International Journal of Applied Earth Observation and Geoinformation 36 (2015) 22–29

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Retrieval of tea polyphenol at leaf level using spectral transformation and multi-variate statistical approach

Dibyendu Dutta[a,*], Prabir Kumar Das[a], Uttam Kumar Bhunia[a], Upasana Singh[a], Shalini Singh[a], Jaswant Raj Sharma[b], Vinay Kumar Dadhwal[b]

[a] *Regional Remote Sensing Centre-East (NRSC), ISRO, Kolkata, West Bengal, India*
[b] *National Remote Sensing Centre, Balanagar, Hyderabad, India*

ARTICLE INFO

Article history:
Received 15 August 2014
Accepted 3 November 2014
Available online 20 November 2014

Keywords:
Tea polyphenol
Hyperspectral
Discriminant analysis
Stepwise multiple linear regression
Partial least square regression
Principal component analysis

ABSTRACT

In the present study, field based hyperspectral data was used to estimate the tea (*Camellia sinensis* L.) polyphenol at Deha Tea garden of Assam state, India. Leaf reflectance spectra were first filtered for noise and then transformed into normalized and first derivative reflectance for further analysis. Stepwise discriminant analysis was carried out to select sensitive bands for a range of polyphenol concentration by minimizing the effects of other factors such as age of the bushes and management practices. The wavelengths at 358, 369, 484, 845, 916, 1387, 1420, 1435, 1621 and 2294 nm were identified as sensitive to tea polyphenol, among which 2294 nm was found to be the most recurring band. The noise removed selected bands, their transformed derivatives and principal components were regressed with the tea polyphenol using univariate and multi-variate analysis. In univariate analysis the correlation was very poor with RMSE more than 3.0. A significant improvement in R^2 values were observed when multivariate analyses like stepwise multiple linear regression (SMLR) and partial least square regression (PLSR) was carried out. The PLSR of first derivative reflectance was most accurate ($R^2 = 0.81$ and RMSE = 1.39 mg g^{-1}) among all the uni- and multivariate analysis for predicting the polyphenol of fresh tea leaves.

Introduction

Tea (*Camellia sinensis* L.) is one of the most popular stable beverages in India and constitutes 21% of global area and 23% of world tea production (Basu Majumder et al., 2010). About 193 million kg processed tea is exported annually from India and is a major foreign exchange earner. Over the years there is stagnation in tea production and decline in tea quality due to old age of tea bushes, declining soil health and increased incidence of pests and diseases (Dutta et al., 2011). To maintain the export quality there is need to periodically monitor green tea leaf quality over time and space. Traditional method of tea quality assessment is either based on the subjective evaluation by the tea tasters, or by destructive sampling and subsequent wet laboratory analysis which is time- and labour intensive and seldom represents the spatial variability of quality. Apparently a rapid, non-destructive, and accurate determination of foliar tea quality over a large area has remained a crucial issue to be addressed. Development of hyperspectral sensors offers possibilities to estimate and monitor vegetation quality in spatial and temporal scale (Knox et al., 2011; Mutanga and Kumar, 2007). The narrow and contiguous spectral bands make it possible to detect subtle variations in the reflectance spectra caused by differences in biochemical composition and physiology of vegetation (Davey, 2009; Schlerf et al., 2010). Many researchers have used reflectance spectra to measure biochemical variables of green vegetation and had shown the potential of the remote sensing technique to capture the spatial bio-chemical variation (Curran, 1989; Schlerf et al., 2010; Skidmore et al., 2010).

The quality of processed tea is primarily determined by leaf chlorophyll, moisture, polyphenols, amino acids and soluble sugars among others. Catechins are the major polyphenols which determine the tea quality and constitutes up to 30% of dry matter in green leaves and are mainly responsible for the characteristic astringent and bitter taste of black tea (Zhang et al., 1992). The younger leaves contain the optimal ratio of polyphenols and amino acids, which forms the special taste of tea beverage. It has been demonstrated by Schulz et al. (1999) that near-infrared spectroscopy can successfully be used as a rapid method for estimation of green tea quality, especially alkaloids and phenolic substances under laboratory

* Corresponding author. Tel.: +91 9434020503.
E-mail addresses: ddisro@gmail.com (D. Dutta), prabir111@gmail.com (P.K. Das), ubhunia.rsvu@gmail.com (U.K. Bhunia), upasana.2703@gmx.com (U. Singh), shalini0110@gmail.com (S. Singh), jrsharma50@gmail.com (J.R. Sharma), dadhwalvk@hotmail.com (V.K. Dadhwal).

