Monitoring Photosynthetic Pigments of Shade-Grown Tea from Hyperspectral Reflectance

Monitoring Photosynthetic Pigments of Shade Grown Tea from Hyperspectral Reflectance

Rei Sonobe

Faculty of Agriculture, Shizuoka University, Shizuoka, Japan

Yuta Miura

Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka, Japan

Tomohito Sano and Hideki Horie

Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Shimada, Japan

Corresponding author

Rei sonobe (E-mail: reysnb@gmail.com)

Monitoring Photosynthetic Pigments of Shade Grown Tea from Hyperspectral Reflectance

The highest quality green tea is cultivated using shading treatments in Japan; however, shading can lead to early mortalities of tea due to excessive environmental stress. The allocation of photosynthetic pigments, chlorophyll a, b and carotenoids, could be a good indicator for evaluating production or environmental stress in plants; thus, developing an in-situ method to monitor photosynthetic pigments is useful for agricultural management. To assess the accuracy of the estimation of photosynthetic pigment contents with existing supervised learning models, four different approaches were compared including random forests, kernel-based extreme learning machine (KELM), deep belief nets and support vector machine. Overall, KELM had the highest performance with a root mean square error of 1.95 ± 0.36 μg cm^{-2}, 1.08 ± 0.11 μg cm^{-2} and 0.68 ± 0.10 μg cm^{-2} for estimating chlorophyll a, b and carotenoid contents, respectively.

Keywords: hyperspectral; machine learning; photosynthetic pigments; shade grown tea

1. Introduction

Green tea, which is extracted from *Camellia sinensis*, has stronger antioxidant abilities than black tea (Lee *et al.*, 2002, Ning *et al.*, 2017), and one major active constituent (epigallocatechin-3-gallate) could inhibit the semen-mediated enhancement of HIV infection (Hauber *et al.*, 2009). As a result, green tea has attracted a great deal of attention.

Appearance, aroma and taste are important factors for assessing the quality of tea (Chen *et al.*, 2008, Bian *et al.*, 2010), and shading of tea increases chlorophyll a content, which is important for evaluating appearance. This increase in chlorophyll a content is caused by reducing natural photosynthesis in the leaves; however, shading can lead to early mortalities of tea due to excessive environmental stresses. Therefore,

detecting environmental stress using field measurements is required for better tea tree management.

Chlorophyll pigments consist of two main types, a and b, and their concentrations relate closely to primary production because these pigments absorb sunlight and use their energy to synthesize carbohydrates from CO_2 and H_2O (Gitelson *et al.*, 2006). Chlorophyll a/b ratios increase sharply in a linear manner at low light intensity, but increase gradually and linearly at higher light intensities (Leong and Anderson, 1984), and the ratio is positively correlated with the amount of the core complex of photosystem II (Terashima and Hikosaka, 1995). Therefore, it is possible to use the ratio to predict the abundance of chlorophyll associated with photosystem I or II. A high ratio would mean less light-harvesting chlorophyll protein complex II and thus lower photosystem II chlorophyll content and higher photosystem I chlorophyll content (Leong and Anderson, 1984), which has been applied for evaluating the response to the changing environmental conditions around vegetation (Tanaka *et al.*, 2001, Hobe *et al.*, 2003, Chen *et al.*, 2010). Carotenoids are also involved in photoprotection and light collection in photosynthesis (DemmigAdams *et al.*, 1996). In addition, they help to protect unsaturated fatty acids, phospholipids and galactolipids from damage (Edge *et al.*, 1997). Therefore, the variation in total chlorophyll/carotenoid ratios could be a good indicator for evaluating environmental stress in plants (Hendry and Price, 1993).

While the SPAD-502 Leaf Chlorophyll Meter (Konica Minolta Inc.) has been widely used for field measurements of leaf chlorophyll content (Leon *et al.*, 2007, Bannari *et al.*, 2008, Lausch *et al.*, 2013), in-situ measurements of leaf carotenoid content are challenging. Spectrophotometric measurements using ultraviolet and visible (UV-VIS) spectroscopy or high-performance liquid chromatography (HPLC) measurements have been used (Bilger *et al.,* 1989). However, there are few applicable

techniques for the in-situ assessment of chlorophyll/carotenoid ratios because many methods are destructive.

