mean, quartile, minimum, maximum and median values of the vegetation indices



REIP represents the highest inter-quartile ranges followed by T_ChI. The maximum frequent values were skewed in the direction of lower median values in almost all the cases. In a case of NDVI and REIP, a high spread was obtained followed by T_ChI, SAVI and SIPI. Except NDVI, for all other cases profoundly high values for minimum and maximum values were found in the dataset. The numbers of outliers were found higher in the case of NDVI as compared to other parameters, which indicates some data points were not following the normal trend of the datasets. This shows that the high NDVI values conditions are may be due to hydrogen intensification in the soil by microbial activities.

The performance of various methods in the terms of correlations can be represented by scatter matrix plot as shown in Fig. 5. In the plots, the correlation between T_ChI and REIP [both r (0.87) and r_s (0.82) values] revealed comparable performances, while between T_ChI and NDVI a linear relationship can be seen with r and r_s values of 0.84 and 0.73 respectively. The T_ChI showed a negative correlation with SIPI with $r = -0.86$ and $r_s = -0.82$. However, T_ChI showed a strong correlation with SAVI with $r = 0.94$ and $r_s = 0.92$. REIP showed a linear relationship with SAVI ($r = 0.78$, $r_s = 0.73$) and NDVI ($r = 0.81$, $r_s = 0.72$) but a negative relationship with SIPI with $r = -0.8$ and

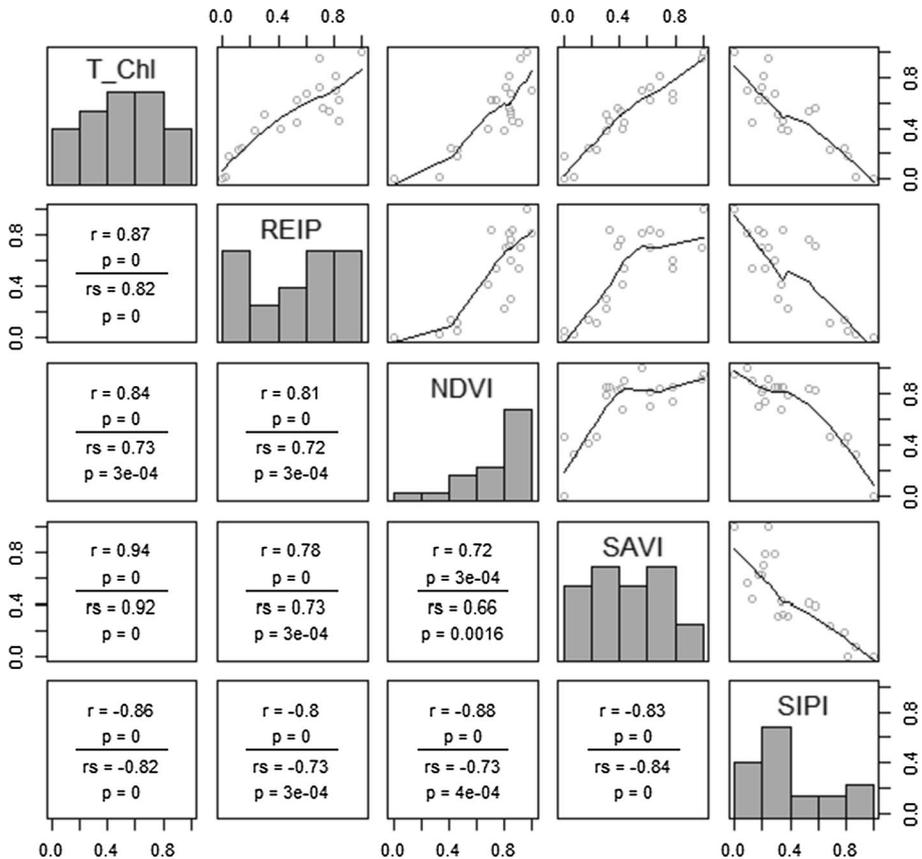


Fig. 5 Pearson and Spearman correlation matrix plot illustrating the relationship among factors (T_ChI, REIP, NDVI, SAVI and SIPI) (where p represents the relationship significance in terms of probability levels)

