Geocarto International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

Development of hyperspectral indices for anti-cancerous Taxol content estimation in the Himalayan region

Ayushi Gupta, Prachi Singh, Prashant K. Srivastava, Manish K. Pandey, Akash Anand, K. S. Chandra Sekar & Karuna Shanker

To cite this article: Ayushi Gupta, Prachi Singh, Prashant K. Srivastava, Manish K. Pandey, Akash Anand, K. S. Chandra Sekar & Karuna Shanker (2021): Development of hyperspectral indices for anti-cancerous Taxol content estimation in the Himalayan region, Geocarto International, DOI: 10.1080/10106049.2021.1983031

To link to this article: https://doi.org/10.1080/10106049.2021.1983031

Development of hyperspectral indices for anti-cancerous Taxol content estimation in the Himalayan region

Ayushi Gupta[1], Prachi Singh[1], Prashant K. Srivastava[1,2*], Manish K. Pandey[1], Akash Anand[1], K. S. Chandra Sekar[3], and Karuna Shanker[4]

[1]Remote Sensing Laboratory, Institute of Environment and Sustainable Development, Banaras Hindu University, U.P

[2]DST-Mahamana Centre for Excellence in Climate Change Research, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India

[3]G.B. Pant, National Institute of Himalayan Environment (NIHE), Kosi-Katarmal, Almora, Uttarakhand, India

[4]CSIR - Central Institute of Medicinal and Aromatic Plants, Lucknow, India

[*]Corresponding Author: prashant.iesd@bhu.ac.in

Abstract

Monitoring and management of rare and economically important species in the highly complex terrain are challenging and thus need advanced technological development. In this study, the hyperspectral radiometer data of *Taxus wallichiana* were acquired at highly complex terrain of the Pindari region of the Himalaya and processed by using several sophisticated algorithms to deduce Taxol content in the plants. The spectroradiometer data were denoised through three different types of smoothing filters such as Average Mean, Savitzky Golay, and Fast Fourier Transform (FFT) followed by feature selection for allocation of best bands for Taxol content estimation. The results showed that the Average Mean filter in combination with feature selection performed best for Taxol spectral indices generation, model development, and Taxol content prediction. The best model showed a correlation of 0.719 with a relative root mean square error (RMSEr) value of 0.678 for Taxol content prediction.

Keywords Taxol spectral indices; Taxol model development; hyperspectral data; smoothening and filtering; *Taxus wallichiana*

1. Introduction

According to WHO (World Health Organization) estimates, 80% of the population worldwide count on herbal medicines for some aspect or the other for their primary health care needs. Approximately 2/3rd of the plants accounted in the modern medical system found their health care origin in Asian countries. Apart from the rural population depending on indigenous systems of medicine (Ekor 2014) even modern medicine derives its inspiration from the indigenous medicinal system (Yuan *et al.* 2016). Many researchers have also emphasized that modern medicine should take the precious experience of natural products and traditional medicine (Pan *et al.* 2014, Yuan *et al.* 2016). In India, around 30% household uses traditional medicine (Srinivasan *et al.* 2017), which is plentiful in India.

Medicinal plants contain phytopigments and bioactive compounds which contribute to their physiological function and medicinal properties (Mohamed *et al.* 2010). The most important phytochemical components that are responsible for medicinal properties for any plant are alkaloids,

tannins, flavonoids, and phenolic compounds (Geetha and Geetha 2014). One of the major plants i.e., *Taxus wallichiana* Zucc. (*T. wallichiana*) also known as Himalayan yew is found very promising for cancer treatment. The needles/leaves of the *T.wallichiana* is one of the valuable sources of taxoid (Appendino *et al.* 1992, Bala *et al.* 1999) Paclitaxel (trade name Taxol) is a tricyclic diterpenoid (alkaloid) and considered as an efficient anti-cancerous drug (Zhu *et al.* 2019). The extraction and refinement of Taxol are time-consuming, difficult, expensive, and tedious because of the low yields (van Rozendaal *et al.* 2000). The taxanes are isolated from the Taxus plant material by complex extraction procedures and analyzed using sophisticated HPLC–UV or LC-MS methods (Fu *et al.* 2009, Chakchak and Zineddine 2013, Sadeghi-Aliabadi *et al.* 2015). However, the population of these species has seen a large reduction due to its excessive demand and collection of this anti-tumor and anti-cancerous drug. A study conducted exhibited that these trees were spoiled due to bark-stripping practices. Moreover, these species are very slow-growing (Suffness 1995). Hence, management of this important resource at a larger scale becomes necessary which can be achieved using remote sensing.

The identification and differentiation among various medicinally important species using remote sensing are often limited, by the ability of spectral variance, which can discriminate the minute spectral differences among species (Clark *et al.* 2005). Hyperspectral remote sensing can serve as a suitable solution since it contains several (mainly between 64 and 256) contiguous fine resolution bands with a bandwidth of 1 to 10 nm, providing noteworthy levels of subtle feature differences to provide fine spectral variations among tree species (Peerbhay *et al.* 2013, Srivastava *et al.* 2020b). These fine spectral bands data are characteristically high-dimensional but also found to be highly correlated with the vegetation parameters tested (Landgrebe 2002). Recently plant phenolics were also characterized in vegetation reflectance at 1660 nm (Kokaly *et al.* 2015). Clearly, retrieval of such canopy information via remotely sensed data involves analytical means which are proficient in translating the spectral response data into practical information.

The high dimensionality of these datasets also cause many problems such as heavy computational processor requirements and high data storage cost. To process high-dimensional data efficiently, dimensionality reduction (DR) becomes essential. Band Selection (BS) is one of the techniques of DR that selects a subset of bands preserving their physical meaningful data with the benefit of keeping intact the relevant original information in the data (Srivastava *et al.* 2020a). The BS method derivative analysis is an amalgamation of a variance-based and shape-based (derivative) approach for feature identification (Tsai and Philpot 1998b) having better separability (larger Jefferies-Matsushita (JM) distances) to differentiate among groups. The apprehension of data dimensions and training data sizes in hyperspectral emphasizes the need to compress valuable information into the least number of bands (Ling *et al.* 2019, Singh *et al.* 2020a).

Uncertainties are often seen in spectral datasets caused due to atmospheric disturbances. Hence, for optimal band selection, smoothening (data pre-processing) becomes a part of the derivative analysis application (Torrecilla *et al.* 2009). Many studies have modified and reviewed these filtering and smoothening techniques to develop a set of cross-platform tools for the analysis of spectral data (Tsai *et al.* 2002). Comparatively few scholars have taken the derivative approach for the analysis of hyperspectral data used in remote sensing due to its limitation (Torrecilla *et al.* 2009). The regression-based models on spectral indices are characteristically empirical formulae aiding the plotting of several biochemical parameters consequential from remotely sensed data. Since it is empirical in nature, it remains undefined to up to what extent this selected regression model works well, till all the band combinations and curve-fitting functions are evaluated (Rivera *et al.* 2014b, Pandey *et al.* 2019).

According to the study of J. P. Rivera et. al (Rivera *et al.* 2014a) many hyperspectral indices have been tested for the retrieval of LAI and Chlorophyll using HyMap sensor data during the

SPARC-2003 campaign in Barrax, Spain. 12 chlorophyll spectral indices for chlorophyll inversion have been developed using ASD spectroradiometer data such as Vogelmann red edge index, Zarco-Tejada, Miller index (ZMI), modified normalized difference vegetation index (mNDVI), modified normalized difference index (mND) etc (Lin *et al.* 2012), (Zagajewski *et al.* 2018). The vast majority of these SIs and their association with desired parameters have been established through experimental work. According to the above studies development of indices can be a successful approach for the retrieval of the biochemical parameters using hyperspectral data. These studies are based on the parametric regression approach (Gupta *et al.* 2014, Verrelst *et al.* 2019). Hence expanding the knowledge of the Hyperspectral for sophisticated biochemical parameters estimation becomes the next logical step in this direction. In the purview of the above, the main aims of this study are 1) Estimation of alkaloid Taxol and reporting its concentration in the Pindari region of Himalaya. 2) Development of indices sensitive for Taxol content estimation through denoised hyperspectral data. 3) Development of robust assessment method to evaluate various indices, spectral band settings, and curve-fitting functions for retrieval of Taxol content in the Himalayan region.