In contrast, hyperspectral remote sensing offers an efficient way to monitor biochemical properties such as chlorophyll content (Zhang *et al.*, 2011, Lausch *et al.*, 2013, Sonobe and Wang, 2017). However, most available datasets are composed of measurements taken under relatively low light-stress conditions (e.g. the coefficients of linear regression models for estimating chlorophyll a content from carotenoid contents were 2.99 (for LOPEX dataset) to 3.45 (for HAWAII dataset) (Féret *et al.*, 2008). In Japan, the highest quality green tea is cultivated using shading treatments, which imposes environmental stress on vegetation and changes the allocation of chlorophyll a, b and carotenoids, and the coefficients of linear regression models for estimating chlorophyll a content from carotenoid contents sometimes exceed five (Sonobe *et al.*, 2018). Therefore, we evaluated estimates of three photosynthetic pigments using the realistic measurements under high stress conditions based on hyperspectral data.

Recently, machine learning algorithms have attracted attention as an approach to quantifying biochemical properties (Doktor *et al.*, 2014, Lv and Yan, 2014). The regression algorithm is also used to estimate the status of vegetation from hyperspectral reflectance. Random forests (RF) has been extremely successful as a general-purpose classification and regression method (Biau and Scornet, 2016) and it is a potential tool for assessing photosynthetic pigments contents. The support vector machine (SVM) has also been a very effective approach, and has been widely used with a Gaussian kernel function (Burges, 1998). In addition, recently developed extensions of machine learning including kernel-based extreme learning machine (KELM) and deep belief nets (DBN) were compared in this study.

The main objectives of this study were to (1) to examine the potential of hyperspectral data for the assessing three photosynthetic pigments under high stress conditions and (2) to identify which machine learning algorithm is most suitable as a regression model.

2. Materials and methods

2.1.Measurements and datasets

Our sampling sites were located at the Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Shimada, Japan. Shade treatment, which is practiced for the top grades (Figure 1), causes tea leaves to synthesise higher levels of chlorophyll and amino acids. To evaluate the influence of shading on environmental stress, tea trees were cultivated under four shading treatments (open-0 % shading, 35 % shading, 75 % shading and 90 % shading using black shade nets beginning 21 April 2017. Spectral reflectance and biochemical properties were measured on 1 and 11 May 2017. In total, 46 (8 samples for open-0 % shading, 12 samples for 35 % shading, 12 samples for 75 % shading and 14 samples for 90 % shading) and 60 measurements (15 samples for each treatment) were obtained on 1 and 11 May 2017, respectively.

<Figure 1>

Following leaf reflectance measurements, leaf discs were punched for pigment concentration measurements and analysed using dual-beam scanning ultraviolet-visible spectrophotometers (UV-1280, Shimadzu, Japan). Wellburn's method (Wellburn, 1994) was applied to calculate chlorophyll and carotenoid contents after absorption. The

equations used in this method for quantifying chlorophyll a (Chla, μg ml^{-1}), chlorophyll b (Chlb, μg ml^{-1}) and carotenoid (Car, μg ml^{-1}) in dimethyl-formamide extracts are as follows:

$$\mathrm{Car} = (1000A_{480} - 1.12\mathrm{Chla} - 34.07\mathrm{Chlb})/245 \quad (1)$$

$$\mathrm{Chla} = 12A_{663.8} - 3.11A_{646.8} \quad (2)$$

$$\mathrm{Chlb} = 20.78A_{646.8} - 4.88A_{663.8} \quad (3)$$

where A is the absorbance and the subscripts are the wavelength (nm).

Reflectance data were obtained using a spectrometer (FieldSpec4, Analytical Spectral Devices Inc., USA) with a leaf clip. The device has three detectors including visible and near-infrared (VNIR), Short Wave Infra-Red (SWIR) 1 and SWIR 2, and some inherent variation in detector sensitivities often causes differences in the spectral drifts at two wavelength locations (1000 and 1800 nm). Thus, the splice correction function of ViewSpec Pro Software (Analytical Spectral Devices Inc., USA) was applied to modify these connections.

2.2.Regression model

A stratified random-sampling approach was applied to select the samples used for training (50 %), validation (25 %) and test data (25 %) (Hastie *et al.*, 2009), and the stratified random-sampling procedure was repeated ten times for more robust results.

