Using spectral reflectance to estimate leaf chlorophyll content of tea with shading treatments

Title

Using spectral reflectance to estimate leaf chlorophyll content of tea with shading treatments

Author names and affiliations

Rei Sonobe[a], Tomohito Sano[b] and Hideki Horie[c]

[a] Faculty of Agriculture, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan

[b]Headquarters, National Agriculture and Food Research Organization, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8517, Japan

[c]Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, 2769 Shishidoi, Kanaya, Shimada, Shizuoka 428-8501, Japan

Corresponding author

Rei Sonobe

reysnb@gmail.com

Abstract

Some stresses are utilised to improve qualities of agricultural products. Low light stress increases the chlorophyll content of tea leaves, which improves appearance. Although chlorophyll content estimation is one of the most common applications of hyperspectral remote sensing, previous studies were based on measurements under relatively low stress conditions. In this study, two methods, machine learning algorithms and the inversion of a radiative transfer model, were evaluated using measurements from tea leaves with shading treatments. According to the ratio of performance to deviation (RPD), PROSPECT-D inversion (RPD=1.71-2.31) had the potential for quantifying chlorophyll content; although, it required some improvements. Overall, the regression models based on machine learning had high performances. The kernel-based extreme learning machine had the highest performance with a root mean square error of 3.04 ± 0.52 μg cm^{-2} and RPD values from 3.38 to 5.92 for the test set, which was used for assessing generalisation error.

Keywords: chlorophyll; green tea; vegetation indices; machine learning; PROSPECT-D

1. Introduction

Green tea is a very healthy beverage because its consumption is associated with reduced mortality, and it has attracted a great deal of attention (Kuriyama *et al.* 2006). Green tea-flavoured sweets have even become popular. Some techniques have been developed for increasing chlorophyll content, which is important for improving tea leaf appearance. Chlorophyll content is strongly related to the colour of dry tea leaves (Wang *et al.* 2004) and the flavour of tea is principally

41 determined by chemical components. Chlorophyll content is positively correlated with the total
42 quality score as well as the scores for appearance and infused leaf (Wang *et al.* 2010). Therefore,
43 various methods are used to increase the chlorophyll content of tea leaves during growth (Lee *et al.*
44 2013). The control of light transmission by shade treatment is the most effective method for
45 increasing chlorophyll content in tea plants (De Costa *et al.* 2007) and shading nets (70–95%
46 shading) have been used in Shizuoka, Japan to increase the chlorophyll content of tea leaves and
47 to improve appearance (Sonobe *et al.* 2018a). However, excessive shading tea can lead to early
48 mortalities due to the excessive environmental stresses caused by reducing natural photosynthesis
49 in the leaves. Both quantifying chlorophyll contents and detecting environmental stresses using
50 field measurements are required for better tea tree management, and no applicable approach has
51 been established.
52
53 As destructive methods, spectrophotometric measurements using ultraviolet and visible (UV-VIS)
54 spectroscopy and high-performance liquid chromatography (HPLC) measurements have been used
55 to quantify pigment content in leaves (Prado-Cabrero *et al.* 2016). However, these techniques are
56 expensive, labour-intensive and not always applicable for in-situ measurements. Alternately, the
57 SPAD-502 Leaf Chlorophyll Meter (Konica Minolta Inc.) has been used for field measurements of
58 leaf chlorophyll content (le Maire *et al.* 2004; Elarab *et al.* 2015). However, light intensity also
59 influences leaf thickness (Murchie *et al.* 2005) and that makes the output of the meter obscure
60 (Yamamoto *et al.* 2002). In contrast, remote sensing using hyperspectral reflectance has been used
61 to evaluate biochemical properties (Whetton *et al.* 2018), especially chlorophyll content
62 estimation, which has received special attention since chlorophyll pigments closely relate to

63 protective activity against a variety of degenerative diseases as well as the photosynthetic process
64 (Korus 2013). Furthermore, because remote sensing is a non-destructive method that can cover
65 large areas and reflect the spatial variability of crop canopies using sensors mounted on airborne
66 drones or satellites, this technique is useful for improving fertiliser management (Gabriel *et al.*
67 2017).
68
69 The numerical inversion of radiative transfer models (RTMs) has been proposed to estimate
70 chlorophyll content using hyperspectral remote sensing (Li *et al.* 2015; Masemola *et al.* 2016).
71 PROpriétés SPECTrales (PROSPECT) is one of the most famous RTMs and has been widely used
72 to assess the biochemical properties of broadleaf species and herbs (Jacquemoud *et al.* 1996; Féret
73 *et al.* 2008; Hernandez-Clemente *et al.* 2014; Sonobe *et al.* 2018b; Sun *et al.* 2018). The latest
74 version, PROSPECT-D, has an improved ability to estimate pigment content (Féret *et al.* 2017).
75 Although le Maire et al. (2004) pointed out that the previous versions of PROSPECT were not
76 accurate enough to evaluate broad leaf chlorophyll content, this version has not been fully
77 evaluated.
78
79 Another recent option for estimating chlorophyll content from hyperspectral reflectance is based
80 on machine learning algorithms (Liang *et al.* 2016; Chemura *et al.* 2017). Random forests (RF) is
81 specifically mentioned as a successful classification and regression method (Biau and Scornet
82 2016), and has been widely used for evaluating the aboveground biomass of C3 and C4 grasses
83 (Shoko *et al.* 2018). Support vector machine (SVM) has also been a very effective approach and is
84 appropriate to express the relationship between reflectance and leaf water status (Das *et al.* 2017).

In addition, the high performances of kernel-based extreme learning machine (KELM) have been shown in some previous studies for solving regression problems (Sonobe et al., 2018a). Therefore, the machine learning algorithms RF, SVM and KELM were applied to estimate the chlorophyll content of shade grown tea from hyperspectral reflectance. Notably, the disadvantages of machine learning algorithms are that they require training data for prediction and not enough training data leads to overfitting and the models could be unsuitable. In this study, the machine learning algorithms which possess only two hyperparameters were evaluated.

Vegetation indices have also been widely used to emphasise the features of vegetation (Sonobe *et al.* 2018c) and a number of vegetation indices have been developed for evaluating chlorophyll content. Most vegetation indices for chlorophyll content are based on wavelengths ranging from 400 to 860 nm, which covers photosynthetically active radiation. There are reflectance value or derivative-based indices and feature-based indices, mainly on the properties of the red edge. However, most indices are only applicable to the specific species or specific leaf types, such as sunlit or shaded leaves, from which they were developed (Sonobe and Wang 2017a). In this study, regression models using vegetation indices were evaluated as well as regression models based on original reflectance and their accuracies were compared.