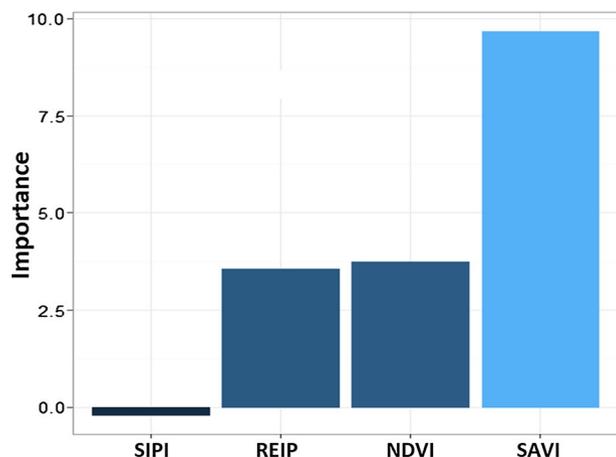
$r_s = -0.73$. SAVI showed a weak correlation with NDVI ($r = 0.72$, $r_s = 0.0016$) and a negative trend with SIPI ($r = -0.83$, $r_s = -0.84$). NDVI and SIPI also showed a negative relation with $r = -0.88$ and $r_s = -0.84$.

3.2 Evaluation of variable importance

For evaluating the variable importance, the Olden function is a more flexible and robust method. This function distinguishes the relative significance of descriptive variable for a solitary reaction variable by considering the model weights. In the network, all weighted links between the layers can be used to determine each variable's importance. All the weight response variables were recognized and pass through the hidden layer. This is recurring process for all the other descriptive variables. This process is repeated continuously until specific input variable is obtained from a list of weights. The connections for every input node are scaled relative to all other inputs that are tallied. In the model, single value describes the relationship with the response variable. This single value is obtained for each descriptive variable. For quantification of variable contributions, the product of the input-hidden and hidden-output connection weights between each input/output neurons are summed across all hidden neurons. The importance of this method maintains the relative contribution of each connection in both terms as the sign and as the magnitude. The other advantage of the Olden's algorithm that it is proficient for evaluating neural network with response variable and multiple hidden layers. Each variable is assigned the importance value in units. Each assigned variable is directly based on the product summation of the connection weight parameters. Based on a magnitude and relative sign between explanatory variable, the actual value can be interpreted.

Independent parameters such as NDVI, SIPI, SAVI and REIP were considered for their importance in T_{Chl} prediction using Olden's method. To assess the sensitivity, each of the measured variables was assessed for its utility for T_{Chl} prediction. The analysis demonstrated that for T_{Chl} prediction, the most sensitive parameters is SAVI followed by NDVI and REIP, while least sensitivity can be detected in case of SIPI (as illustrated in Fig. 6).

Fig. 6 Variable importance estimation using Olden's method for all the vegetation indices employed in this study