Variable selection is useful for removing non-informative variables to obtain better and simpler prediction models (Villar *et al.*, 2017). Villar et al. (2017) evaluated the performances of three variable selection techniques including Martens uncertainty test, interval partial least square regression and genetic algorithm (GA), which is adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics, using a visible-near infrared sensor system and showed that GA was the

best among them. In this study, GA simulated the survival of the fittest among individuals over consecutive five generation for estimating chlorophyll and carotenoid contents. Each generation were composed of a population of character strings (i.e. band combinations) that were analogous to the chromosome that we see in our DNA. After that the individuals were made to go through a process of evolution and the best band combination would be revealed. Following Villar et al. (2017), GA was also applied to select several wavelengths using training data in this study.

Next, the regression models were generated using the supervised learning models RF, KELM, DBN and SVM.

RF is an ensemble learning technique that builds multiple trees based on random bootstrapped samples of the training data (Breiman, 2001). Nodes are split using the best split variable from a group of randomly selected variables (Liaw and Wiener, 2002). This strategy provides robustness against over-fitting and handles thousands of dependent and independent input variables without variable deletion. The strong performances of this algorithm have been documented in previous studies (Montillo and Ling, 2009, Johansson *et al.*, 2014), so it was used as a benchmark in this study. RF regression was implemented using R version 3.3.1 (R Core Team 2016) and the 'randomForest' package (Liaw and Wiener, 2002).

Extreme learning machine (ELM) is a single hidden layer feed-forward neural network that has a fixed hidden layer, which is composed of a vast number of nonlinear nodes, and the hidden layer bias is defined randomly (Huang *et al.*, 2006). ELM has been successfully used in many practical tasks, such as prediction, fault diagnosis, recognition, classification and signal processing (Li *et al.*, 2016). Furthermore, the kernel trick has been applied to ELM instead of attempting to fit a non-linear model (Huang *et al.*, 2012). We applied the Radial Basis Function (RBF) kernel and the

regulation coefficient (*Cr*) and the kernel parameter (*Kp*) were tuned. KELM was applied using MATLAB and Statistics Toolbox Release 2016a (The MathWorks, Inc., Natick, Massachusetts, United States) and the source code was downloaded from http://www.ntu.edu.sg/home/egbhuang/.

Restricted Boltzmann machines (RBMs) are two-layer neural networks including the first layer called the visible, or input layer, and the second one called the hidden layer. DBN consists of a multi-layer unsupervised RBMs and a layer of a supervised back-propagation network (Hinton *et al.*, 2006). Dropout refers to ignoring units and was also used during the training phase, as it was shown to contribute good predictions. DBN regression was implemented using R version 3.3.1 and the 'darch' package (Drees, 2013).

SVM is an effective machine learning method based on statistical learning theory that been successfully applied to solve the problem of high dimension and local minima (Ding *et al.*, 2016). We applied the Gaussian radial basis function (RBF) kernel using R version 3.3.1 and the 'e1071' package (Meyer *et al.*, 2017). It possesses two hyperparameters including the regularization parameter *C* and the kernel bandwidth γ. High *C* values lead to high penalties for inseparable points, which may result in overfitting. In contrast, low *C* values lead to under-fitting. The γ value defines the reach of a single training example, with low values indicating ‘far’ and high values indicating ‘close’ reach.

For optimizing the hyperparameters of these machine learning algorithms, Bayesian optimisation, which is a framework used to optimise hyperparameters of noisy, expansive black-box functions and defines a principled approach to modelling uncertainty, was applied with the Gaussian process (Bergstra and Bengio, 2012, Snoek *et al.*, 2015).

2.3.Statistical criteria

Statistical criteria to evaluate the performance of these models relied on the root mean square error (RMSE, Equation (4)).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(\widehat{y_i} - y_i)^2} \qquad (4)$$

where n is number of samples, y_i is the measured value and $\widehat{y_i}$ is the estimated value.

Further, a stepwise linear discriminant analysis (Draper, 1998, Burns and Burns, 2008) was conducted to identify which wavelengths had significant differences ($p < 0.05$). This technique for selecting suitable predictor wavelengths based on a multiple regression model was confirmed among the four shading treatments. Although there are models including forward, backward and combination, a combination of forward and backward stepwise regression was adopted in this study.

The sensitivity of the regression models was assessed based on a data-based sensitivity analysis (DSA). DSA is similar to a computationally efficient one-dimensional sensitivity analysis (Kewley *et al.*, 2000) where only one input is changed at a time and the others are kept at their average values. However, DSA uses several training samples instead of a baseline vector (Cortez and Embrechts, 2013) and then it performs a pure black box use of the fitted models by querying the fitted models with sensitivity samples and recording their responses.