In leaves, there are two types of chlorophyll pigments (chlorophyll a and b) and the chlorophyll a/b ratio is positively correlated with the ratio of photosystem II cores to light harvesting chlorophyll-protein complex (Terashima and Hikosaka 1995). As a result, the cultivation using shading treatments imposes environmental stress on vegetation and changes the balance among

chlorophyll a and b contents. However, some previous studies have used datasets 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 content were 2.99 (Hosgood *et al.* 1994) to 3.45 (Féret *et al.* 2008)). Therefore, some approaches in previous studies for estimating chlorophyll content are not valid for evaluating the chlorophyll content of shade grown tea, and these approaches were evaluated in this study.

The objective of this study was to examine the potential of hyperspectral remote sensing approaches including radiative transfer model inversion and machine learning algorithms for quantifying chlorophyll content of tea that was grown under low-light stress.

2. Materials and methods

2.1. Measurements and datasets

The first flush of leaves, which are harvested from mid-April to mid-May, have the highest quality and, therefore, we focused on this period. The experiments were conducted at the Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Shimada, Japan. Daily temperatures and precipitation varied between 12.5–19.2 °C and 0–17.5 mm, respectively, during the experiment (Japan Meteorological Agency, 2017). Four shading treatments were conducted using no net (0% shading), shading net #410 (35% shading), #1210 (75% shading) and #1220 (90% shading) (Dio Chemicals, Ltd., Japan) to assess the influence of shading on tea chlorophyll content from 21 April 2017 to 11 May 2017.

Forty-six samples (8 samples for 0% shading, 12 samples for 35% shading, 12 samples for 75%

130 shading and 14 samples for 90% shading) and 60 measurements (15 samples for each treatment)
131 were collected on 1 and 11 May 2017, respectively. After sampling, we used the spectral
132 reflectance and biochemical properties including chlorophyll a, b and carotenoid content for 106
133 leaf samples.
134
135 A spectrometer (FieldSpec4, Analytical Spectral Devices Inc., USA) with three detectors (VNIR,
136 SWIR 1 and SWIR 2) was used to obtain reflectance data with a leaf clip. The drifts depended on
137 some inherent variation in detector sensitivities and were confirmed at these connections (i.e. the
138 wavelengths of 1000 and 1800 nm). Thus, the splice correction function was applied to modify
139 these connections using ViewSpec Pro Software (Analytical Spectral Devices Inc., USA).
140
141 Leaf discs were used for pigment concentration measurements in dimethyl-formamide extracts
142 using dual-beam scanning ultraviolet-visible spectrophotometers (UV-1280, Shimadzu, Japan).
143 The following equations were used to quantify chlorophyll content (Wellburn 1994):

144 $$\text{Chlorophyll} = \text{Chlorophyll a} + \text{Chlorophyll b} \quad (1)$$

145 $$\text{Chlorophyll a} = 12A_{663.8} - 3.11A_{646.8} \quad (2)$$

146 $$\text{Chlorophyll b} = 20.78A_{646.8} - 4.88A_{663.8} \quad (3)$$

147 where A is the absorbance and the subscripts are the wavelength (nm).
148
149 2.2. Vegetation indices
150 Vegetation indices calculated from various remote sensing sensors are effective for removing
151 variability caused by other features, such as soil background and atmospheric conditions

(Blackburn and Steele 1999). They are also effecting for reducing the data saturation problem (Mutanga and Skidmore 2004) and in quantitative and qualitative evaluations of vegetation cover, vigour and growth dynamics, among other applications, that are important components of precision agriculture (Panda *et al.* 2010). Many studies have focused on chlorophyll content estimation based on hyperspectral remote sensing techniques, and a number of vegetation indices have been developed for this purpose. In this study, 96 published vegetation indices (Table 1) were used as inputs of machine learning models for estimating the chlorophyll content of shaded tea. Five (i.e. Chl_{green}, $Chl_{red\ edge}1$, $Chl_{red\ edge}2$, chlorophyll vegetation index (CVI) and global imager vegetation index (GLI)) are based on broadband reflectance and the mean reflectance values were used in this study. Although most of the indices are based on original reflectance or a first derivative at a given wavelength ($R_{wavelength}$ or $D_{wavelength}$), there are eight feature-based indices that focus on the red edge (i.e. edge-green first derivative normalized difference (EGFN), edge-green first derivative ratio (EGFR), wavelength of the red edge (RE), Red-Edge Inflection Point (REIP), Red-edge position liner extrapolation method proposed by Cho and Skidmore (2006) (REP1), Red-edge position liner extrapolation method proposed by Guyot and Baret (1988) (REP2), sum of the amplitudes between 680 and 780 nm in the first derivative of the reflectance spectra (Sum1) and sum of derivative values between 626 nm and 795 nm (Sum2)). They are calculated as a sum value of the reflectance (Elvidge and Chen 1995; Filella *et al.* 1995) or a wavelength value (Collins 1978; Horler *et al.* 1983; Miller *et al.* 1990; Filella *et al.* 1995). In addition, the triangular vegetation index (TVI) (Broge and Leblanc 2001) and spectral polygon vegetation index (SPVI) (Vincini *et al.* 2006) are categorised as feature-based indices.

174 The indices based on original reflectance or first derivative are divided into four groups:
175 reflectance or first derivative at a given wavelength or inverse reflectance (R, D or 1/R) (Boochs
176 *et al.* 1990; Gitelson *et al.* 1999; Carter and Knapp 2001); difference in reflectance (DR) (Jordan
177 1969; Buschmann and Nagel 1993); simple ratio (SR) (Jordan 1969; Chappelle *et al.* 1992;
178 Vogelmann *et al.* 1993; Carter 1994; McMurtrey *et al.* 1994; Peñuelas *et al.* 1995a; Gitelson and
179 Merzlyak 1996; Lichtenthaler *et al.* 1996; Zarco-Tejada *et al.* 2003a; Zarco-Tejada *et al.* 2003c;
180 Delalieux *et al.* 2009; Gong *et al.* 2014) or modified simple ratio (mSR) (Sims and Gamon 2002);
181 and normalized differences (ND or dND) (Gong *et al.* 2014; Sonobe and Wang 2018; Sonobe and
182 Wang 2017b). In addition, complicated indices based on soil line (Rondeaux *et al.* 1996; Wu *et al.*
183 2008), chlorophyll absorption ratio indices (Kim *et al.* 1994; Daughtry *et al.* 2000; Wu *et al.* 2008;
184 Guan and Liu 2009) and integrated forms (Daughtry *et al.* 2000; Wu *et al.* 2008; Guan and Liu
185 2009) were evaluated in this study.

186 <Table 1>

187 2.3. Regression models based on machine learning algorithms

188 Machine learning algorithms including random forests (RF), support vector machine (SVM) and
189 kernel-based extreme learning machine (KELM) were applied to estimate the chlorophyll content
190 of shade grown tea from hyperspectral reflectance.