2. Material and Methods

2.1. Study Area

The Nanda Devi biosphere reserve is in Chamoli, Pithoragarh, and Bageshwar districts of the state of Uttarakhand is located in Western Himalayas lying in Highland Biogeographic Zone (2a). The climatic year of the Nanda Devi biosphere reserve has been distinguished into three seasons mainly- summer (April- June), rainy season (June-September), and winter (October-March). The average annual rainfall is 930 mm, out of which 48% occurs in two months (July-August). The maximum temperature range varies between 11 to 24°C and the minimum temperature varies between 3 to 7°C. The present study site selected is rich in medicinal herbs as well as trees. *T. wallichiana* is one of the species which is prominent and highly medicinal in nature. The medicinal compound is one of the major reasons for the plant`s declining population. The sampling was done in the rainy season as it is the most favorable season for plant growth. Broadly the area is divided into two climatic zones that could be categorized as (i) Lower montane zone: elevation range of 1800-2400 m above mean sea level (amsl), (ii) Upper montane zone: elevation range of 2400- 3000 m amsl. The precipitation is more in the upper zone is more in terms of snowfall than showers (Gaur, 1999). The samples of *Taxus wallichiana* were collected are as shown in figure 1**.**

2.2. Sample collection and analysis

Ground sample was collected in the Pindari region of Himalaya during the dates 26/09/2019 and 29/09/2019 at different locations with a varying altitude of 2292-3039 m in the Nanda Devi Biosphere Reserve (NDBR) as shown in figure 1 along with hyperspectral radiometer measurements. The leaves of the plant sample were collected from two to three locations from the same tree to make the leaf sample homogenous for each located tree of *T. wallichiana*. The plant samples collected over the NDBR were of the same phenological stage. The spectra of these leaves were recorded using a handheld ASD spectroradiometer. The properties of the selected plants are displayed in table 1. The samples were then were crushed in liquid nitrogen for further analysis. For Taxol extraction, 1 g of crushed leaves was deflated with hexane using sonication. The deflated samples were filtered and then percolated using 25 ml methanol, each time repeated thrice using sonication. The hexane portions were rejected and methanol aliquots were collected together and then concentrated using a rotary evaporator. The samples were extracted in distilled water (50 ml). For the chloroform partitioning, the extracted water sample was then successively extracted by the solvent extraction process five times with 50 ml of chloroform each time. The chloroform extracted sample was then

pooled together (250 ml) for each sample and dried under reduced pressure using a rotary evaporator, then re-dissolved in methanol (1 ml) (Shanker *et al.* 2008).

The liquid chromatography was done at room temperature on a Symmetry® C18, (both 250 mm × 4.6 mm i.d, 5.0 µm particle size) with an ultraviolet-diode array detector (UV-DAD). Chromatographic surroundings were augmented by regulating the composition and potential of hydrogen (pH) of the mobile phase for replicable results. The chromatographic solvents used for isocratic runs were: (a) Methanol and (b) Water (0.05% Acetic Acid) (62:38, v/v). The flow rate for the mobile phase was 1.0 mL min^{-1}. The working solution of paclitaxel was prepared from standard using methanol. Insertions of samples were done using a sample injector of a 20 µL loop. The UV-DAD scanned acquisitions of Taxol at 230 nm. The percentage of Taxol was calculated using equation (1) (Shanker *et al.* 2008).

$$\text{Taxol Content (\%)} = \frac{\text{Ar.}_{\text{sample}} * \text{Conc.}_{\text{std.}} \left({}^{\text{mg}}/_{\text{ml}}\right)}{\text{Ar.}_{\text{std.}} * 1000 * \text{Conc.}_{\text{sample}} \left({}^{\text{g}}/_{\text{ml}}\right)} * 100 \quad (1)$$

where Ar_{std} and Ar_{sample} are the areas under peak associated with the standard or reference and sample taxoid, respectively, and $Conc._{sample}$ and $Conc._{std}$ are the concentration of sample and reference taxoid, respectively (Shanker *et al.* 2008).

2.3. Data pre-processing

Data pre-processing is a crucial step. It has been stated that a key issue of applying filters for pre-processing is to allow the smoothening techniques to match the scale of the spectral features of interest (Bruce *et al.* 2001).

2.3.1. Savitzky–Holay Smoothing

Savitzky and Golay uses simplified least square fit intricacy, smoothing and derivatives. The general equation of the simplified least square convolution can be represented as equation (2)

$$S^* = \frac{\sum_{i=-m}^{m} C_i S_{j+i}}{n} \quad (2)$$

where S is the original spectral information, S* is the resultant (smoothed) spectral information, C_i is the coefficient for the i^{th} spectral value of the filter (smoothing window), and n is the number of convoluting integers. The index j is the running index of the original ordinate data table. The smoothing array (filter size) consists of 2 m + 1 points, where m is the half-width of the smoothing window (Tsai and Philpot 1998a).

2.3.2. Mean filter Smoothening

A mean filter takes the mean spectral value of nearest points within the considered window and the new value of j is the midpoint of the chosen window as given in equation (3)

$$S_j = \frac{\sum s_i}{n} \quad (3)$$

where n is number of sampling points. If the user specifies an even number of points as the filter size, the mean is assigned to the new value of the nearest point right of the center (longer wavelength) (Tsai and Philpot 1998a).

2.3.3. Fast Fourier transform (FFT)

The Fast Fourier process of digital filtering has been used for many years to process chemical signals. The basic equation for digital filtering is the correlation equation (4):

$$c(\pm n\Delta x) = \sum_x a(\mathrm{x}) b(x \pm n\Delta x)\ n = 0, 1, 2\ . \qquad (4)$$

where an (x) is the original signal, b(x) is the filter function, c(nΔx) is the filtered signal, and Δx is the sampling interval. Eq. (4) points to the filtered signal which is obtained by estimating the sum of the products of the signal and the filter function when the filter function is shifted across the whole signal waveform. In simple terms written as equation (5) (Singh *et al.* 2020b).

$$a(x) * b(x) = c(x) \qquad (5)$$

$$\mathrm{A(f)} * \mathrm{B(f)} = \mathrm{C(f)} \qquad (6)$$

The digital filters were implemented using the Fourier transform route as illustrated by Equations (5) and (6). The output of the FFT subroutine consists of two series, X(J) and Y(J), which are the real and imaginary components of the transform. X(J) resembles to A(f) in Eq. (6) (Betty and Horlick 1976).

2.4. Feature Selection

Derivative Spectral Analysis (DSA) algorithms were implemented to facilitate the extraction of "useful" information from hyperspectral data. Absorption features in reflectance spectra are enhanced using derivative spectroscopy. A derivative of a set of consecutive values (a spectrum). Adding the derivatives as features in the identification process optimizes and minimizes the number of bands required to achieve acceptable results due to larger JM distances (Tsai *et al.* 2002).