3. Results and Discussion

3.1. Chlorophyll and Carotenoid Contents of Each Treatment

Figure 2 shows the boxplots of chlorophyll a, b and carotenoid contents of the different shading treatments. The contents in leaf area (cm^2) ranged from 7.81-29.53 μg for chlorophyll a, 1.43-8.98 μg for chlorophyll b and 3.46-7.49 μg for carotenoids on 1 May 2017, and from 18.85-44.62 μg for chlorophyll a, 5.73-16.39 μg for chlorophyll b and 6.16-11.29 μg for carotenoids on 11 May 2017.

Although the mean contents of chlorophyll a and b were greater with more shading, the differences were not significant for 35 % - 90 % shading and 75 % - 90 % shading for either date ($p < 0.05$, based on the Tukey-Kramer test). The differences in carotenoid contents were not significant among shading treatments ($p < 0.05$, based on the Tukey-Kramer test).

While the chlorophyll/carotenoid ratios ranged from 2.56 to 5.53 on 1 May, they ranged from 4.04 to 7.63 on 11 May and the increases were confirmed during the experiment. The significant differences were confirmed, except for the combination between 75 % and 90 % shading on both dates ($p < 0.05$, based on the Tukey-Kramer test). Shading treatment results in higher leaf protein content and thicker leaves (Poorter *et al.*, 2006). Generally, shaded leaves contain more photosynthetic pigments than sunlit leaves, and to harvest more light and nitrogen, the shaded leaves increase chlorophyll a content (Suzuki and Shioi, 2003). Consequently, the chlorophyll/carotenoid ratio would be increased.

< Figure 2>

3.2.Spectral Reflectance of Different Treatments

The mean reflectance and standard deviations of each treatment are shown in Figure 3. The reflectance near the green peak was larger when the shading treatments were lighter and this tendency was clear in measurements acquired on 1 May. In addition, more shading made the reflectance lower at the red edge inflection point (REIP). On the other hand, there were little difference in the reflectance near the start of the red edge (near 680 nm) and this tendency was clear in measurements acquired on 11 May. The shifts of the green peak and REIP were confirmed within the four shading treatments during the experimental period (Figure 4).

The results of a stepwise discriminant analysis ($p < 0.05$) shows that reflectance at 519, 569, 686, 797 and 798 nm was useful for identification of the four shading treatments on 1 May (overall accuracy = 0.739), while reflectance at 564 and 701 nm was useful for treatments on 11 May (overall accuracy = 0.650).

Figure 5 represents the correlations of each wavelength of spectral reflectance with chlorophyll a, b or carotenoid content. For all pigment contents, negative correlations were confirmed over visible to REIP and two troughs were identified near 510 nm (R = -0.87 at 514 nm for chlorophyll a, -0.84 at 521 nm for chlorophyll b and -0.83 at 504 nm for carotenoid) and 740 nm (R = -0.96 at 741 nm for chlorophyll a, -0.94 at 739 nm for chlorophyll b and -0.90 at 746 nm for carotenoid). Although the correlation coefficients near 680 nm were relatively weak for chlorophyll a and b, they were obscure for carotenoids. Table 1 lists the selected wavelengths for each photosynthetic pigment. The numbers of wavelengths were 11, 17 and 13 for chlorophyll a, b and carotenoid, respectively. The reflectance values at 740 – 750 nm were not selected for chlorophyll a and b, even though they had high correlation coefficients. Some reflectance values at 450–500 nm were selected, despite the tendencies of reflectance with the shape treatments to be visually obscure (Figure 3). In

the previous studies, these values have been applied to emphasize the reflectance at 680-690 nm for estimating total chlorophyll contents (Penuelas *et al.*, 1995, Lichtenthaler *et al.*, 1996) and carotenoid contents (Blackburn, 1998). In this case, these values were also used as references.

< Figure 3>

< Figure 4>

< Figure 5>

<Table 1>

3.3.Accuracy Validation

Accuracy results are tabulated in Table 2. RMSEs of chlorophyll a content estimations were 3.87 ± 0.50 μg cm^{-2} for RF, 1.95 ± 0.36 μg cm^{-2} for KELM, 3.41 ± 0.73 μg cm^{-2} for DBN and 3.20 ± 0.98 μg cm^{-2} for SVM. KELM had the best performance, although SVM had a better RMSE for round 2. Similar results were confirmed in chlorophyll b content estimations; however, RF was the best algorithm in some cases (ex. round 2). For estimating carotenoids, KELM was also the best algorithm; although, it was the worst in round 7. Overall, KELM was the most robust algorithm for estimating the three photosynthetic pigment contents of shade grown tea.