191

192 The optimisations of each machine learning algorithm were conducted based on Bayesian
193 optimisation, which is a framework used to optimise hyperparameters of noisy, expansive
194 black-box functions and it defines a principled approach to modelling uncertainty (Bergstra and
195 Bengio 2012). These processes were conducted with the Gaussian process (GP), which is a

196 continuous stochastic process commonly used for Bayesian optimisation (Snoek *et al.* 2015). All
197 the processes were conducted using R 3.4.3 (R Core Team 2017). While KELM was conducted
198 using MATLAB and Statistics Toolbox Release 2016a (MathWorks, Inc., USA) and the source
199 code was downloaded from http://www.ntu.edu.sg/home/egbhuang/index.html, RF and SVM were
200 assessed using the 'randomForest' package (Liaw and Wiener 2002) and 'kernlab' package
201 (Karatzoglou *et al.* 2004), respectively. For applying Bayesian optimisation, the
202 'rBayesianOptimization' package (Yan 2016) was applied.
203
204 2.3.1 Random forests
205 RF is an ensemble learning technique that builds multiple trees (the classification and regression
206 tree, CART) (Breiman 2001) and its two user-defined parameters, number of trees (ntree) and the
207 number of variables used to split the nodes (mtry), are normally optimised. Each tree is built using
208 training data and the nodes are split using the best split variable out of a group of randomly
209 selected variables (Liaw and Wiener 2002). This strategy guards against over-fitting and can
210 handle thousands of dependent and independent input variables without variable deletion.
211 Although the two main user-defined parameters are the number of trees (k) and the number of
212 variables used to split the nodes (m), the generalisation error always converges, and overfitting is
213 not a problem if the number of trees is increased. However, randomising the splitting rule can
214 improve the performance for ensembles (Ishwaran 2015; Sonobe *et al.* 2017). Therefore, three
215 hyperparameters including the minimum number of unique cases in a terminal node (nodesize), the
216 maximum depth to which a tree should be grown (nodedepth) and the number of random splits
217 (nsplit) were optimised in this study, as well as ntree and mtry.

2.3.2 Support vector machine

SVM has been successfully applied to solve the problem of high dimension and local minima (Ding *et al.* 2016). The 'kernel trick' has frequently been applied instead of attempting to fit a non-linear model in previous studies for classification and regression, and the Gaussian Radial Basis Function (RBF) kernel was most used (Chatziantoniou *et al.* 2017). In this study, the RBF kernel was applied, and the regularisation parameter C and the kernel bandwidth σ were tuned to control the flexibility.

2.3.3 Kernel based extreme learning machine

ELM (Huang *et al.* 2006), which is expressed as a single hidden layer feed-forward neural network, has been widely applied for many practical tasks, such as prediction, fault diagnosis, recognition, classification and signal processing (Li *et al.* 2016). Since a fixed hidden layer is composed of a vast number of nonlinear nodes and the hidden layer bias is defined randomly in this algorithm (Huang *et al.* 2006), it possesses fewer hyperparameters than deep learning, such as deep belief networks. Similar to SVM, the RBF kernel was applied in this study and the hyperparameters of the regulation coefficient (Cr) and kernel parameter (Kp) were optimised.

2.4. Inversion of the radiative transfer model

The PROSPECT model, which is based on the plate model (Allen *et al.* 1969), simulates leaf reflectance and transmittance. The first version was expressed as a function of three input parameters, including the internal structure parameter of the leaf mesophyll (N), chlorophyll content and water content. The input parameters of the latest version (PROSPECT-D) are leaf

mass per area, brown pigments, carotenoid and anthocyanin contents, plus the above three parameters. Féret et al (2017) reported that PROSPECT-D outperforms all previous versions.

The model inversion of PROSPECT-D was conducted using MATLAB and Statistics Toolbox Release 2016a and the source codes (PROSPECT-D_Matlab.rar) downloaded from the portal site (Institut de physique du globe de Paris, 2017). The code for the inversion of PROSPECT-5 was modified for the application of PROSPECT-D. The absorption coefficients of this model were calibrated to avoid potential systematic bias and error propagation in the inversion process following Féret et al. (2008).

2.5. Performance assessment

All measurements were divided into three groups, training (50%), validation (25%) and test data (25%) for assessing the potential of machine learning algorithms (Hastie *et al.* 2009). To apply this strategy, all measurements were divided into four groups based on the shading treatments and, 50% of the groups were selected as training data, which were used for generating regression models, based on random number for each group. Next, 50% of the remaining measurements were selected as validation data, which were used for optimising hyperparameters of the machine learning algorithms. Finally, the last group was used as test data for evaluating accuracies. This procedure was repeated ten times to increase the robustness of the results.

After dividing the data, variable selection based on the genetic algorithm (GA), which is an adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and

genetics, was applied to remove non-informative variables to generate better and simpler prediction models (Villar *et al.* 2017) using the training data. This process was conducted using R 3.4.3 and the 'plsVarSel' package (Mehmood *et al*. 2012) and the preliminary parameters were set to the default values, which were proposed by Hasegawa *et al* (1999). Then, the hyperparameters of the regression models based on machine learning algorithms were optimised based on the estimation errors for the validation data.

For assessing vegetation indices, training data and validation data, sets were merged and used to generate regression models based on linear or exponential regression. This merged dataset was also used to calibrate the absorption coefficients of PROSPECT-D. Finally, chlorophyll contents were estimated for the test data and the accuracies of each model were evaluated based on them.

For evaluating performances, the ratio of performance to deviation (RPD, Equation (4)) (Williams 1987) was applied. Each method was classified into three categories based on RPD: Category 'A' (RPD > 2.0), Category 'B' (1.4 ≤ RPD ≤ 2.0) and Category 'C' (RPD < 1.4). Models classified as Categories 'A' or 'B' were assumed to have the potential to estimate chlorophyll content (Chang *et al.* 2001). Category 'A' is divided into three levels including approximate (2.0 ≤ RPD ≤ 2.5), good (2.5 < RPD ≤ 3.0) and excellent (RPD < 3.0) (Saeys *et al.* 2005).

$$\mathrm{RPD} = \mathrm{SD}/\mathrm{RMSE} \qquad (4)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(\hat{y}_i - y_i)^2} \qquad (5)$$

where SD is the standard deviation of the chlorophyll content in the test data, n is number of samples, y_i is the measured chlorophyll content and $\hat{y}_i$ is the estimated chlorophyll content.

A data-based sensitivity analysis (DSA) (Cortez and Embrechts 2013), which uses several training samples instead of a baseline vector and DSA has been applied for the regression or classification models based on machine learning algorithms by querying the fitted models with sensitivity samples, was applied for analysing the sensitivity of the regression models using R 3.4.3 and the 'rminer' package (Cortez and Embrechts 2013).

3. Results and Discussion

3.1. Chlorophyll content after each treatment

Chlorophyll a and b contents were determined based on measurements of the absorbance of the supernatant dimethyl-formamide extracts. Table 2 summarises the main characteristics of the measurements. The numbers of samples collected were not the same since some leaf disks were too thin to measure chlorophyll content using UV-1280. Figure 1 shows histograms of the chlorophyll content of the different shading treatments on the two dates. The mean values of chlorophyll content by leaf area (μg/cm²) were 15.74, 23.37, 27.40 and 27.02 on 1 May, and 35.10, 44.79, 51.05 and 49.39 on 11 May for 0%, 35%, 75% and 90% shading, respectively.