In the process here, spectral derivatives were assessed using a finite approximation algorithm. For the first-order derivative of a spectrum, s(**λ**), the estimation is based on equation (7).

$$\frac{\partial s}{\partial \lambda} \approx \frac{s(\lambda_j) - s(\lambda_i)}{\Delta\lambda} \qquad (7)$$

where Δλ is the separation between adjacent bands, i.e., $\Delta\lambda = \lambda_j - \lambda_i$ and $\lambda_j > \lambda_i$

However, the procedures do not work at the ends of the spectrum therefore, the resultant spectrum is shorter than the original by the width of the filter. It is noteworthy to remember that no new information is created by using derivatives (Tsai *et al.* 2002). Qualitative information regarding pigment concentration has been obtained based on the wavelength position of absorption features in derivative spectra (Louchard *et al.* 2002)

2.5. Automated Radiative Transfer Models Operator

The Spectral Index (SI) assessment (Verrelst *et al.* 2011, Verrelst *et al.* 2013) through Automated Radiative Transfer Models Operator (ARTMO) has been implemented in this study. It is based on parametric and non-parametric regression along with physically-based inversion using a lookup table (LUT). Because of complicated processing steps, a combined approach of the Radiative

Transfer Model (RTM) and vegetation indices have been introduced in the ARTMO on the MATLAB platform. With the help of the Spectral Index (SI) assessment, new generic indices have been developed in this study. Afterward, a statistical regression model with various curve fitting was implemented, which allows the relation of satellite data between desired biochemical parameters by using (ex-situ) calibration data. First, for each SI, all spectral band combinations are correlated against the generated dataset. The results are generated with the dataset that has formerly been segregated into a calibration and validation set. The obtained SI models are evaluated with multiple linear goodness-of-fit measures like the r, and the relative root means squared error (RMSEr). By investigating all bands against each other in a correlation matrix, ARTMO helps to identify redundant bands and to overcome Hughes' phenomenon or "curse of dimensionality" (Rivera *et al.* 2014b). The workflow of the entire methodology is shown in **figure 2.**

3. Results and Discussion

3.1. Statistical analysis of biochemical properties

Here, chlorophyll content, carotenoid content, total polyphenolic content (TPC), and taxol content (TC) were statistically analysed. The outcomes of the biochemical analysis are shown in **figure 3**, which suggests that common biochemical properties such as total chlorophyll and carotenoids are present in *T. wallichiana* in a certain specified range. The chlorophyll content varied between 2.014 to 4.195 mg/g with an average of 3.541 +/- 0.501. Carotenoids varied between 0.704 to 0.983 mg/g with an average of 0.836 +/- 0.087 respectively. TPC and TC varied between 94.676 to 72.656 mg GAE/g of its fresh weight (FW) with an average of 79.901 ± 6.271 mg GAE/g of FW and 0 to 0.037 mg/g FW with an average of 0.011 +/- 0.012 mg/g FW. TPC showed a significant correlation coefficient (0.672) with altitude. The other biochemical properties did not show any statistically significant correlation with elevation. The frequency of the *T. wallichiana* becomes lesser after 3040 m as more of a grassland ecosystem naturally exists at Nanda Devi Biosphere Reserve. The highest Taxol content concerning elevation is recorded between 2850 to 3000 m in Nanda Devi Biosphere Reserve. The TPC values and their correlation with elevation are strong unlike taxol due to low temperatures at higher altitudes up to elevation 3100 m. TPC and TC relations show that medicinal plants also carry phenolic content in them that indirectly relates to redox properties which are responsible for their antioxidant effects (Heinig and Jennewein 2009, Baba and Malik 2015). *T. wallichiana* plant showed lowest Taxol content near timberline in the upper montane zone beyond which grassland ecosystem (3040 m) at Phurkia, which is the ecotone region and near lower montane zone at the point of human intervention Khati (last habitable point in the valley) (Rai *et al.* 2019). The lowest TC measured was 0.001 mg/g and 0 mg/g of FW at the upper montane zone and lower montane zone respectively.

3.2. Denoising and feature selection of hyperspectral data

The smoothening and filtering technique used here Average Mean filter, Savitzky Golay, and FFT paired with derivative analysis further to select the most optimal band for Taxol. The derivative analysis, a feature selection process is an efficient process in capturing the subtle difference in the spectra required to locate any specific feature present in the spectra. It works best when paired with an optimal pre-processing technique. The spectra depicted in figure 4(a) are the raw spectra of T. wallichiana while figures 4(c), (e), (f) are denoised spectra. The derivative of all the spectra is depicted in figure 4(b), (d), (f), (h). A high-resolution clear sectional view of individual spectra of *T. wallichiana* is represented in figure 5(a), (b) which showed a significant difference among the three methods of denoising.

In **figure 5(a)** it is clearly visible how each smoothening algorithm transformed in regards to the raw spectra of *T. wallichiana*. In the case of moving averages, a least-squares fit is made to a zero-order polynomial (i.e., a straight horizontal line or a constant value). Typically, these features are flattened by other (simpler) averaging points within the filter window (Tsai and Philpot 1998b). The primary factor controlling the degree of smoothing is the size (bandwidth) of the filter window used for convolution or averaging. It can be seen from table 2 from TC 4 - TC 8 that all the bands lies in-between 420-610 nm range, which is the initial reflectance wavelength of the spectra. It was found that the Average Mean filter may not have presented the nearest value to the raw spectra but maintained the peak of the spectra as shown in figure 5(a).

The Savitzky-Golay filtering technique makes use of frequency data or spectroscopic (peak) data. For frequency data, this smoothening method is more effective at conserving the high-frequency components of the signal while upholding the profile and height of waveform peaks (in their case, Gaussian-shaped spectral peaks) (Persson and Strang 2003). The Savitzky-Golay method was the least successful compared to moving average filter and FFT to de-noise the spectra with small disturbance as shown in **table 2.** The wavelength allocated for Taxol content after applying the Savitzky-Golay belongs to the SWIR region (TC 9 to TC 14) of the spectra. The signal after denoising in figure 5(a), (b) have lost the pattern, as well as its reflectance magnitude, was also changed.

In the case of the FFT filter as can be seen from **figures 5(a) and (b**) it is neither enhanced nor reduced the raw signal after application keeping the information in the signal intact but it did lose the patterns (peaks and dips). The wavelength selected using FFT filter showed a negative correlation of 0.320 at 960-970 nm with the SR indices generated values. A negative correlation was observed in the NIR region. Similarly, the derivative analysis of raw spectra shows a negative correlation of 0.370 near the 958-968 nm range. This implies terpenes have a negative correlation in the NIR region and FFT pre-processing is close to raw spectra. FFT pre-processing does modify the signal in such a way that the subtle difference is preserved in the spectra after feature selection. This filter proves to have better correlation results for the Taxol indices generation than the Savitzky Golay filter as shown in table 2 (TC 15 and TC 16).

Contrary to the above filter techniques used, Savitzky Golay changes the spectra magnitude and pattern creating more loss of information. Previously, many studies have stated that the mean filter algorithm is not as good as Savitzky Golay but in our case, retrieval of biochemical variables like Taxol from hyperspectral data was found most suitable. Usually, the first parameter which is retrieved from electromagnetic spectra using Hyperspectral remote sensing is Chlorophyll. This chlorophyll is majorly allocated in electromagnetic spectra after 480 nm (Yang *et al.* 2015), but the visible region expands between 370–700 nm. Hence the information between (370- 450) nm is discarded as noise. TC is majorly allocated in the visible region where noise is much. Moreover, TC is very less hence locating these small peaks for the same was most appropriately done using an Average Mean filter in combination with derivative analysis. The Average Mean filter removed the noise without compromising the ability to resolve fine spectral detail. FFT data provided a smooth spectrum preserving the magnitude of the signal but during absorbance band selection, the number of bands selected due to FFT transformation was very less (Betty and Horlick 1976) making it the second-best filter. These filtered spectra followed by feature selection led to the selection of wavelengths. The advantage and disadvantages of each filter technique were judged based on a statistical correlation between the indices generated values with real estimated values from the field data.