<Table 2>

3.4.Comparison between KELM and RF

The strong performances of RF have been documented in previous studies (Montilla *et al.*, 2009, Johansson *et al.*, 2014). Therefore, this algorithm was used as a benchmark, and its sensitivity against reflectance was compared with that of KELM. Figure 6 shows the results of the DSA. For RF, the reflectance at the green peak had a greater influence on chlorophyll a content estimations than on chlorophyll b estimations; this tendency was obscure for the other algorithm. Figure 1 shows that chlorophyll a and b contents decreased with higher shading treatments and the extensive dependence on the green peak and red edge seemed useful for good estimations. For estimating chlorophyll b contents, the importance of the reflectance at 708, 713 and 753 nm occupied 53.3 % when RF was applied and the performance of RF for chlorophyll b content estimation was the second best averaged RMSE. However, the shifts of the green peak and REIP should be considered and that could be a risky strategy. Half of the estimation results had the worst accuracies (rounds 1, 4, 5, 7 and 10), although it sometimes gave precise estimations (rounds 2, 9 and 10). For estimating chlorophyll a, the influences were more serious and the results of RF, whose importance of the reflectance at 523, 700 and 769 nm occupied 38.5 %, were worse than those of the others.

The reflectance at the green peak and the end of the red edge also has great potential for estimating carotenoid content for RF; however, the influence on KELM regression was not clear. Although the differences in carotenoid contents were not significant among the four shading treatments, the shifts in reflectance were confirmed at the green peak and red edge (Figure 3). That may mislead the estimation based on RF, since the importance of the reflectance at only two wavelengths (start of the green peak and end of the red edge) occupied 48.1 %.

< Figure 6>

The chlorophyll to carotenoid ratio is an indicator for environmental stress in plants (Hendry and Price, 1993). The RMSE values of the estimated chlorophyll/carotenoid ratio for each machine learning algorithm were evaluated (Table 3). KELM was the best algorithm for assessing the chlorophyll to carotenoid ratio with an RMSE of 0.36 ± 0.08. KELM shows great potential for detecting environmental stress based on field measurements, which is needed for better tea tree management.

<Table 3>

4. Conclusions

We examined the relationships between reflectance and three photosynthetic pigment contents (chlorophyll a, b and carotenoid) using measurements from shade grown tea including measurements of plants exposed to high environmental stress. This study compared photosynthetic pigment content estimation accuracies among four algorithms (RF, KELM, DBN and SVM) from hyperspectral data.

Among them, KELM yielded the most accurate estimation with 1.95 ± 0.36 μg cm^{-2}, 1.08 ± 0.11 μg cm^{-2} and 0.68 ± 0.10 μg cm^{-2} for chlorophyll a, b and carotenoid, respectively. The results of this study demonstrate hyperspectral data obtained from FieldSpec4 can be used for in-situ measurements of leaf photosynthetic pigment contents.

As the results, reflectance values at some wavelengths allowed for evaluation the risk of early mortalities as well as the appearance and ingredients of green tea before their cultivation. The results will also improve the evaluation of physiological properties of vegetation and usability for agricultural management using hyperspectral reflectance.

Funding

This work was supported by the science and technology research promotion program for agriculture, forestry, fisheries and food industry [grant number 27015C].

Disclosure statement

No potential conflicts of interest are reported by the authors.

References

Bannari, A., Khurshid, K.S., Staenz, K. & Schwarz, J., 2008. Potential of hyperion EO-1 hyperspectral data for wheat crop chlorophyll content estimation. *Canadian Journal of Remote Sensing,* 34**,** S139-S157.

Bergstra, J. & Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research,* 13**,** 281-305.

Bian, M., Skidmore, A.K., Schlerf, M., Fei, T., Liu, Y.F. & Wang, T.J., 2010. Reflectance Spectroscopy of Biochemical Components as Indicators of Tea (Camellia Sinensis) Quality. *Photogrammetric Engineering and Remote Sensing,* 76**,** 1385-1392.

Biau, G. & Scornet, E., 2016. A random forest guided tour. *Test,* 25**,** 197-227.