Shading treatment makes leaf protein content higher and leaves become thicker (Poorter *et al.* 2006). So, shaded leaves contain more photosynthetic pigments, especially chlorophyll a, than sunlit leaves to harvest more light and nitrogen (Suzuki and Shioi 2003). As a result, the mean values of chlorophyll a and b contents were greater with more shading and a significant difference in chlorophyll content was confirmed among the four shading treatments when all the

306 measurements from the two dates were combined ($p < 0.01$, based on ANOVA test). However, the
307 differences in chlorophyll content were not significant for 35%, 75% or 90% shading for either
308 date ($p < 0.05$, based on the Tukey-Kramer test) because the relative amount of chlorophyll b
309 decreased and chlorophyll a/b ratios increased sharply in a linear manner at low light intensity
310 (Leong and Anderson 1984).

311 <Table 2>

312 <Figure 1>

313 3.2. Spectral reflectance of different treatments

314 The reflected spectra of each shading treatment are presented in Figure 2. The reflectance of
315 tea leaves near the green peak on 1 May was larger than that on 11 May and it was larger for
316 lighter shading treatments. The red edge inflection points were confirmed around 740 nm as with
317 previous studies (Vogelmann *et al.* 1993) and they became greater from 1 May to 11 May for all
318 shading treatments, even though the differences in the reflectance at the start of the red edge (near 680 nm)
319 were not significant among the four treatments. The differences in reflectance values between 800 nm and
320 1400 nm were unclear for 0% and 35% shading, while those of 75% and 90% shading on 11 May were
321 apparently higher than those on 1 May. A similar trend was found in reflectance values of tea leaves
322 between 1600 nm and 1800 nm.

323 <Figure 2>

324 For tea leaves, significant negative correlations were confirmed between chlorophyll content and
325 reflectance values between 500 and 750 nm (Figure 3). The lowest correlations were obtained at
326 741 nm, 735 nm and 518 nm for all measurements acquired on 1 May and on 11 May, respectively.

327 <Figure 3>

328 3.3. Selected wavelengths based on GA

329 Table 3 lists the selected wavelengths for each round to estimate chlorophyll content based on GA.

330 The numbers of wavelengths were 10–16. Mainly, selected wavelengths were near the green peak
331 and red edge inflection point of the tea leaves, and they were sensitive to chlorophyll content.
332 However, some wavelengths shorter than 500 nm or greater than 750 nm, for which there were few
333 differences in reflectance values among the four shading treatments, were selected and these
334 values were used as references. These values have been applied to emphasise the reflectance at
335 680–690 nm for estimating total chlorophyll content in previous studies (Peñuelas *et al.* 1995b;
336 Lichtenthaler *et al.* 1996).

337 <Table 3>

338 3.4. Selected indices based on GA

339 The indices, which were selected for estimating chlorophyll content based on GA, are listed in
340 Table 4. From 3 to 6 indices were selected and used to reduce numbers of explanatory variables for
341 regression models. MSAVI was selected three times; mSR2, TCI and EGFR were selected twice
342 and most of the indices were selected only once. Chlorophyll strongly absorbs light at blue and red
343 spectra and does not include light in green and orange spectra (Mattos *et al.* 2015). Most
344 vegetation indices for chlorophyll content use wavelengths ranging from 400 to 860 nm.
345 Furthermore, some stresses influence reflectance at specific wavelengths and reflectance between
346 healthy and stressed vegetation can be detected in changes to the green peak and the red edge
347 (Zarco-Tejada *et al.* 2001). As a result, some combinations consist of red edge-related indices and
348 indices covering photosynthetically active radiation (PAR) domain.

349 <Table 4>

350 3.5. Accuracy validation

351 The statistical criteria of accuracies including the RPD, RMSE and R^2 from the test data are given

352 in Table 5.

353 <Table 5>

354 Among all the machine learning algorithms with RPD values always greater than 2.0 when
355 original reflectance values were used, KELM showed the best performance with RPD values of
356 more than 3.38. This means that it is an excellent model for quantifying chlorophyll content.
357 However, there is no advantage of using vegetation indices for machine learning regression and
358 that made their RPD values smaller, except for round 2 and 9 of SVM. Rounds 4, 6 and 10 of SVM
359 and round 6 of KELM were unacceptable as regression models for estimating chlorophyll
360 contents.

361

362 Broadleaf trees are composed of two distinctive leaf groups of shaded and sunlit leaves and most
363 of indices were divided into two groups including the sunlit leaf-oriented indices and the
364 shaded-oriented indices (Sonobe and Wang 2017a). The light shading treatments, such as 0 % or
365 35 % shading, made sunlit leaves, while the heavy shading treatments, such as 75 % or 90%
366 shading, made shaded leaves. As the result, few vegetation indices could be used as generally
367 applicable indices to express the differences in leaf chlorophyll content and the combination of
368 machine learning algorithms and vegetation indices led the worse results. The comparisons among
369 the three algorithms were conducted based on the results of the original reflectance values.

370

371 To clarify which wavelengths were sensitive to their accuracies, a sensitivity analysis was
372 conducted and Figure 4 shows the results of the DSA. Although there were specific differences in

373 strategies in which wavelengths were attached weights to estimate chlorophyll contents, they were
374 obscure between SVM and KELM and both used RBF kernel in this study, the accuracies of
375 KELM were superior to those of SVM. Some previous studies showed the selection of the kernel
376 function parameters can negatively affect their accuracies (Horvath 2003). The optimal values of σ
377 ranged from 2^{-41} to 2^{-1} while Kp ranged from 2^{-8} to 2^{-2} based on Bayesian optimisation and that led
378 the differences. (Huang *et al.* 2010) showed that the extreme learning machine has less
379 optimisation constraints and its superiorities have been confirmed in regression (Maliha *et al.*
380 2018). The reflectance at 701–750 nm had the greatest influence on chlorophyll content
381 estimations for all the algorithms; it occupied an importance of 45.1% for RF, while, no
382 wavelength that occupied an importance of more than 20% was confirmed for SVM or KELM.
383 Thus, RF achieved its high performances based on a few explanatory variables. This strategy
384 might be effective than SVM; however, the red edge inflection points of tea leaves were increasing
385 and the green peek was decreasing with the shading treatment (Figure 2). The excessive
386 partialities of RF's importance made its estimation accuracies lower than KELM.