Each spectrum of *T. wallichiana* after application of filter was followed by first derivative to select the absorption bands. These selected wavebands were then included in the ARTMO SI generation toolbox. The derivative analysis presented a different range of wavelengths specified

under VIs (370-700 nm), Near SWIR (NSWIR-1350-1450 nm), and Far SWIR (FSWIR- 1800-2500 nm) regions (Hennessy *et al.* 2020). Some of these wavelengths are represented in table 2. The derivative analysis in figure 5(b) suggests that the uneven raw spectra were much smoother after filter application.

3.3. Taxol model development

The regression models are based on spectral indices are typically empirical equations enabling the mapping of biophysical and biochemical parameters over a large area. The specified retrieval strategies within the SI toolbox were first analyzed and then the wavelength selected using the derivative analysis from the raw spectral spectroradiometer data are provided in the text file format.

Novel Taxol indices are generated by identifying various combinations from the spectroradiometer raw data between a spectral range of 350-2500 nm as shown in table 2. The index for each selected combination wavelength was tested. ARTMO model spectral indices (SI) were specified with the preselected wavelength using feature selection (derivative analysis). Using statistical techniques, the Taxol content (TC) retrieval accuracies of newly developed Taxol models were investigated. The best-selected wavelength for band two-band combination yielded a statistically significant correlation for the Average Mean filter. Based on statistical performance two best models were selected. Various curve fitting was also tested with real observed and model-generated Taxol content data, among which the linear curve fitting was found best as shown in **table 3**. The visual comparative representation of all the models is shown in **figure 6** using a Taylor plot. Model LTC-TC 5 and LTC-TC 8 showed a very high correlation of 0.719, 0.718. RMSEr values of both the models are found relatively equivalent i.e., 0.578 and 0.576 for model LTC-TC 5 and LTC-TC 8 respectively. After testing different indices at the selected wavelength, Average Mean denoised spectra found to be the best filter for indices generation, which in combination with the feature selection showed the best statistical performance among all other models. LTC-TC 5 and LTC-TC 8 are the top-performing models which are formed using the combination of wavelengths selected from the visible range values. These values provided ideal results for deriving TC from hyperspectral reflectance data. Similar high relation of this class of compounds i.e., Taxol is recently reported to have an association with the visible region between 400-500nm (Fine *et al.* 2021). Hence, the results obtained are consistent with Taxol indices generated using the visible region reflectance values.

The model generation with ARTMO also gives the best results when the most appropriate bands after smoothening were given as input. The linear curve fitting performed best between modelled data values from ARTMO and TC estimated with HPLC analysis. Hence, the overall results suggests that Taxol content can be quantified using the hyperspectral reflectance in the visible range of 415-421 +/- 5 nm.

4. Conclusion

The result of statistical analysis suggests that the elevation along with its ecosystem climatic conditions plays an important role in variation in phytoconstituents. The highest Taxol content concerning elevation is recorded between 2850 to 3000 m in Nanda Devi Biosphere Reserve. For retrieval of biomedicinal molecule Taxol from hyperspectral data, the average mean filter was found most suitable. TC found in leaves of *T. wallichiana* was very less in terms of quantity hence locating these small peaks corresponding to it in the reflectance curve was most appropriately done using an Average Mean filter in combination with derivative analysis for indices generation. The SI assessment through ARTMO provides a systematized approach in a streamlined way for the selection as well as the assessment of the most precise and sensitive SI formulations which can be

used for parameter retrieval using hyperspectral datasets. ARTMO generated SI clearly shows that the TC can be quantified using spectral indices and the model developed using these indices shows the best results in terms of r and RMSEr. The linear curve fitting with modelled data values from ARTMO correlated best with measured TC. Empirical methods like the regression-based model are the best tool to monitor the health of the plant on a real-time basis as it takes less time to compute and is easy to use.

In the future, sampling at more locations in the Himalayas will be performed with the inclusion of seasonality to check the robustness of the model developed. The other region between 370-480 nm of the spectra also needs to be rigorously analysed as it holds more valuable information. This requires a network for the collection of ground samples which becomes quite difficult due to elevation and harsh weather in the region. The current study will reduce the time and tedious effort of researchers and will make the management of canopy-level information much simpler. Compared to conventional labour-intensive on-site measurements, the proposed method will deliver quick information about the Taxol content and thus can help in protecting and managing forest resources more realistically.

Funding

A.G.'s is funded under the University Grant Commission`s Junior Research Fellowship program. This work is funded by the *National Mission on Himalayan Studies*, G.B. Pant National Institute of Himalayan Environment (NIHE), Ministry of Environment, Forest & Climate Change (MoEF&CC), Government of India.

Acknowledgment

The authors are thankful to the University Grant Commission and National Mission for Himalayan Studies for the necessary financial assistance and support throughout. The authors also acknowledge the Institute of Environment and Sustainable Development, Banaras Hindu University, and Central Institute of Medicinal and Aromatic Plants, Lucknow, India for providing the necessary laboratory support for the study. The authors also extend their sincere thanks to NMHS, G.B. Pant National Institute of Himalayan Environment (NIHE) for their constant support in this work. The authors also extend their gratitude to Dr. Jochem Verrelst, Imaging Processing Laboratory (IPL) at the University of València, Spain for providing the ARTMO tool and guidance for the presented work.

Conflict of interest

There is no conflict of interest

References

Appendino, G., Gariboldi, P., Gabetta, B., Pace, R., Bombardelli, E. & Viterbo, D.J.J.O.T.C.S., Perkin Transactions 1, 1992. 14β-hydroxy-10-deacetylbaccatin iii, a new taxane from himalayan yew (taxus wallichiana zucc.). (21), 2925-2929.

Baba, S.A. & Malik, S.A., 2015. Determination of total phenolic and flavonoid content, antimicrobial and antioxidant activity of a root extract of arisaema jacquemontii blume. *Journal of Taibah University for Science,* 9 (4), 449-454.

Bala, S., Uniyal, G., Chattopadhyay, S., Tripathi, V., Sashidhara, K., Kulshrestha, M., Sharma, R., Jain, S., Kukreja, A. & Kumar, S.J.J.O.C.A., 1999. Analysis of taxol and major taxoids in himalayan yew, taxus wallichiana. 858 (2), 239-244.

Betty, K.R. & Horlick, G.J.a.S., 1976. A simple and versatile fourier domain digital filter. 30 (1), 23-27.