Bilger, W., Bjorkman, O. and Thayer, S. S. 1989. Light-induced spectral absorbance changes in relation to photosynthesis and epoxidation state of xanthophyll cycle components in cotton leaves. *Plant Physiology*, 91, 542-551.

Blackburn, G.A., 1998. Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment,* 66**,** 273-285.

Breiman, L., 2001. Random forests. *Machine Learning,* 45**,** 5-32.

Burges, C.J.C., 1998. A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery,* 2**,** 121-167.

Burns, R.P. & Burns, R., 2008. *Business research methods and statistics using SPSS*: SAGE Publications.

Chen, Q.S., Zhao, J.W. & Vittayapadung, S., 2008. Identification of the green tea grade level using electronic tongue and pattern recognition. *Food Research International,* 41**,** 500-504.

Chen, W.M., Jin, N., Shi, Y., Su, Y.Q., Fei, B.J., Li, W., Qiao, D.R. & Cao, Y., 2010. Coordinate expression of light-harvesting chlorophyll a/b gene family of photosystem II and chlorophyll a oxygenase gene regulated by salt-induced phosphorylation in Dunaliella salina. *Photosynthetica,* 48**,** 355-360.

Cortez, P. & Embrechts, M.J., 2013. Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences,* 225**,** 1-17.

Demmigadams, B., Gilmore, A.M. & Adams, W.W., 1996. Carotenoids .3. In vivo functions of carotenoids in higher plants. *Faseb Journal,* 10**,** 403-412.

Ding, S.F., Shi, Z.Z., Tao, D.C. & An, B., 2016. Recent advances in Support Vector Machines. *Neurocomputing,* 211**,** 1-3.

Doktor, D., Lausch, A., Spengler, D. & Thurner, M., 2014. Extraction of Plant Physiological Status from Hyperspectral Signatures Using Machine Learning Methods. *Remote Sensing,* 6**,** 12247-12274.

Draper, N., H, 1998. *Applied regression analysis (Wiley Series in Probability and Statistics)*: Wiley-Interscience.

Drees, M., 2013. Implementierung und Analyse von tiefen Architekturen in R. Fachhochschule Dortmund.

Edge, R., Mcgarvey, D.J. & Truscott, T.G., 1997. The carotenoids as anti-oxidants - a review. *Journal of Photochemistry and Photobiology B-Biology,* 41**,** 189-200.

Féret, J.-B., Francois, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., Le Maire, G. & Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment,* 112**,** 3030-3043.

Gitelson, A.A., Vina, A., Verma, S.B., Rundquist, D.C., Arkebauer, T.J., Keydan, G., Leavitt, B., Ciganda, V., Burba, G.G. & Suyker, A.E., 2006. Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research-Atmospheres,* 111.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition* The United States: Springer-Verlag New York.

Hauber, I., Hohenberg, H., Holstermann, B., Hunstein, W. & Hauber, J., 2009. The main green tea polyphenol epigallocatechin-3-gallate counteracts semen-mediated enhancement of HIV infection. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 9033-9038.

Hendry, G.A.F. & Price, A.H., 1993. *Stress Indicators: Chlorophylls and Carotenoids*.

Hinton, G.E., Osindero, S. & Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural Computation,* 18**,** 1527-1554.

Hobe, S., Fey, H., Rogl, H. & Paulsen, H., 2003. Determination of relative chlorophyll binding affinities in the major light-harvesting chlorophyll a/b complex. *Journal of Biological Chemistry,* 278**,** 5912-5919.

Huang, G.B., Zhou, H.M., Ding, X.J. & Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics,* 42**,** 513-529.

Huang, G.B., Zhu, Q.Y. & Siew, C.K., 2006. Extreme learning machine: Theory and applications. *Neurocomputing,* 70**,** 489-501.

Johansson, U., Bostrom, H., Lofstrom, T. & Linusson, H., 2014. Regression conformal prediction with random forests. *Machine Learning,* 97**,** 155-176.

Kewley, R.H., Embrechts, M.J. & Breneman, C., 2000. Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks,* 11**,** 668-679.

Lausch, A., Pause, M., Schmidt, A., Salbach, C., Gwillym-Margianto, S. & Merbach, I., 2013. Temporal hyperspectral monitoring of chlorophyll, LAI, and water content of barley during a growing season. *Canadian Journal of Remote Sensing,* 39**,** 191-207.