387 <Figure 4>

388 The RPD values of PROSPECT-D were relatively stable between 1.71 and 2.31; therefore, this
389 method was assumed to have the potential to estimate chlorophyll content according to the
390 criterion by Chang et al. (2001). In this model, the wavelength near the green peak and the red
391 edge inflection point are not fully considered and the statistics of PROSPECT-D were inferior to
392 those of the machine learning algorithms. However, it is useful to estimate other biochemical
393 properties including carotenoid content, anthocyanin content, water content and leaf mass per area
394 simultaneously. PROSPECT-D has a high potential for quantifying carotenoid content (Sonobe *et*

al. 2018b). Since the chlorophyll to carotenoid ratio is an indicator for environmental stress in plants (Hendry and Price 1993), this model might be useful to assess physiological properties as well as biochemical properties. Furthermore, in some versions of PROSPECT, the influence of reflectance in total chlorophyll content is separated into chlorophyll a and b. However, a consideration of anthocyanin information and improvement in the measurement of individual photosynthetic pigment concentrations are needed since the measurement of carotenoid was not accurate enough (Zhang *et al.* 2017). Thus, there is some potential for using PROSPECT to assess vegetation properties following some improvements.

In this study, it was revealed that the combination of leaf scale spectroscopy and machine learning has exhibited relatively high accuracy and has been found to be useful for quick estimation of chlorophyll content. However, satellite- or unmanned aerial vehicle (UAV) remote sensing techniques are more of professional applications concerning large scale assessment. There remain many gaps to be crossed over from the study reported here to large scale satellite-borne applications.

4. Conclusions

The potential of vegetation reflectance for quantifying chlorophyll content of shaded tea was evaluated. In this study, two methods were assessed, the inversion of a radiative transfer model and machine learning algorithms, and the estimations based on machine learning algorithms were superior. Specifically, KELM yielded the most accurate estimation with a RMSE of 3.04 ± 0.52 μg cm^{-2} and RPD values from 3.38 to 5.92, which means the regression models based on KELM were

417 excellent and were confirmed for quantifying chlorophyll content.

418

419 PROSPECT-D possessed some potential for this purpose; its RPD values ranged from 1.71 to 2.62
420 and more improvements were required to apply this method to shade grown tea cultivation.
421 However, it is not sufficient to evaluate chlorophyll content of tea trees under high light stress and
422 further improvements of the PROSPECT model are required. Our results showed that applying
423 machine learning algorithms is a unique solution to conduct in-situ measurements of green tea.

424

425 **Funding**

426 This work was supported by the Science and Technology Research Promotion Program for
427 Agriculture, Forestry, Fisheries, and Food Industry [grant number 27015C].

428

429 **Disclosure statement**

430 No potential conflicts of interest are reported by the authors.

431

432 **References**

433 Allen, WA, Gausman, HW, Richardson, AJ, Thomas, JR (1969) Interaction of isotropic light with a
434 compact plant leaf. Journal of the Optical Society of America 59, 1376-1379.

435 Bergstra, J, Bengio, Y (2012) Random search for hyper-parameter optimization. Journal of
436 Machine Learning Research 13, 281-305.

437 Biau, G, Scornet, E (2016) A random forest guided tour. Test 25, 197-227.

438 Blackburn, GA (1998a) Quantifying chlorophylls and caroteniods at leaf and canopy scales: An

439 evaluation of some hyperspectral approaches. Remote Sensing of Environment 66,
440 273-285.
441 Blackburn, GA (1998b) Spectral indices for estimating photosynthetic pigment concentrations: a
442 test using senescent tree leaves. International Journal of Remote Sensing 19, 657-675.
443 Blackburn, GA, Steele, CM (1999) Towards the remote sensing of matorral vegetation physiology:
444 Relationships between spectral reflectance, pigment, and biophysical characteristics of
445 semiarid bushland canopies. Remote Sensing of Environment 70, 278-292.
446 Boochs, F, Kupfer, G, Dockter, K, Kuhbauch, W (1990) Shape of the red edge as vitality indicator
447 for plants. International Journal of Remote Sensing 11, 1741-1753.
448 Breiman, L (2001) Random forests. Machine Learning 45, 5-32.
449 Broge, NH, Leblanc, E (2001) Comparing prediction power and stability of broadband and
450 hyperspectral vegetation indices for estimation of green leaf area index and canopy
451 chlorophyll density. Remote Sensing of Environment 76, 156-172.
452 Buschmann, C, Nagel, E (1993) In vivo spectroscopy and internal optics of leaves as basis for
453 remote sensing of vegetation. International Journal of Remote Sensing 14, 711-722.
454 Carter, GA (1994) Ratios of leaf reflectances in narrow wavebands as indicators of plant stress.
455 International Journal of Remote Sensing 15, 697-703.
456 Carter, GA, Knapp, AK (2001) Leaf optical properties in higher plants: Linking spectral
457 characteristics to stress and chlorophyll concentration. American Journal of Botany 88,
458 677-684.
459 Chang, CW, Laird, DA, Mausbach, MJ, Hurburgh, CR (2001) Near-infrared reflectance
460 spectroscopy-principal components regression analyses of soil properties. Soil Science

461 Society of America Journal 65, 480-490.
462 Chappelle, EW, Kim, MS, McMurtrey, JE (1992) Ratio analysis of reflectance spectra (RARS): An
463 algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B,
464 and carotenoids in soybean leaves. Remote Sensing of Environment 39, 239-247.
465 Chatziantoniou, A, Petropoulos, GP, Psomiadis, E (2017) Co-Orbital Sentinel 1 and 2 for LULC
466 Mapping with Emphasis on Wetlands in a Mediterranean Setting Based on Machine
467 Learning. Remote Sensing 9,
468 Chemura, A, Mutanga, O, Odindi, J (2017) Empirical Modeling of Leaf Chlorophyll Content in
469 Coffee (Coffea Arabica) Plantations With Sentinel-2 MSI Data: Effects of Spectral
470 Settings, Spatial Resolution, and Crop Canopy Cover. IEEE Journal of Selected Topics in
471 Applied Earth Observations and Remote Sensing 10, 5541-5550.
472 Chen, JM (1996) Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal
473 Applications. Canadian Journal of Remote Sensing 22, 229-242.
474 Cho, MA, Skidmore, AK (2006) A new technique for extracting the red edge position from
475 hyperspectral data: The linear extrapolation method. Remote Sensing of Environment 101,
476 181-193.
477 Collins, W (1978) Remote sensing of crop type and maturity. Photogrammetric Engineering and
478 Remote Sensing 44, 43-55.
479 Cortez, P, Embrechts, MJ (2013) Using sensitivity analysis and visualization techniques to open
480 black box data mining models. Information Sciences 225, 1-17.
481 Das, B, Sahoo, RN, Pargal, S, Krishna, G, Verma, R, Chinnusamy, V, Sehgal, VK, Gupta, VK
482 (2017) Comparison of different uni- and multi-variate techniques for monitoring leaf water