Bruce, L.M., Li, J.J.I.T.O.G. & Sensing, R., 2001. Wavelets for computationally efficient hyperspectral derivative analysis. 39 (7), 1540-1546.
Chakchak, H. & Zineddine, H.J.a.J.O.C., 2013. Extraction and identification of new taxoids from the moroccan yew. 25 (9), 4861-4864.
Clark, M.L., Roberts, D.A. & Clark, D.B.J.R.S.O.E., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. 96 (3-4), 375-398.
Ekor, M.J.F.I.P., 2014. The growing use of herbal medicines: Issues relating to adverse reactions and challenges in monitoring safety. 4, 177.
Fine, P.V., Salazar, D., Martin, R.E., Metz, M.R., Misiewicz, T.M. & Asner, G.P.J.E., 2021. Exploring the links between secondary metabolites and leaf spectral reflectance in a diverse genus of amazonian trees. 12 (2), e03362.
Fu, Y.J., Sun, R., Zu, Y.G., Li, S.M., Liu, W., Efferth, T., Gu, C.B., Zhang, L. & Luo, H.J.B.C., 2009. Simultaneous determination of main taxoids in taxus needles extracts by solid-phase extraction-high-performance liquid chromatography with pentafluorophenyl column. 23 (1), 63-70.
Geetha, T. & Geetha, N.J.I.J.O.P.R., 2014. Phytochemical screening, quantitative analysis of primary and secondary metabolites of cymbopogan citratus (dc) stapf. Leaves from kodaikanal hills, tamilnadu. 6 (2), 521-529.
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. *Remote sensing applications in environmental research.* Springer, Cham, 177-194.
Heinig, U. & Jennewein, S., 2009. Taxol: A complex diterpenoid natural product with an evolutionarily obscure origin. *African Journal of Biotechnology*, 8 (8).
Hennessy, A., Clarke, K. & Lewis, M.J.R.S., 2020. Hyperspectral classification of plants: A review of waveband selection generalisability. 12 (1), 113.
Juyal, D., Thawani, V., Thaledi, S., Joshi, M.J.J.O.T. & Medicine, C., 2014. Ethnomedical properties of taxus wallichiana zucc.(himalayan yew). 4 (3), 159-161.
Khan, M., Verma, S., Srivastava, S., Shawl, A., Syamsundar, K., Khanuja, S. & Kumar, T., 2006a. Essential oil composition of taxus wallichiana zucc. From the northern himalayan region of india. *Flavour and fragrance journal*, 21 (5), 772-775.
Khan, M., Verma, S., Srivastava, S., Shawl, A., Syamsundar, K., Khanuja, S., Kumar, T.J.F. & Journal, F., 2006b. Essential oil composition of taxus wallichiana zucc. From the northern himalayan region of india. 21 (5), 772-775.
Kokaly, R.F., Skidmore, A.K.J.I.J.O.a.E.O. & Geoinformation, 2015. Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm. 43, 55-83.
Landgrebe, D.J.I.S.P.M., 2002. Hyperspectral image data analysis. 19 (1), 17-28.
Lin, P., Qin, Q., Dong, H. & Meng, Q., Year. Hyperspectral vegetation indices for crop chlorophyll estimation: Assessment, modeling and validationed.^eds. *2012 IEEE International Geoscience and Remote Sensing Symposium*IEEE, 4841-4844.
Ling, B., Goodin, D.G., Raynor, E.J. & Joern, A.J.F.I.P.S., 2019. Hyperspectral analysis of leaf pigments and nutritional elements in tallgrass prairie vegetation. 10, 142.
Louchard, E.M., Reid, R.P., Stephens, C.F., Davis, C.O., Leathers, R.A., Downes, T.V. & Maffione, R.J.O.E., 2002. Derivative analysis of absorption features in hyperspectral remote sensing data of carbonate sediments. 10 (26), 1573-1584.
Mohamed, A.A., Khalil, A.A. & El-Beltagi, H.E.J.G.Y.A., 2010. Antioxidant and antimicrobial properties of kaff maryam (anastatica hierochuntica) and doum palm (hyphaene thebaica). 61 (1), 67-75.
Pan, S.-Y., Litscher, G., Gao, S.-H., Zhou, S.-F., Yu, Z.-L., Chen, H.-Q., Zhang, S.-F., Tang, M.-K., Sun, J.-N., Ko, K.-M.J.E.-B.C. & Medicine, A., 2014. Historical perspective of traditional indigenous medical practices: The current renaissance and conservation of herbal resources. 2014.

Pandey, P.C., Anand, A., Srivastava, P.K.J.B. & Conservation, 2019. Spatial distribution of mangrove forest species and biomass assessment using field inventory and earth observation hyperspectral data. 28 (8-9), 2143-2162.
Peerbhay, K.Y., Mutanga, O., Ismail, R.J.I.J.O.P. & Sensing, R., 2013. Commercial tree species discrimination using airborne aisa eagle hyperspectral imagery and partial least squares discriminant analysis (pls-da) in kwazulu–natal, south africa. 79, 19-28.
Persson, P.-O. & Strang, G., 2003. Smoothing by savitzky-golay and legendre filters. *Mathematical systems theory in biology, communications, computation, and finance.* Springer, 301-315.
Purohit, A., Maikhuri, R., Rao, K. & Nautiyal, S., 2001a. Impact of bark removal on survival of taxus baccata l.(himalayan yew) in nanda devi biosphere reserve, garhwal himalaya, india. *Current Science*, 586-590.
Purohit, A., Maikhuri, R., Rao, K. & Nautiyal, S.J.C.S., 2001b. Impact of bark removal on survival of taxus baccata l.(himalayan yew) in nanda devi biosphere reserve, garhwal himalaya, india. 586-590.
Rai, I.D., Singh, G., Pandey, A. & Rawat, G., 2019. Ecology of treeline vegetation in western himalaya: Anthropogenic and climatic influences. *Tropical ecosystems: Structure, functions and challenges in the face of global change.* Springer, 173-192.
Rivera, J.P., Verrelst, J., Delegido, J., Veroustraete, F. & Moreno, J., 2014a. On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. *Remote Sensing,* 6 (6), 4927-4951.
Rivera, J.P., Verrelst, J., Delegido, J., Veroustraete, F. & Moreno, J.J.R.S., 2014b. On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. 6 (6), 4927-4951.
Sadeghi-Aliabadi, H., Asghari, G., Mostafavi, S. & Esmaeili, A.J.D.J.O.P.S., 2015. Solvent optimization on taxol extraction from taxus baccata l., using hplc and lc-ms. 17 (3), 192-198.
Shanker, K., Negi, A.S., Chattopadhyay, S.K., Sashidhara, K., Kaur, T., Gupta, M., Agrawal, P., Misra, A.J.J.O.H., Spices & Plants, M., 2008. Determination of paclitaxel, 10-dab, and related taxoids in himalayan yew using reverse phase hplc. 13 (4), 25-44.
Singh, P., Pandey, P.C., Petropoulos, G.P., Pavlides, A., Srivastava, P.K., Koutsias, N., Deng, K.a.K. & Bao, Y., 2020a. Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends. *Hyperspectral remote sensing.* Elsevier, 121-146.
Singh, P., Srivastava, P.K., Malhi, R.K.M., Chaudhary, S.K., Verrelst, J., Bhattacharya, B.K. & Raghubanshi, A.J.I.S.J., 2020b. Denoising aviris-ng data for generation of new chlorophyll indices.
Srinivasan, R., Sugumar, V.R.J.J.O.E.-B.C. & Medicine, A., 2017. Spread of traditional medicines in india: Results of national sample survey organization's perception survey on use of ayush. 22 (2), 194-204.
Srivastava, P.K., Gupta, M., Singh, U., Prasad, R., Pandey, P.C., Raghubanshi, A. & Petropoulos, G.P., 2020a. Sensitivity analysis of artificial neural network for chlorophyll prediction using hyperspectral data. *Environment, Development and Sustainability*, 1-16.
Srivastava, P.K., Malhi, R.K.M., Pandey, P.C., Anand, A., Singh, P., Pandey, M.K. & Gupta, A., 2020b. Revisiting hyperspectral remote sensing: Origin, processing, applications and way forward. *Hyperspectral remote sensing.* Elsevier, 3-21.
Suffness, M., 1995. *Taxol: Science and applications*: CRC press.
Torrecilla, E., Piera, J. & Vilaseca, M., 2009. Derivative analysis of hyperspectral oceanographic data. *Advances in geoscience and remote sensing.* IntechOpen.
Tsai, F. & Philpot, W., 1998a. Derivative analysis of hyperspectral data. Remote sensing environment.
Tsai, F., Philpot, W.D.J.I.T.O.G. & Sensing, R., 2002. A derivative-aided hyperspectral image analysis system for land-cover classification. 40 (2), 416-425.
Tsai, F. & Philpot, W.J.R.S.O.E., 1998b. Derivative analysis of hyperspectral data. 66 (1), 41-51.