Lee, K.W., Lee, H.J. & Lee, C.Y., 2002. Antioxidant activity of black tea vs. green tea. *Journal of Nutrition,* 132**,** 785-785.

Leon, A.P., Vina, S.Z., Frezza, D., Chaves, A. & Chiesa, A., 2007. Estimation of chlorophyll contents by correlations between SPAD-502 meter and chroma meter in butterhead lettuce. *Communications in Soil Science and Plant Analysis,* 38**,** 2877-2885.

Leong, T.Y. & Anderson, J.M., 1984. Adaptation of the thylakoid membranes of pea chloroplasts to light intensities. I. Study on the distribution of chlorophyll-protein complexes. *Photosynthesis research,* 5**,** 105-15.

Li, X.D., Mao, W.J. & Jiang, W., 2016. Multiple-kernel-learning-based extreme learning machine for classification design. *Neural Computing & Applications,* 27**,** 175-184.

Liaw, A. & Wiener, M., 2002. Classification and regression by random Forest. *R News,* 2**,** 18-22.

Lichtenthaler, H.K., Lang, M., Sowinska, M., Heisel, F. & Miehe, J.A., 1996. Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology,* 148**,** 599-612.

Lv, J. & Yan, Z.G., 2014. Estimating Leaf Chlorophyll Concentration in Soybean Using Random Forests and Field Imaging Spectroscopyed. *3rd International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Beijing, China, 1-4.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C. & Lin, C.-C., 2017. *Misc Functions of the Department of Statistics, Probability* [online]. [Accessed Access Date 2017].

Montillo, A. & Ling, H.B., 2009. Age regression from faces using random forests. *2009 16th IEEE International Conference on Image Processing,* 2437-2440.

Ning, J.M., Hou, G.G., Sun, J.J., Wan, X.C. & Dubat, A., 2017. Effect of green tea powder on the quality attributes and antioxidant activity of whole-wheat flour pan bread. *Lwt-Food Science and Technology,* 79**,** 342-348.

Penuelas, J., Filella, I., Lloret, P., Munoz, F. & Vilajeliu, M., 1995. Reflectance Assessment of Mite Effects on Apple-Trees. *International Journal of Remote Sensing,* 16**,** 2727-2733.

Poorter, H., Pepin, S., Rijkers, T., De Jong, Y., Evans, J.R. & Korner, C., 2006. Construction costs, chemical composition and payback time of high- and low-irradiance leaves. *Journal of Experimental Botany,* 57**,** 355-371.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Accessed 1 June 2016. http://www.R-project.org/.

Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M.M.A., Prabhat & Adams, R.P., 2015. Scalable Bayesian optimization using deep neural networks. *the 32nd International Conference on Machine Learning (ICML)*, Paris, 2171–2180.

Sonobe, R., Miura, Y., Sano, T. & Horie, H., 2018. Estimating leaf carotenoid contents of shade grown tea using hyperspectral indices and PROSPECT-D inversion. *International Journal of Remote Sensing*, 39, 1306-1320

Sonobe, R. & Wang, Q., 2017. Towards a universal hyperspectral index to assess chlorophyll content in deciduous forests. *Remote Sensing,* 9**,** 191.

Suzuki, Y. & Shioi, Y., 2003. Identification of chlorophylls and carotenoids in major teas by high-performance liquid chromatography with photodiode array detection. *Journal of Agricultural and Food Chemistry,* 51**,** 5307-5314.

Tanaka, R., Koshino, Y., Sawa, S., Ishiguro, S., Okada, K. & Tanaka, A., 2001. Overexpression of chlorophyllide a oxygenase (CAO) enlarges the antenna size of photosystem II in Arabidopsis thaliana. *Plant Journal,* 26**,** 365-373.

Terashima, I. & Hikosaka, K., 1995. Comparative ecophysiology of leaf and canopy photosynthesis. *Plant Cell and Environment,* 18**,** 1111-1128.

Villar, A., Vadillo, J., Santos, J.I., Gorritxategi, E., Mabe, J., Arnaiz, A. & Fernandez, L.A., 2017. Cider fermentation process monitoring by Vis-NIR sensor system and chemometrics. *Food Chemistry,* 221**,** 100-106.

Wellburn, A.R., 1994. The spectral determination of chlorophyll a and chlorophyll b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *Journal of Plant Physiology,* 144**,** 307-313.