483 status as an indicator of water-deficit stress in wheat through spectroscopy. Biosystems
484 Engineering 160, 69-83.
485 Dash, J, Curran, PJ (2004) The MERIS terrestrial chlorophyll index. International Journal of
486 Remote Sensing 25, 5403-5413.
487 Datt, B (1998) Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total
488 carotenoid content in eucalyptus leaves. Remote Sensing of Environment 66, 111-121.
489 Datt, B (1999a) A new reflectance index for remote sensing of chlorophyll content in higher
490 plants: Tests using Eucalyptus leaves. Journal of Plant Physiology 154, 30-36.
491 Datt, B (1999b) Visible/near infrared reflectance and chlorophyll content in Eucalyptus leaves.
492 International Journal of Remote Sensing 20, 2741-2759.
493 Daughtry, CST, Walthall, CL, Kim, MS, de Colstoun, EB, McMurtrey, JE (2000) Estimating corn
494 leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of
495 Environment 74, 229-239.
496 De Costa, WAJM, Mohotti, AJ, Wijeratne, MA (2007) Ecophysiology of tea. Brazilian Journal of
497 Plant Physiology 19, 299-332.
498 Delalieux, S, Somers, B, Verstraeten, WW, Van Aardt, JAN, Keulemans, W, Coppin, P (2009)
499 Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf
500 phenology. International Journal of Remote Sensing 30, 1887-1912.
501 Ding, SF, Shi, ZZ, Tao, DC, An, B (2016) Recent advances in Support Vector Machines.
502 Neurocomputing 211, 1-3.
503 Elarab, M, Ticlavilca, AM, Torres-Rua, AF, Maslova, I, McKee, M (2015) Estimating chlorophyll
504 with thermal and broadband multispectral high resolution imagery from an unmanned

505 aerial system using relevance vector machines for precision agriculture. International
506 Journal of Applied Earth Observation and Geoinformation 43, 32-42.
507 Elvidge, CD, Chen, ZK (1995) Comparison of broad-band and narrow-band red and near-infrared
508 vegetation indices. Remote Sensing of Environment 54, 38-48.
509 Filella, I, Amaro, T, Araus, JL, Peñuelas, J (1996) Relationship between photosynthetic
510 radiation-use efficiency of Barley canopies and the photochemical reflectance index (PRI).
511 Physiologia Plantarum 96, 211-216.
512 Filella, I, Serrano, L, Serra, J, Peñuelas, J (1995) Evaluating Wheat Nitrogen Status with Canopy
513 Reflectance Indices and Discriminant Analysis. Crop Science 35, 1400-1405.
514 Féret, J-B, Francois, C, Asner, GP, Gitelson, AA, Martin, RE, Bidel, LPR, Ustin, SL, le Maire, G,
515 Jacquemoud, S (2008) PROSPECT-4 and 5: Advances in the leaf optical properties model
516 separating photosynthetic pigments. Remote Sensing of Environment 112, 3030-3043.
517 Féret, J-B, Gitelson, AA, Noble, SD, Jacquemoud, S (2017) PROSPECT-D: Towards modeling
518 leaf optical properties through a complete lifecycle. Remote Sensing of Environment 193,
519 204-215.
520 Gabriel, JL, Zarco-Tejada, PJ, Lopez-Herrera, PJ, Perez-Martin, E, Alonso-Ayuso, M, Quemada,
521 M (2017) Airborne and ground level sensors for monitoring nitrogen status in a maize crop.
522 Biosystems Engineering 160, 124-133.
523 Gamon, JA, Surfus, JS (1999) Assessing leaf pigment content and activity with a reflectometer.
524 New Phytologist 143, 105-117.
525 Gandia, S, Fernandez, G, Moreno, J (2005) 'Retrieval of Vegetation Biophysical Variables from
526 CHRIS/PROBA Data in the SPARC Campaign, the 2nd CHRIS/Proba Workshop.' Frascati,

527 Italy. (ESA/ESRIN:
528 Gitelson, A, Merzlyak, MN (1994) Spectral Reflectance Changes Associated with Autumn
529 Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral
530 Features and Relation to Chlorophyll Estimation. Journal of Plant Physiology 143,
531 286-292.
532 Gitelson, AA, Buschmann, C, Lichtenthaler, HK (1999) The chlorophyll fluorescence ratio
533 F-735/F-700 as an accurate measure of the chlorophyll content in plants. Remote Sensing
534 of Environment 69, 296-302.
535 Gitelson, AA, Gritz, Y, Merzlyak, MN (2003) Relationships between leaf chlorophyll content and
536 spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher
537 plant leaves. Journal of Plant Physiology 160, 271-282.
538 Gitelson, AA, Kaufman, YJ, Merzlyak, MN (1996) Use of a green channel in remote sensing of
539 global vegetation from EOS-MODIS. Remote Sensing of Environment 58, 289-298.
540 Gitelson, AA, Keydan, GP, Merzlyak, MN (2006) Three-band model for noninvasive estimation of
541 chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical
542 Research Letters 33,
543 Gitelson, AA, Merzlyak, MN (1996) Signature analysis of leaf reflectance spectra: Algorithm
544 development for remote sensing of chlorophyll. Journal of Plant Physiology 148, 494-500.
545 Gobron, N, Pinty, B, Verstraete, MM, Widlowski, JL (2000) Advanced vegetation indices
546 optimized for up-coming sensors: Design, performance, and applications. IEEE
547 Transactions on Geoscience and Remote Sensing 38, 2489-2505.
548 Gong, Z, Zhao, Y, Zhao, W, Lin, C (2014) Estimation model for plant leaf chlorophyll content

549 based on the spectral index content. Acta Ecologica Sinica (in Chinese) 34, 5736-5745.
550 Guan, L, Liu, X (2009) Hyperspectral recognition models for physiological ecology
551 characterization of rice in Cd pollution stress. Ecology and Environmental Sciences (in
552 Chinese) 18, 488-493.
553 Guyot, G, Baret, F TDGaJJ Hunt (Ed.) (1988) 'Utilisation de la haute resolution spectrale pour
554 suivre l'etat des couverts vegetaux, Spectral Signatures of Objects in Remote Sensing.'
555 Aussois (Modane), France. (European Space Agency:
556 Haboudane, D, Miller, JR, Tremblay, N, Zarco-Tejada, PJ, Dextraze, L (2002) Integrated
557 narrow-band vegetation indices for prediction of crop chlorophyll content for application
558 to precision agriculture. Remote Sensing of Environment 81, 416-426.
559 Hasegawa, K, Kimura, T, Funatsu, K (1999) GA Strategy for Variable Selection in QSAR Studies:
560 Application of GA-Based Region Selection to a 3D-QSAR Study of Acetylcholinesterase
561 Inhibitors. Journal of Chemical Information and Computer Sciences 39, 112-120.
562 Hastie, T, Tibshirani, R, Friedman, J (Eds P Bickel, P Diggle, SE Fienberg, U Gather, S Zeger
563 (2009) 'The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
564 Second Edition.' (Springer-Verlag New York: The United States)
565 Hendry, GAF, Price, AH (1993) 'Stress Indicators: Chlorophylls and Carotenoids.'
566 Hernandez-Clemente, R, Navarro-Cerrillo, RM, Zarco-Tejada, PJ (2014) Deriving Predictive
567 Relationships of Carotenoid Content at the Canopy Level in a Conifer Forest Using
568 Hyperspectral Imagery and Model Simulation. IEEE Transactions on Geoscience and
569 Remote Sensing 52, 5206-5217.
570 Horler, DNH, Dockray, M, Barber, J (1983) The red edge of plant leaf reflectance. International