http://dx.doi.org/10.1016/j.jag.2014.11.001
0303-2434/

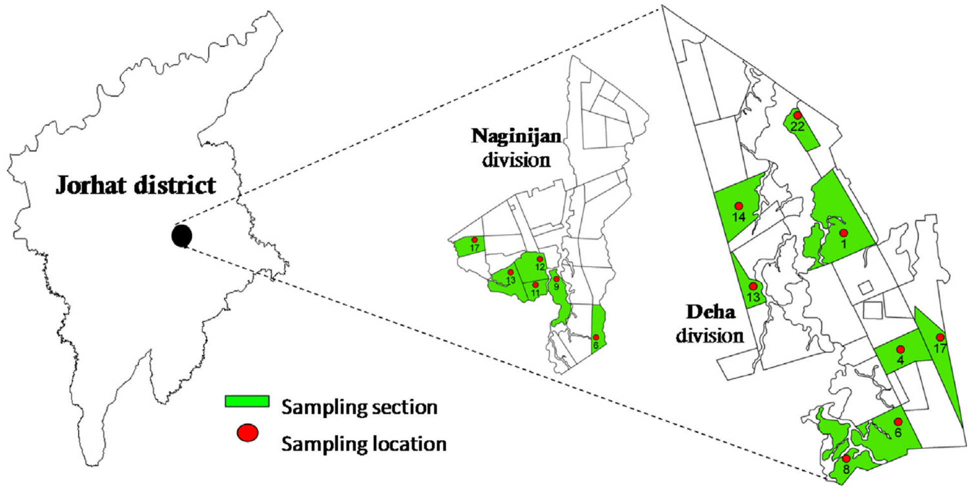

Fig. 1. Map sowing Deha Tea Estate with the selected sections and sample points.

condition. They had used the entire wavelength region from 1100 nm to 2500 nm for estimation of caffeine in green and black tea along with 3 selected wavelengths of 1470 nm, 1646 nm and 2084 nm. Near-infrared spectroscopy has also been used by Hall et al. (1988) for measuring theaflavin and moisture contents as well as for the prediction of black tea quality. Limited studies had been carried out to retrieve the tea quality parameters using remote sensing techniques. Dutta et al. (2011) used the multispectral satellite data for measuring the green and black tea quality and inferred that the NDVI can be effectively utilized for tea quality estimation. Bian et al. (2013) estimated the tea polyphenols, amino acids and soluble sugars at powder, leaf and canopy level using reflectance spectroscopy over an experimental tea plantation in China. They applied a partial least square regression technique on the noise removed reflectance data and the bands sensitive to polyphenol concentration were identified. The prediction accuracy of the leaf level polyphenol content was found to be lower than at canopy and powder level.

In the present study, tea polyphenol concentration was retrieved using the ground based hyperspectral data collected from Deha Tea Estate of Assam state, India. The stepwise discriminant analysis was adopted to identify the wavebands sensitive to polyphenol by effectively minimizing the age effect of bushes and management practices. A set of univariate and multivariate analyses have been carried out over the noise removed original spectral data and also on transformed spectral data, i.e. normalized reflectance and first derivative reflectance, to test the improvement in the prediction accuracy. The potential of principal component analysis on original and transformed dataset was also explored in combination with partial least square regression technique. The model with minimum prediction error was selected from various combinations and validated using independent data sets.

Materials and methods

Study area

The present study was conducted in Deha Tea Estate located in Jorhat district of Assam state, India. The garden is covered by the Survey of India Topomap no 83J-5, 6 and bounded by the coordinate 26°44′45″ N to 26°47′18″ N latitude and 94°20′35″ E to 94°22′25″ E longitude. Tea gardens are confined to the southern half of the district and constitute about 11% of the district. The climate of the district is favourable for tea cultivation with the temperature varying between 9 °C and 39 °C along with mean annual rainfall of 2244 mm. Major part of the district is characterized by plain interspersed by small hills. The southern part of the district extends to the foothills of Nagaland state and the soil here is extremely suitable for tea plantation. There are 2 administrative divisions in the tea estate (i) Deha division (consists of 29 sections) and (ii) Naginijan (consists of 22 sections). To account for the wide variability of tea polyphenol concentration, bushes of different age groups and farming practices (organic/inorganic) and cultivars viz. TS446 (old seed) and TV1 (clone) were selected. The study area along with the selected sections is given in Fig. 1. Details of the sections of the tea estate where observation were taken are given in Table 1.

Spectral measurements and pre-processing

A field campaign was carried out during 2nd week of March 2012 which matches with the onset of first flush of tea in the region. The field measurements were carried out using ASD FieldSpec® Pro (Analytical Spectral Devices, Inc.) Spectroradiometer that measures surface radiance in 971 spectral bands (347–2506 nm) with varying bandwidth of minimum 1 nm in the visible to as high as 3.8 nm in the SWIR region. The fibre optic probe was used for collection of spectra of first three leaves from the tip of the crown. Four locations in each section were chosen for hyperspectral observation and the average of 20 measurements from each location was recorded as representative of each location. Spectral measurements at each sampling location were preceded by a white spectral on reference panel for normalization of illumination and to convert radiance values into reflectance. The observation window was strictly kept between 11:00 and 13:00 h when the sun is almost overhead.

Several pre-processing techniques were applied before statistical analyses that include signature file overlap/matching, Savitzky and Golay (1964) filtering, reflectance normalization and first derivative of the reflectance. A Savitzky–Golay (SG) filter with five window size and 2nd order polynomial was adopted for noise minimization and retaining the shape of the original spectra. The normalization of the reflectance spectra was carried out on the smoothened reflectance data by using mean and standard deviation of each dataset to minimize the impact of illumination and time of data collection. The SG filtered spectra was transformed into first order derivative spectra as the prediction accuracy of the foliar biochemicals improves with transformed spectra (Ferwerda et al., 2006). The smoothed reflectance (SR), normalized reflectance (NR) and first derivative of reflectance (FDR) of the tea leaves with