3.3 Sensitivity analysis of input variables using Lek's profile method

The SA method was proposed by Lek et al. (1995, 1996). The principle of Lek profile algorithm is to create an unreal matrix be relevant to the range of all input variables. The generated matrix consists of explanatory variables values where rows symbolize observations numbers and columns represent explanatory variables. This data frame or matrix is used to anticipate the value of response variable through fitted object model. The variable of this data frame or matrix is divided into several intervals between minimum values and maximum values. These intervals are called scale. The value of interval from maximum to minimum show median, third quartile and maximum. For the study of this variable, value for each of the scale point is required. After that, the profile of the outcome variable can be plotted and the same process or calculation are provided for the other variables. A curve is obtained for each variable, then according to the input variable, it gives a set of profiles of the variables. This gives a final product that is a set of response curves. The calculation of relative significance for respective input variables is illustrated by the magnitude of its range to predict response value, i.e. maximum to minimum. This process is called "profile" method by Gevrey et al. (2003) and the plots display the bivariate nature in one explanatory variable and one response variable. The several lines in plots show the relationship variation, and the explanatory variable held constant to their minimum 20th, 40th, 60th, 80th and the maximum quartile values. Influence of four independent environment variables on T_Ch1 in the ANNs can be illustrated by the six curves (as shown in Fig. 7). After the evaluation of each independent parameter—NDVI, SIPI, SAVI and REIP, the T_Ch1 is predicted. The concentration of T_Ch1 is the minimum at the low value of the independent variable, and then, it enhances very rapidly to reach the maximum level in the case of SAVI, which demonstrated that for T_Ch1 prediction, SAVI, NDVI and REIP are the most sensitive parameters as compared to SIPI.

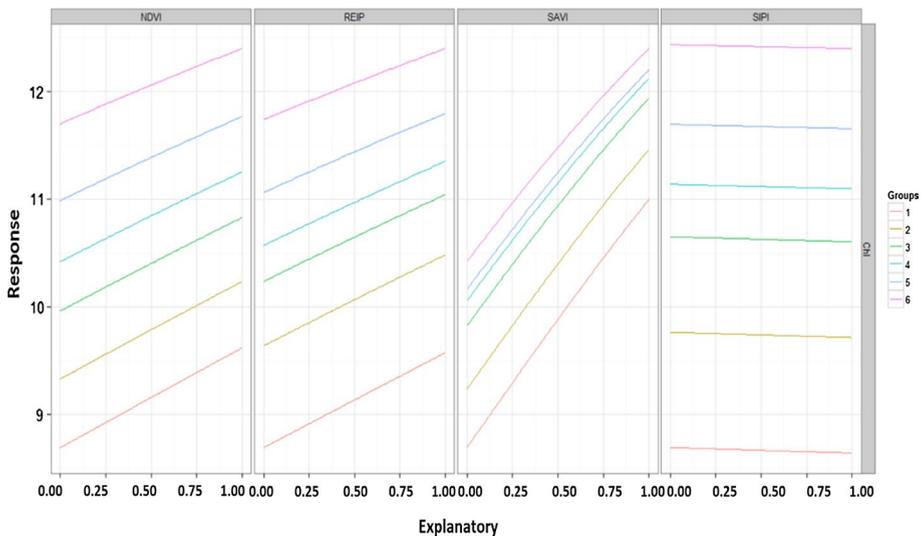


Fig. 7 Lek's "profile" method for sensitivity analysis of ANNs with the input variables NDVI, SIPI, SAVI and REIP for T_Ch1 prediction

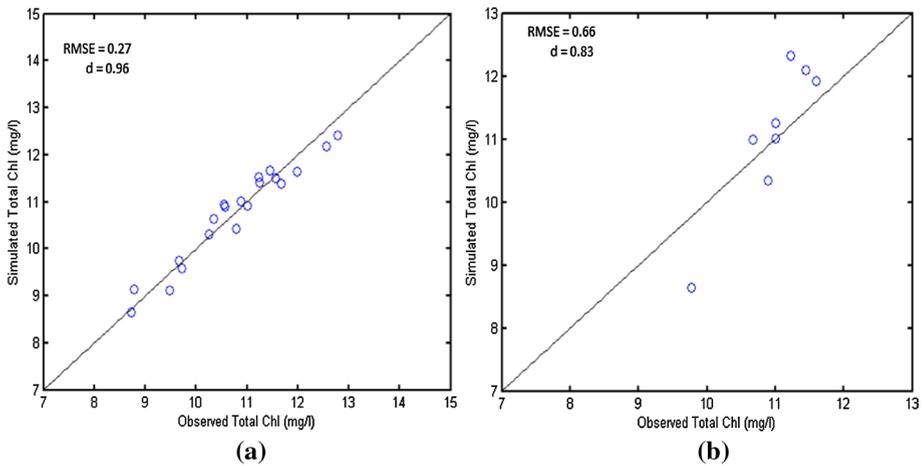


Fig. 8 Scatter plot of observed and simulated total chlorophyll during a) calibration and b) validation