571 Journal of Remote Sensing 4, 273-288.
572 Horvath, G (2003) CMAC neural network as an SVM with B-spline kernel functions, 20th IEEE
573 Instrumentation and Measurement Technology Conference, IMTC30, 1108-1113.
574 Hosgood, B, Jacquemoud, S, Andreoli, G, Verdebout, J, Pedrini, G, Schmuck, G (1994) Leaf
575 Optical Properties EXperiment 93 (LOPEX93). European Commission, Joint Research
576 Centre, Institute for Remote Sensing Applications. Report EUR 16095 EN, pp. 21.
577 Huang, GB, Ding, XJ, Zhou, HM (2010) Optimization method based extreme learning machine for
578 classification. Neurocomputing 74, 155-163.
579 Huang, GB, Zhu, QY, Siew, CK (2006) Extreme learning machine: Theory and applications.
580 Neurocomputing 70, 489-501.
581 Huete, A, Didan, K, Miura, T, Rodriguez, EP, Gao, X, Ferreira, LG (2002) Overview of the
582 radiometric and biophysical performance of the MODIS vegetation indices. Remote
583 Sensing of Environment 83, 195-213.
584 Hunt, ER, Daughtry, CST, Eitel, JUH, Long, DS (2011) Remote Sensing Leaf Chlorophyll Content
585 Using a Visible Band Index. Agronomy Journal 103, 1090-1099.
586 Ishwaran, H (2015) The effect of splitting on random forests. Machine Learning 99, 75-118.
587 Jacquemoud, S, Ustin, SL, Verdebout, J, Schmuck, G, Andreoli, G, Hosgood, B (1996) Estimating
588 leaf biochemistry using the PROSPECT leaf optical properties model. Remote Sensing of
589 Environment 56, 194-202.
590 Japan Meteorological Agency (2017) https://www.jma.go.jp/jma/indexe.html
591 Jordan, CF (1969) Derivation of leaf area index from quality of light on the forest floor. Ecology
592 50, 663-666.

593 Karatzoglou, A, Smola, A, Hornik, K, Zeileis, A (2004) kernlab - An S4 Package for Kernel
594 Methods in R. Journal of Statistical Software 11, 1-20.
595 Kim, MS, Daughtry, CST, Chappelle, EW, Mcmurtrey, JE, Walthall, CL (1994) 'The use high
596 spectral resolution bands for estimating absorbed photo synthetically active radiation
597 (APAR) 6th International Symposium on Physical Measurements and Signatures in
598 Remote Sensing ' France.
599 Korus, A (2013) Effect of preliminary and technological treatments on the content of chlorophylls
600 and carotenoids in kale (Brassica oleracea L. var. Acephala). Journal of Food Processing
601 and Preservation 37, 335-344.
602 Kuriyama, S, Hozawa, A, Ohmori, K, Shimazu, T, Matsui, T, Ebihara, S, Awata, S, Nagatomi, R,
603 Arai, H, Tsuji, I (2006) Green tea consumption and cognitive function: a cross-sectional
604 study from the Tsurugaya Project. American Journal of Clinical Nutrition 83, 355-361.
605 le Maire, G, Francois, C, Dufrene, E (2004) Towards universal broad leaf chlorophyll indices
606 using PROSPECT simulated database and hyperspectral reflectance measurements.
607 Remote Sensing of Environment 89, 1-28.
608 le Maire, G, Francois, C, Soudani, K, Berveiller, D, Pontailler, J-Y, Breda, N, Genet, H, Davi, H,
609 Dufrene, E (2008) Calibration and validation of hyperspectral indices for the estimation of
610 broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf
611 canopy biomass. Remote Sensing of Environment 112, 3846-3864.
612 Lee, LS, Choi, JH, Son, N, Kim, SH, Park, JD, Jang, DJ, Jeong, Y, Kim, HJ (2013) Metabolomic
613 Analysis of the Effect of Shade Treatment on the Nutritional and Sensory Qualities of
614 Green Tea. Journal of Agricultural and Food Chemistry 61, 332-338.

615 Leong, TY, Anderson, JM (1984) Adaptation of the thylakoid membranes of pea chloroplasts to
616 light intensities. I. Study on the distribution of chlorophyll-protein complexes.
617 Photosynthesis research 5, 105-15.
618 Li, XD, Mao, WJ, Jiang, W (2016) Multiple-kernel-learning-based extreme learning machine for
619 classification design. Neural Computing & Applications 27, 175-184.
620 Li, ZH, Jin, XL, Wang, JH, Yang, GJ, Nie, CW, Xu, XG, Feng, HK (2015) Estimating winter wheat
621 (Triticum aestivum) LAI and leaf chlorophyll content from canopy reflectance data by
622 integrating agronomic prior knowledge with the PROSAIL model. International Journal of
623 Remote Sensing 36, 2634-2653.
624 Liang, L, Qin, ZH, Zhao, SH, Di, LP, Zhang, C, Deng, MX, Lin, H, Zhang, LP, Wang, LJ, Liu, ZX
625 (2016) Estimating crop chlorophyll content with hyperspectral vegetation indices and the
626 hybrid inversion method. International Journal of Remote Sensing 37, 2923-2949.
627 Liaw, A, Wiener, M (2002) Classification and regression by random Forest. R News 2, 18-22.
628 Lichtenthaler, HK, Lang, M, Sowinska, M, Heisel, F, Miehe, JA (1996) Detection of vegetation
629 stress via a new high resolution fluorescence imaging system. Journal of Plant Physiology
630 148, 599-612.
631 Maccioni, A, Agati, G, Mazzinghi, P (2001) New vegetation indices for remote measurement of
632 chlorophylls based on leaf directional reflectance spectra. Journal of Photochemistry and
633 Photobiology B-Biology 61, 52-61.
634 Main, R, Cho, MA, Mathieu, R, O'Kennedy, MM, Ramoelo, A, Koch, S (2011) An investigation
635 into robust spectral indices for leaf chlorophyll estimation. Isprs Journal of
636 Photogrammetry and Remote Sensing 66, 751-761.