Van Rozendaal, E.L., Lelyveld, G.P. & Van Beek, T.a.J.P., 2000. Screening of the needles of different yew species and cultivars for paclitaxel and related taxoids. 53 (3), 383-389.
Verrelst, J., Malenovský, Z., Van Der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P. & Moreno, J.J.S.I.G., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. 40 (3), 589-629.
Verrelst, J., Rivera, J., Alonso, L. & Moreno, J., Year. Artmo: An automated radiative transfer models operator toolbox for automated retrieval of biophysical parameters through model inversioned.^eds. *Proc. EARSeL 7th SIG-Imag. Spectrosc. Workshop*Citeseer, 11-13.
Verrelst, J., Rivera, J.P., Guadalarjara, A., Delegido, J. & Moreno, J., Year. Artmo's new spectral indices (si) module to rapidly evaluate a multitude of sis for mapping of biophysical parametersed.^eds. *EARSeL 8th SIG-Imaging Spectroscopy Workshop*, 08-10.
Yang, X., Yu, Y., Fan, W.J.E.M. & Assessment, 2015. Chlorophyll content retrieval from hyperspectral remote sensing imagery. 187 (7), 1-13.
Yuan, H., Ma, Q., Ye, L. & Piao, G.J.M., 2016. The traditional medicine and modern medicine from natural products. 21 (5), 559.
Zagajewski, B., Kycko, M., Tømmervik, H., Bochenek, Z., Wojtun, B., Bjerke, J.W. & Kłos, A., 2018. Feasibility of hyperspectral vegetation indices for the detection of chlorophyll concentration in three high arctic plants: Salix polaris, bistorta vivipara, and dryas octopetala.
Zhu, L., Chen, L.J.C. & Letters, M.B., 2019. Progress in research on paclitaxel and tumor immunotherapy. 24 (1), 40.

Table 1. Common attributes of *Taxus wallichiana*.

SI No.	**Attributes**	***Taxus wallichiana***	**References**
1.	Common Name	Himalayan yew	(Khan *et al.* 2006b)
2.	Family	Taxaceae.	(Khan *et al.* 2006b)
3.	Plant Height	10-28m	(Juyal *et al.* 2014)
4.	Worldwide distribution	Europe, North America, Northern India, Pakistan, China, and Japan	(Khan *et al.* 2006a)
5.	Distribution in India	Meghalaya Manipur, and Nanda Devi Biosphere Reserve (NDBR) of Garhwal Himalayas	(Purohit *et al.* 2001b, Khan *et al.* 2006b)
6.	Altitude	Himalayan altitude of 1800-3300amsl	(Purohit *et al.* 2001a)
7.	Forest type	Temperate Forests type	(Juyal *et al.* 2014)
8.	Concern status	Endangered	http://envis.frlht.org/plantdetails/d6e172fda7fed3241a4ea444f83e7d82/2f2b9537310d77c9cacad37c1442f696. . . .

Table 2: New Taxol Indices Developed using optimal Bands.

SI No	Data	Band 1	Band 2	New Indices for Taxol Content (TC)
1	Raw data	968	958	TC 1 = (R958-R968/R958+R968)
		1757	1767	TC 2 = (R1767-R1757/R1767+R1757)
		2262	2272	TC 3 = (R2272-R2262/R2272+R2262)
2	Average Mean	421	426	TC 4 = (R426-R421/R426+R421)
		415	421	TC 5 = (R415-R421/R415+R421)
		608	601	TC 6 = (R601-R608/R601-R608)
		421	426	TC 7 = (R421/R426)
		421	415	TC 8 = (R415/R421)
3	Savitzky Golay	1728	1738	TC 9 = (R1738-R1728/R1738+R1728)
		1738	1748	TC 10 = (R1748-R1738/R1748+R1738)
		1738	1758	TC 11 = (R1758-R1738/R1758+R1738)
		1728	1738	TC 12 = (R1728/R1738)
		1738	1748	TC 13 = (R1738/R1748)
		1748	1758	TC 14 = (R1758/R1748)
4	Fast Fourier Transform	370	374	TC 15 = (R374-R370/R374+R370)
		370	374	TC 16 = (R370/R374)

Table 3: Taxol Models Established Using New Taxol Indices Developed from Spectroradiometer Data.

Model Name	Data	Formula	R	RMSEr
LTC-TC 1	Raw Data	TC = 0.0565 * TC 1 + 0.0388	0.402	0.766
LTC-TC 2		TC = 0.1752 * TC 2 + 0.0559	0.340	0.826
LTC-TC 3		TC = 0.1065 * TC 3 + 0.0174	0.488	0.701
LTC-TC 4	Average Mean	TC = 0.4133 * TC 4 + 0.0514	0.512	0.819
LTC-TC 5		TC = 0.4154 * TC 5 + 0.0251	0.719	0.678
LTC-TC 6		TC = 0.0884 * TC 6 + 0.0173	0.331	0.721
LTC-TC 7		TC = 0.4432 * TC 7 + 0.0523	0.513	0.823
LTC-TC 8		TC = 0.4175 * TC 8 + 0.0248	0.718	0.676
LTC-TC 9	Savitzky Golay	TC = 0.0409 * TC 9 + 0.0561	0.433	0.823
LTC-TC 10		TC = 0.0176 * TC 10 + 0.0601	0.324	0.833
LTC-TC 11		TC = 0.0413 * TC 11 + 0.0771	0.327	0.865
LTC-TC 12		TC = 0.5882 * TC 12 + 0.0406	0.317	0.884
LTC-TC 13		TC = 0.0231 * TC 13 + 0.0615	0.391	0.836
LTC-TC 14		TC = 0.0413 * TC 14 + 0.0771	0.327	0.865
LTC-TC 15	Fast Fourier Transform	TC = 0.6063 * TC 15 + 0.0817	0.468	0.886
LTC-TC 16		TC = 0.0318 * TC 17 − 0.0119	0.438	-2.232