The scatter plots during calibration and validation of ANN are shown in Fig. 8. For prediction of T_Chlorophyll using ANNs, the size of network and decay functions for an optimal performance were obtained as 10 and 0.1, respectively. The goodness of fit of measured and simulated total chlorophyll is represented using the RMSE and index of agreement. The statistical performance indices were obtained as (RMSE = 0.27; index of agreement = 0.96) during the calibration and validation (RMSE = 0.66; index of agreement = 0.83), suggesting that the model is appropriate for chlorophyll prediction with high efficiency and thus can be used for chlorophyll estimation. Similar results of high ANN performance were also obtained by Suo et al. (2010), in which the authors used ANN for chlorophyll prediction in cotton plant. Results obtained through their method were in sound agreement with those obtained in the laboratory. In study by Broge and Mortensen (2002), it was shown that vegetation indices are needed for canopy leaf chlorophyll content of winter wheat crops, as they reduce spectral effects caused by external factors such as the atmosphere and the soil background. In research by Tian et al. (2017), traditional ANN model for chlorophyll prediction was presented. They showed that proper optimization of ANN is needed for chlorophyll estimation, and they concluded that the optimized ANN is showing better results than traditional one. Further, in order to reduce the dimensionality, the PCA method was utilized by Zhou et al. (2015) with hyperspectral datasets. Their results showed that the PCA, when combined with ANN model, can improve the accuracy. In overall, the results produced by this paper are in agreement with the studies as mentioned above.

4 Conclusion

The study evaluated the ANNs for chlorophyll prediction by providing appropriate SA for the indices developed using the hyperspectral datasets. For the first time, the Lek's profile method has been used for T_Chlorophyll prediction. The Lek's profile method not only classifies the input variables by evaluating relative importance but also defines how these inputs contribute to the results. The techniques adopted in the present research illustrate that the maximum information of plant biochemical variables, i.e. T_Chlorophyll content, can be obtained

using the hyperspectral radiometer datasets. The hyperspectral radiometer measurements were then used for the generation of indices like NDVI, REIP, SAVI and SIPI. The results obtained from Lek's profile method revealed that the SIPI is the least sensitive parameter for T_Ch1 as compared to the other indices—SAVI, NDVI and REIP. SAVI was found to be the most sensitive parameter for T_Ch1 and should be carefully estimated from radiometer datasets. The SA-ANN approach provided through this work can be used as an effective method for chlorophyll estimation from hyperspectral sensor and therefore reduce our dependency on destructive techniques. The method is useful for an effective crop management and forecasting as well as significant in providing quality control information to estimate uncertainty and their associated assumptions. This research work can be implemented to estimate the biochemical parameters in conjunction with the upcoming missions such as Airborne Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) and Hyperspectral Infrared Imager (HypIRI). In future, vegetation indices that include water content, leaf structure and others will be incorporated for T_Ch1 prediction. Additionally, in future, attempts will be made to evaluate the sensitivity of other artificial intelligence techniques like support vector machines and relevance vector machines for crop biochemical parameters estimation.

Acknowledgements Authors would like to thank University Grant Commission and NMHS, MOEF and CC, Government of India, for providing Funds for this research. Authors would also like to thank Department of Civil Engineering, IIT Roorkee, for providing the necessary instrumentation for this research.