Zhang, K.W., Hu, B.X., Wang, J.G., Pattey, E. & Smith, A.M., 2011. Improving the retrieval of the biophysical parameters of vegetation canopies using the contribution index. *Canadian Journal of Remote Sensing,* 37**,** 643-652.

Table

Table 1. Selected wavelengths (nm) based on GA.

Photosynthetic pigment	Wavelength (nm)
Chlorophyll a	469, 474, 488, 502, 523, 581, 636, 696, 700, 769, 774
Chlorophyll b	451, 462, 493, 526, 572, 581, 584, 614, 631, 634, 652, 672, 708, 713, 753, 770, 771
Carotenoid	476, 488, 503, 521, 539, 540, 548, 661, 689, 700, 712, 736, 743

Table 2. RMSEs (μg cm^{-2}) of estimation results for each machine learning algorithm.

Repetition	Chlorophyll a				Chlorophyll b				Carotenoid			
	RF	KELM	DBN	SVM	RF	KELM	DBN	SVM	RF	KELM	DBN	SVM
Round01	2.94	1.15	2.65	5.49	1.34	0.96	1.19	1.25	0.74	0.53	0.75	0.67
Round02	3.81	2.38	2.58	2.02	1.07	1.15	2.21	1.15	0.68	0.56	0.95	0.61
Round03	3.38	2.26	4.65	4.06	1.53	1.23	1.70	1.55	0.81	0.82	0.94	0.95
Round04	3.71	1.96	4.38	3.32	1.63	1.11	1.35	1.46	1.10	0.78	0.83	0.99
Round05	4.52	1.69	3.98	1.89	1.68	0.98	1.30	1.25	0.70	0.77	0.59	0.58
Round06	4.36	2.41	2.64	3.49	1.56	1.30	1.61	1.34	0.89	0.73	0.95	0.79
Round07	4.61	2.04	2.74	2.87	1.31	1.00	1.20	1.23	0.97	0.54	0.84	0.55
Round08	3.55	1.77	3.63	3.18	1.40	0.97	1.30	1.88	0.70	0.64	0.81	0.61
Round09	4.06	1.70	3.78	2.93	1.42	1.07	1.54	2.78	0.86	0.73	0.77	0.56
Round10	3.71	2.09	3.11	2.74	1.56	1.01	1.45	1.50	0.78	0.67	0.72	1.24
Mean	3.87	1.95	3.41	3.20	1.45	1.08	1.48	1.54	0.82	0.68	0.81	0.76
Standard deviation	0.50	0.36	0.73	0.98	0.17	0.11	0.29	0.46	0.13	0.10	0.12	0.38

Table 3. RMSEs of chlorophyll/carotenoid ratio for each machine learning algorithm.

	Measured value			RMSE			
	Max.	Min.	Standard deviation	RF	KELM	SVM	DBN
Round01	5.83	2.95	0.79	0.62	0.27	0.67	0.35
Round02	5.99	2.67	0.81	0.52	0.36	0.50	0.42
Round03	5.93	2.67	0.86	0.53	0.29	0.87	0.79
Round04	5.99	2.67	0.78	0.73	0.40	0.57	0.86
Round05	5.99	2.95	0.86	0.66	0.55	0.35	0.60
Round06	5.89	2.96	0.80	0.57	0.35	0.56	0.70
Round07	5.83	2.95	0.78	0.66	0.35	0.58	0.61
Round08	5.83	3.30	0.73	0.55	0.27	0.79	0.46
Round09	5.93	2.96	0.74	0.61	0.39	0.52	0.49
Round10	5.89	2.56	0.90	0.70	0.40	0.68	0.70
Mean	5.91	2.86	0.80	0.61	0.36	0.61	0.60
Standard deviation	0.06	0.21	0.05	0.07	0.08	0.14	0.16

Figures

Figure 1. Shading treatments conducted in this study.

Figure 2. Boxplots of (a) chlorophyll a, (b) chlorophyll b, (c) carotenoid content and (d) the ratio of total chlorophyll to carotenoid.

Figure 3. Mean reflectance spectra (solid lines) and standard deviations (thinner zones) on (a) 1 and (b) 11 May.

Figure 4. Shifts of (a) green peak and (b) red edge positions. The central solid line represents the median; bars extend to the 95% confidence limits; dots represent outliers.

Figure 5. Correlations between reflectance and chlorophyll a, b or carotenoid content for all measurements.

Figure 6. DSA results for RF and KELM regression models.