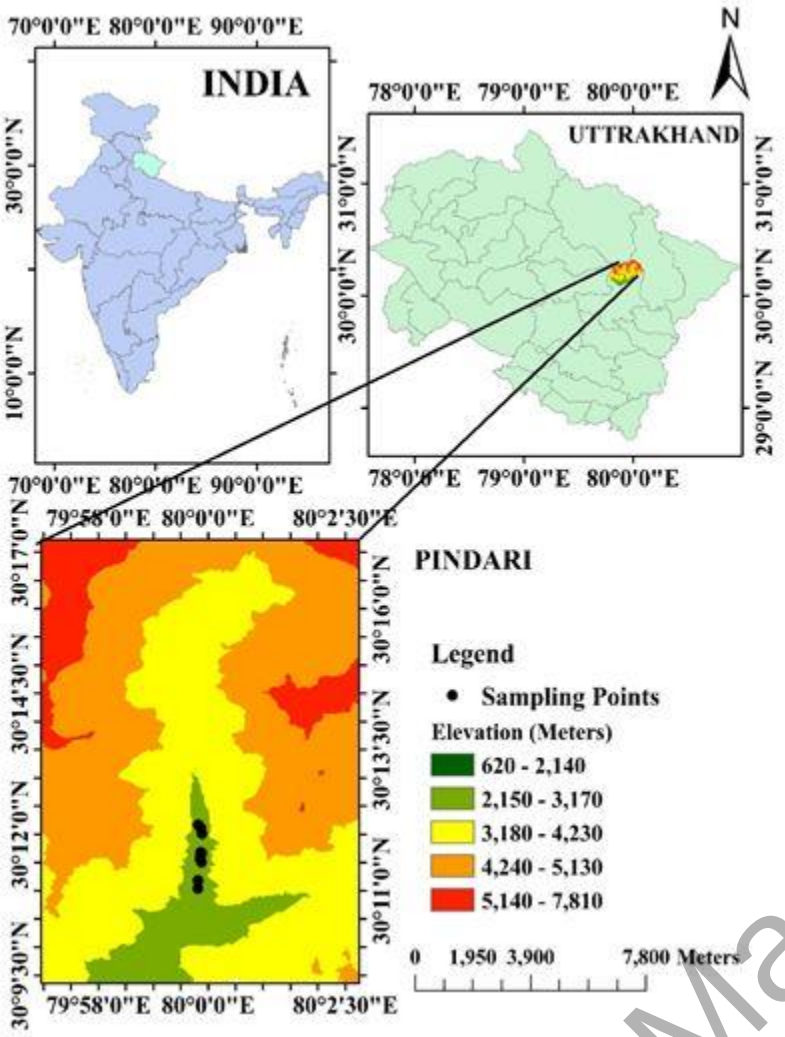

Figure 1. Map of study location with elevation.

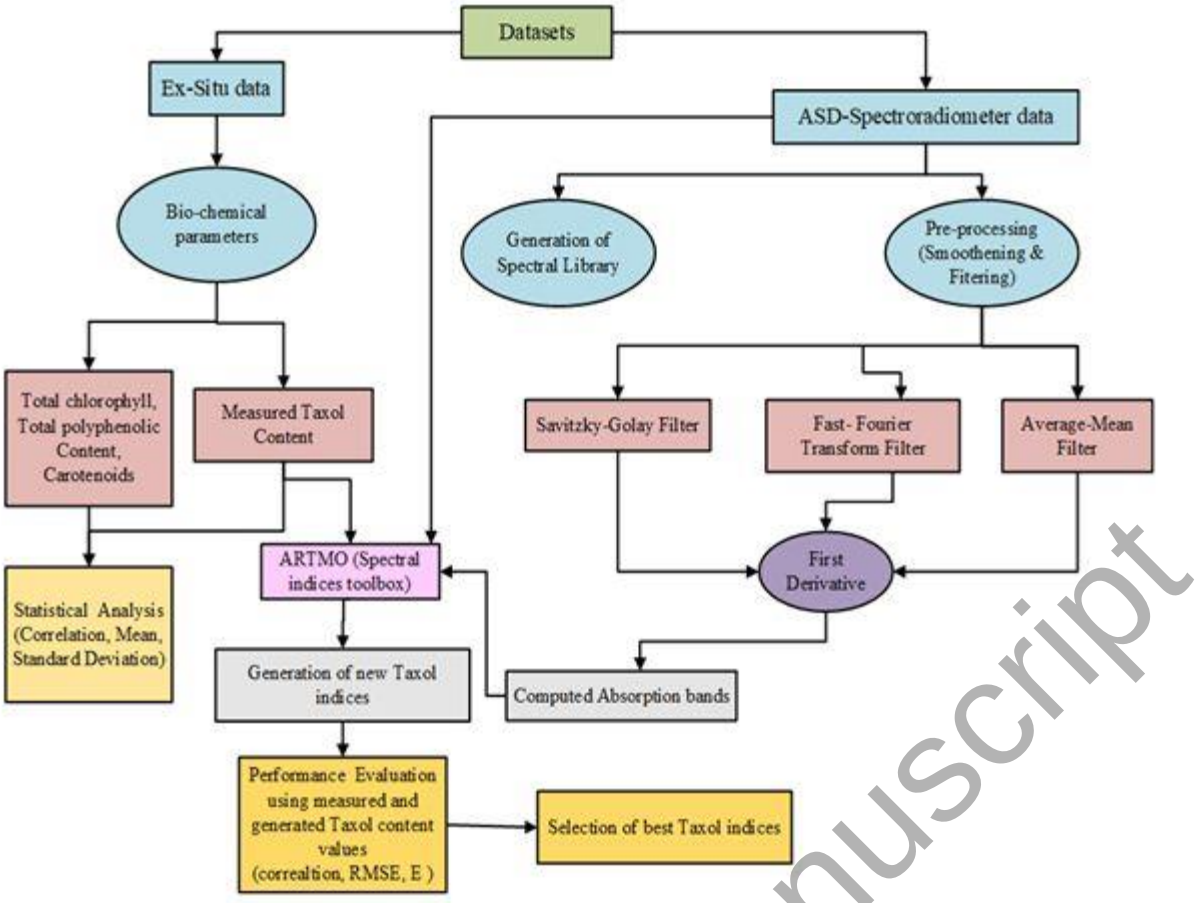

Figure 2. Workflow chart for the study.

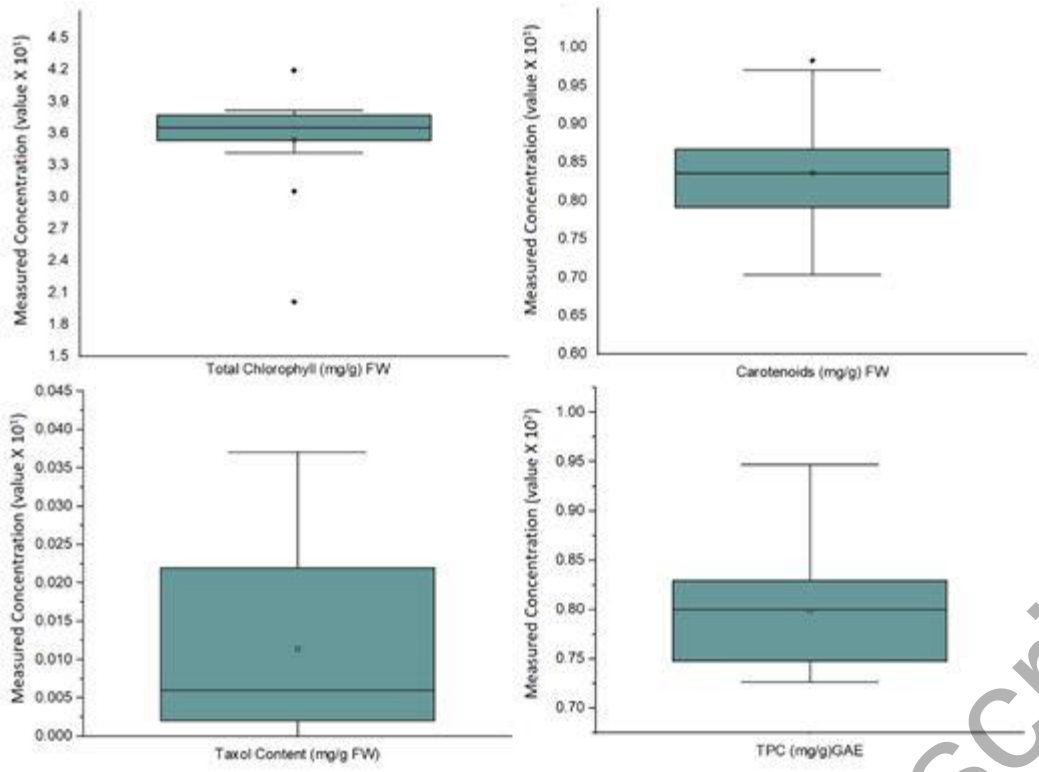

Figure 3. Box Plot for the biochemical parameter's depiction.