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

References

- Aleksander, I., & Morton, H. (1990). *An introduction to neural computing*. London: Chapman & Hall.
- Anand, A., et al. (2020). Use of hyperion for mangrove forest carbon stock assessment in Bhitarkanika forest reserve: A contribution towards blue carbon initiative. *Remote Sensing*, 12(4), 597.
- Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. *Plant Physiology*, 24(1), 1.
- Asrar, G., Fuchs, M., Kanemasu, E., & Hatfield, J. (1984). Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal*, 76(2), 300–306.
- Barton, C. V. M., & North, P. (2001). Remote sensing of canopy light use efficiency using the photochemical reflectance index: Model and sensitivity analysis. *Remote Sensing of Environment*, 78(3), 264–273.
- Becker, F., & Choudhury, B. J. (1988). Relative sensitivity of normalized difference vegetation index (NDVI) and microwave polarization difference index (MPDI) for vegetation and desertification monitoring. *Remote Sensing of Environment*, 24(2), 297–311.
- Ben-Dor, E., Schläpfer, D., Plaza, A. J., & Malthus, T. (2013). Hyperspectral remote sensing. Airborne measurements for environmental research: Methods and instruments (pp. 413–456)
- Bharose, R., Lal, S., Singh, S., & Srivastava, P. (2013). Heavy metals pollution in soil–water–vegetation continuum irrigated with ground water and untreated sewage. *Bulletin of Environmental and Scientific Research*, 2(1), 1–8.
- Blackburn, G. A. (1998). Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66(3), 273–285.
- Blackburn, G. A. (2006). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4), 855–867.
- Broge, N. H., & Mortensen, J. V. (2002). Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote Sensing of Environment*, 81(1), 45–57.

- CÁrcamo, H. A., & Spence, J. R. (1994). Crop type effects on the activity and distribution of ground beetles (Coleoptera: Carabidae). *Environmental Entomology*, 23(3), 684–692.
- Cho, M. A., Skidmore, A., Corsi, F., Van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9(4), 414–424.
- Clevers, J., De Jong, S., Epema, G., Van Der Meer, F., Bakker, W., Skidmore, A., et al. (2002). Derivation of the red edge index using the MERIS standard band setting. *International Journal of Remote Sensing*, 23(16), 3169–3184.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3), 249–264.
- Goel, P. K., Prasher, S. O., Landry, J. A., Patel, R. M., Viau, A. A., & Miller, J. R. (2003). Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing. *Transactions of the ASAE*, 46(4), 1235.
- Gupta, H. V., Beven, K. J., & Wagener, T. (2006). Model calibration and uncertainty estimation. Part 11: Rainfall Run-off modelling. In M. G. Anderson (Ed.), *Encyclopedia of hydrological sciences* (pp. 1–17). Wiley. <https://doi.org/10.1002/0470848944.hsa138>.
- Gupta, M., Srivastava, P. K., Mukherjee, S., & Kiran, G. S. (2014). Chlorophyll retrieval using ground based hyperspectral data from a tropical area of India using regression algorithms. In P. Srivastava, S. Mukherjee, M. Gupta & T. Islam (Eds.), *Remote sensing applications in environmental research. Society of earth scientists series* (pp. 177–194). Springer. https://doi.org/10.1007/978-3-319-05906-8_10.
- Gustafson, P., Srinivasan, C., & Wasserman, L. (1996). Local sensitivity analysis. *Bayesian Statistics*, 5, 197–210.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309.
- Ireland, G., Petropoulos, G. P., Carlson, T. N., & Purdy, S. (2015). Addressing the ability of a land biosphere model to predict key biophysical vegetation characterisation parameters with global sensitivity analysis. *Environmental Modelling & Software*, 65, 94–107.
- Kisi, O., & Shiri, J. (2011). Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resources Management*, 25(13), 3135–3152.
- Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2002). Imaging spectrometry and vegetation science. In *Imaging spectrometry* (pp. 111–155). Springer.
- Lamine, S., Petropoulos, G. P., Brewer, P. A., Bachari, N.-E.-I., Srivastava, P. K., Manevski, K., et al. (2019). Heavy metal soil contamination detection using combined geochemistry and field spectroradiometry in the United Kingdom. *Sensors*, 19(4), 762.
- Lamine, S., Petropoulos, G. P., Singh, S. K., Szabó, S., Bachari, N. E. I., Srivastava, P. K., et al. (2018). Quantifying land use/land cover spatio-temporal landscape pattern dynamics from Hyperion using SVMs classifier and FRAGSTATS®. *Geocarto International*, 33(8), 862–878.
- Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., & Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*, 46(8), 1229–1236.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., & Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, 90(1), 39–52.
- Malhi, R. K. M., Anand, A., Mudaliar, A. N., Pandey, P. C., Srivastava, P. K., & Sandhya Kiran, G. (2020). Synergetic use of in situ and hyperspectral data for mapping species diversity and above ground biomass in Shoolpaneshwar Wildlife Sanctuary, Gujarat. *Tropical Ecology*, 61, 106–115. <https://doi.org/10.1007/s42965-020-00068-8>.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371.
- Nagy, H., Watanabe, K., & Hirano, M. (2002). Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering*, 128(6), 588–595.
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1), 135–150.
- Ortenberg, F., Thenkabail, P., Lyon, J., & Huete, A. (2011). *Hyperspectral sensor characteristics: Airborne, spaceborne, hand-held, and truck-mounted; integration of hyperspectral data with Lidar*. Boca Raton, FL: CRC Press.

- Pandey, P. C., Anand, A., & Srivastava, P. K. (2019). Spatial distribution of mangrove forest species and biomass assessment using field inventory and earth observation hyperspectral data. *Biodiversity and Conservation*, 28(8–9), 2143–2162. <https://doi.org/10.1007/s10531-019-01698-8>.
- Pandey, P. C., Tate, N. J., & Balzter, H. (2014). Mapping tree species in coastal Portugal using statistically segmented principal component analysis and other methods. *IEEE Sensors Journal*, 14(12), 4434–4441. <https://doi.org/10.1109/JSEN.2014.2335612>.
- Pandey, P., Manevski, K., Srivastava, P. K., & Petropoulos, G. P. (2018). The use of hyperspectral earth observation data for land use/cover classification: Present status, challenges and future outlook. In P. Thenkabail (Ed.), *Hyperspectral remote sensing of vegetation* (1st edn, pp. 147–173).
- Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2), 221–230.
- Petropoulos, G. P., Arvanitis, K., & Sigrimis, N. (2012a). Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3), 3800–3809. <https://doi.org/10.1016/j.eswa.2011.09.083>.
- Petropoulos, G. P., Griffiths, H. M., Ioannou-Katidis, P., & Srivastava, P. K. (2014). Sensitivity exploration of SimSphere land surface model towards its use for operational products development from Earth observation data. In P. Srivastava, S. Mukherjee, M. Gupta & T. Islam (Eds.), *Remote sensing applications in environmental research. Society of earth scientists series* (pp. 35–56). Springer.
- Petropoulos, G. P., Kalaitzidis, C., & Vadrevu, K. P. (2012b). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, 41, 99–107.
- Petropoulos, G. P., Kalivas, D. P., Georgopoulou, I. A., & Srivastava, P. K. (2015). Urban vegetation cover extraction from hyperspectral imagery and geographic information system spatial analysis techniques: Case of Athens, Greece. *Journal of Applied Remote Sensing*, 9(1), 096088.
- Quarmby, N., Milnes, M., Hindle, T., & Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing*, 14(2), 199–210.
- Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition. Volume 1—Foundations*. Cambridge, MA: The MIT Press.
- Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis*. New York: Wiley.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis: The primer*. Hoboken: Wiley.
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Hoboken: Wiley.
- Schaap, M. G., Leij, F. J., & Van Genuchten, M. T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal*, 62(4), 847–855.
- Singh, S. K., Srivastava, P. K., Szabó, S., Petropoulos, G. P., Gupta, M., & Islam, T. (2017). Landscape transform and spatial metrics for mapping spatiotemporal land cover dynamics using Earth Observation data-sets. *Geocarto International*, 32(2), 113–127.
- Smith, M. (1992). *CROPWAT: A computer program for irrigation planning and management*. Rome: Food & Agriculture Org.
- Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6), 568–576.
- Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Al-Shrafany, D., & Islam, T. (2013). Data fusion techniques for improving soil moisture deficit using SMOS satellite and WRF-NOAH land surface model. *Water Resources Management*, 27(15), 5069–5087.
- Srivastava, P. K., Han, D., Rico-Ramirez, M. A., & Islam, T. (2014). Sensitivity and uncertainty analysis of mesoscale model downscaled hydro-meteorological variables for discharge prediction. *Hydrological Processes*, 28(15), 4419–4432.
- Suo, X.-M., Jiang, Y.-T., Mei, Y., Li, S.-K., Wang, K.-R., & Wang, C.-T. (2010). Artificial neural network to predict leaf population chlorophyll content from cotton plant images. *Agricultural Sciences in China*, 9(1), 38–45.
- Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71(2), 158–182.
- Tian, W., Liao, Z., & Zhang, J. (2017). An optimization of artificial neural network model for predicting chlorophyll dynamics. *Ecological Modelling*, 364, 42–52.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150.
- Tzeng, F.-Y., & Ma, K.-L. (2005). Opening the black box-data driven visualization of neural networks. In *Visualization, 2005. VIS 05. IEEE*. IEEE (pp. 383–390).