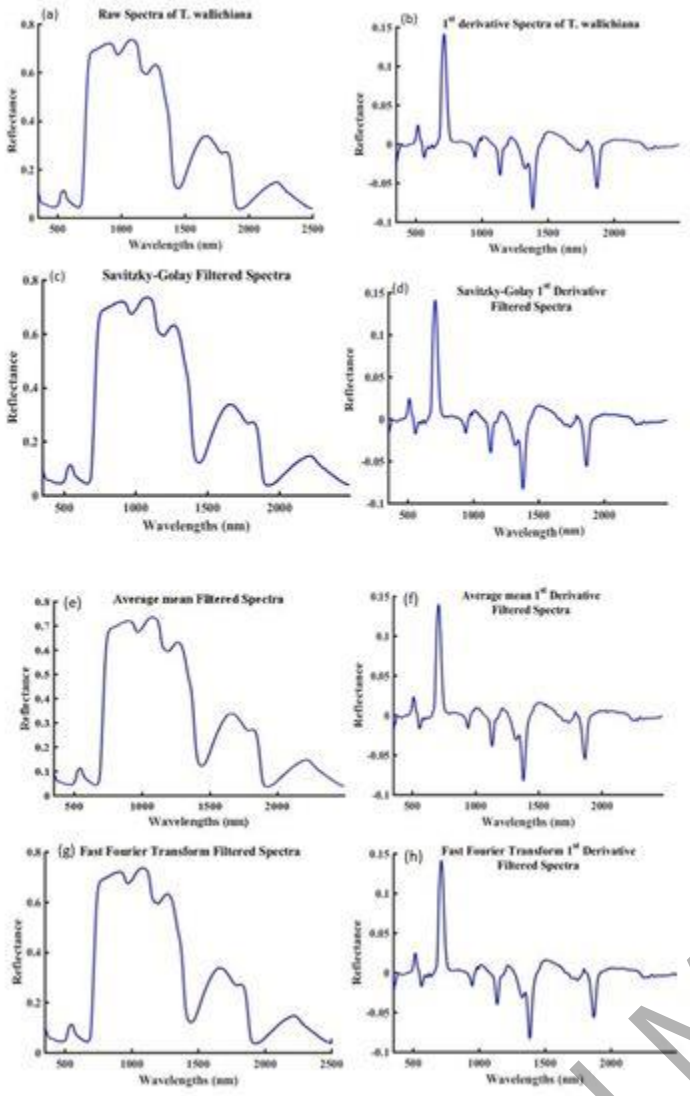

Figure 4. (a), (c), (e), (g) represents raw spectra, Savitzky Golay, Average Mean, Fast Fourier transformed spectra of *T. wallichiana* and respectively and (b), (d), (f), (h) is its 1st derivative i.e., transformed spectra respectively.

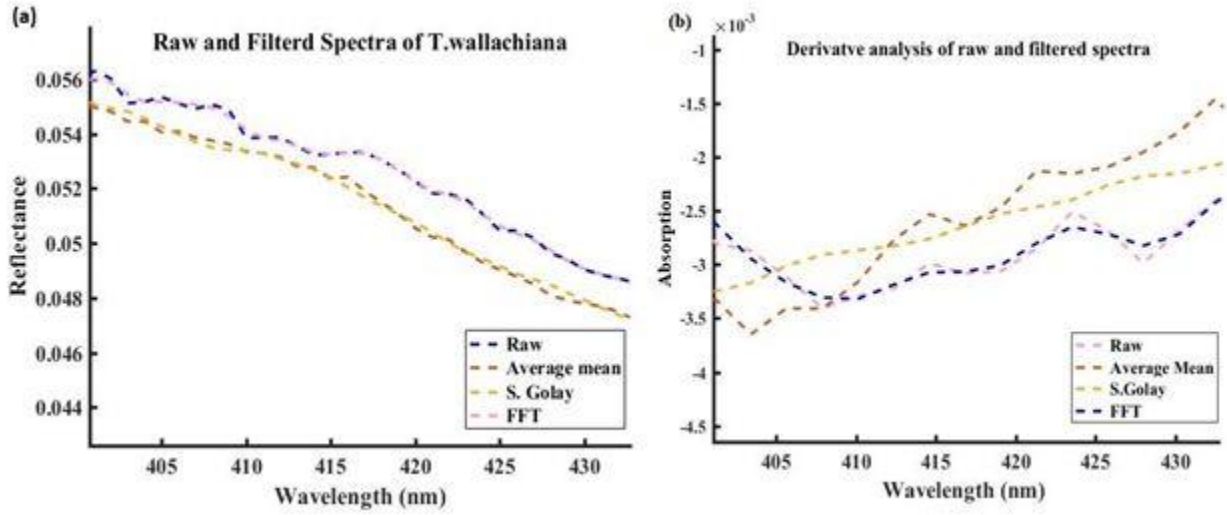

Figure 5. (a) Reflectance values of raw and filter applied spectra & (b) Absorption values of derivative applied raw and filtered spectra.

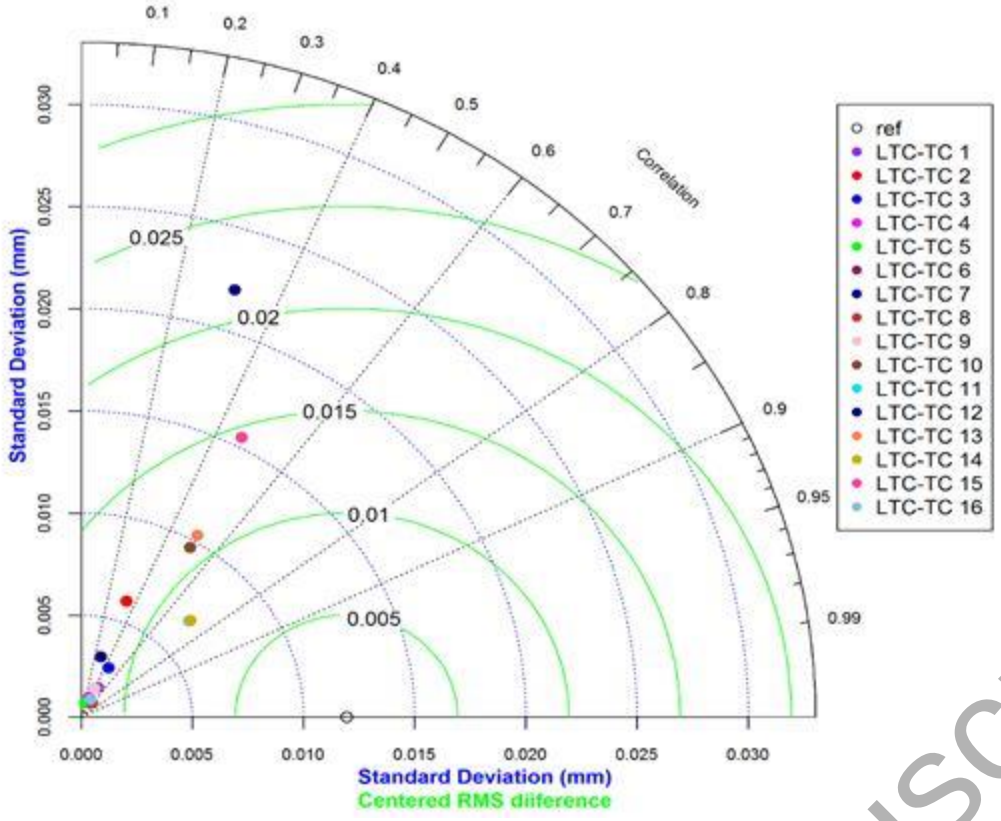

Figure 6. Evaluation of all the models using the Taylor plot.