- Verrelst, J., Rivera, J. P., & Moreno, J. (2015). ARTMO's global sensitivity analysis (GSA) toolbox to quantify driving variables of leaf and canopy radiative transfer models. *EARSeL eProc*, 14, 1–11.
- Wagner, T., McIntyre, N., Lees, M., Wheeler, H., & Gupta, H. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, 17(2), 455–476.
- Wei, F., Yan, Z., Yongchao, T., Weixing, C., Xia, Y., & Yingxue, L. (2008). Monitoring leaf nitrogen accumulation in wheat with hyper-spectral remote sensing. *Acta Ecologica Sinica*, 28(1), 23–32.
- Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., et al. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5), 8995–9005.
- Wu, C., & Chau, K. (2011). Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology*, 399(3), 394–409.
- Wu, C., Chau, K., & Fan, C. (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, 389(1), 146–167.
- Zhang, G., Eddy Patuwo, B., & Hu, Y. M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62.
- Zhou, L., Ma, W., Zhang, H., Li, L., & Tang, L. (2015). Developing a PCA–ANN model for predicting chlorophyll a concentration from field hyperspectral measurements in Dianshan Lake, China. *Water Quality, Exposure and Health*, 7(4), 591–602.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Prashant K. Srivastava^{1,2}  · Manika Gupta³ · Ujjwal Singh¹ · Rajendra Prasad⁴ · Prem Chandra Pandey⁵ · A. S. Raghubanshi¹ · George P. Petropoulos⁶

Prem Chandra Pandey
pre26bit@gmail.com

- ¹ Remote Sensing Laboratory, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India
- ² DST-Mahamana Centre for Excellence in Climate Change Research, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India
- ³ Department of Geology, University of Delhi, New Delhi, India
- ⁴ Department of Physics, IIT (BHU), Varanasi, India
- ⁵ Center for Environmental Sciences and Engineering, School of Natural Sciences, Shiv Nadar University, Greater Noida, Uttar Pradesh, India
- ⁶ Department of Geography, Harokopio University of Athens, El. Venizelou 70, Kallithea, 17671 Athens, Greece