637 Maliha, A, Yusof, R, Shapiai, MI (2018) Extreme learning machine for structured output spaces.
638 Neural Computing & Applications 30, 1251-1264.
639 Masemola, C, Cho, MA, Ramoelo, A (2016) Comparison of Landsat 8 OLI and Landsat 7 ETM+
640 for estimating grassland LAI using model inversion and spectral indices: case study of
641 Mpumalanga, South Africa. International Journal of Remote Sensing 37, 4401-4419.
642 Mattos, ER, Singh, M, Cabrera, ML, Das, KC (2015) Enhancement of biomass production in
643 Scenedesmus bijuga high-density culture using weakly absorbed green light. Biomass &
644 Bioenergy 81, 473-478.
645 McMurtrey, JE, Chappelle, EW, Kim, MS, Meisinger, JJ, Corp, LA (1994) Distinguishing nitrogen
646 fertilization levels in field corn (Zea mays L.) with actively induced fluorescence and
647 passive reflectance measurements. Remote Sensing of Environment 47, 36-44.
648 Mehmood, T, Liland, KH, Snipen, L, Sæbø, S (2012) A review of variable selection methods in
649 Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems 118,
650 62-69.
651 Miller, JR, Hare, EW, Wu, J (1990) Quantitative characterisation of the red edge reflectance 1. An
652 inverted-Gaussian model. International Journal of Remote Sensing 11, 1755-1773.
653 Murchie, EH, Hubbart, S, Peng, S, Horton, P (2005) Acclimation of photosynthesis to high
654 irradiance in rice: gene expression and interactions with leaf development. Journal of
655 Experimental Botany 56, 449-460.
656 Mutanga, O, Skidmore, AK (2004) Narrow band vegetation indices overcome the saturation
657 problem in biomass estimation. International Journal of Remote Sensing 25, 3999-4014.
658 Panda, SS, Ames, DP, Panigrahi, S (2010) Application of Vegetation Indices for Agricultural Crop

659 Yield Prediction Using Neural Network Techniques. Remote Sensing 2, 673-696.
660 Peñuelas, J, Filella, I, Gamon, JA (1995a) Assessment of photosynthetic radiation‐use efficiency
661 with spectral reflectance. New Phytologist 131, 291-296.
662 Peñuelas, J, Filella, I, Lloret, P, Munoz, F, Vilajeliu, M (1995b) Reflectance assessment of mite
663 effects on apple trees. International Journal of Remote Sensing 16, 2727-2733.
664 Peñuelas, J, Gamon, JA, Fredeen, AL, Merino, J, Field, CB (1994) Reflectance indices associated
665 with physiological changes in nitrogen- and water-limited sunflower leaves. Remote
666 Sensing of Environment 48, 135-146.
667 Poorter, H, Pepin, S, Rijkers, T, de Jong, Y, Evans, JR, Korner, C (2006) Construction costs,
668 chemical composition and payback time of high- and low-irradiance leaves. Journal of
669 Experimental Botany 57, 355-371.
670 Prado-Cabrero, A, Beatty, S, Howard, A, Stack, J, Bettin, P, Nolan, JM (2016) Assessment of
671 lutein, zeaxanthin and meso-zeaxanthin concentrations in dietary supplements by chiral
672 high-performance liquid chromatography. European Food Research and Technology 242,
673 599-608.
674 Qi, J, Chehbouni, A, Huete, AR, Kerr, YH, Sorooshian, S (1994) A modified soil adjusted
675 vegetation index. Remote Sensing of Environment 48, 119-126.
676 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
677 Statistical Computing, Vienna, Austria. https://www.R-project.org/.
678 Rondeaux, G, Steven, M, Baret, F (1996) Optimization of soil-adjusted vegetation indices. Remote
679 Sensing of Environment 55, 95-107.
680 Roujean, JL, Breon, FM (1995) Estimating PAR absorbed by vegetation from bidirectional

681 reflectance measurements. Remote Sensing of Environment 51, 375-384.
682 Saeys, W, Mouazen, AM, Ramon, H (2005) Potential for onsite and online analysis of pig manure
683 using visible and near infrared reflectance spectroscopy. Biosystems Engineering 91,
684 393-402.
685 Shoko, C, Mutanga, O, Dube, T, Slotow, R (2018) Characterizing the spatio-temporal variations of
686 C3 and C4 dominated grasslands aboveground biomass in the Drakensberg, South Africa.
687 International Journal of Applied Earth Observation and Geoinformation 68, 51-60.
688 Sims, DA, Gamon, JA (2002) Relationships between leaf pigment content and spectral reflectance
689 across a wide range of species, leaf structures and developmental stages. Remote Sensing
690 of Environment 81, 337-354.
691 Smith, RCG, Adams, J, Stephens, DJ, Hick, PT (1995) Forecasting wheat yield in a
692 Mediterranean-type environment from the NOAA satellite. Australian Journal of
693 Agricultural Research 46, 113-125.
694 Snoek, J, Rippel, O, Swersky, K, Kiros, R, Satish, N, Sundaram, N, Patwary, MMA, Prabhat,
695 Adams, RP F Bach, D Blei (Eds) (2015) 'Scalable Bayesian optimization using deep neural
696 networks, the 32nd International Conference on Machine Learning (ICML).' Paris.
697 Sonobe, R, Miura, Y, Sano, T, H., H (2018a) Monitoring photosynthetic pigments of shade grown
698 tea from hyperspectral reflectance. Canadian Journal of Remote Sensing
699 Sonobe, R, Miura, Y, Sano, T, Horie, H (2018b) Estimating leaf carotenoid contents of shade
700 grown tea using hyperspectral indices and PROSPECT-D inversion. International Journal
701 of Remote Sensing 39, 1306-1320.
702 Sonobe, R, Wang, Q (2018) Nondestructive assessments of carotenoids content of broadleaved

plant species using hyperspectral indices Computers and Electronics in Agriculture 145, 18-26.

Sonobe, R, Wang, Q (2017a) Hyperspectral indices for quantifying leaf chlorophyll concentrations performed differently with different leaf types in deciduous forests. Ecological Informatics 37, 1-9.

Sonobe, R, Wang, Q (2017b) Towards a Universal Hyperspectral Index to Assess Chlorophyll Content in Deciduous Forests. Remote Sensing 9,

Sonobe, R, Yamaya, Y, Tani, H, Wang, XF, Kobayashi, N, Mochizuki, K (2017) Mapping crop cover using multi-temporal Landsat 8 OLI imagery. International Journal of Remote Sensing 38, 4348-4361.

Sonobe, R, Yamaya, Y, Tani, H, Wang, XF, Kobayashi, N, Mochizuki, K (2018c) Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. Journal of Applied Remote Sensing 12,

Sun, J, Shi, S, Yang, J, Du, L, Gong, W, Chen, BW, Song, SL (2018) Analyzing the performance of PROSPECT model inversion based on different spectral information for leaf biochemical properties retrieval. Isprs Journal of Photogrammetry and Remote Sensing 135, 74-83.

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.

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

Tucker, CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation.

725 Remote Sensing of Environment 8, 127?150.
726 Villar, A, Vadillo, J, Santos, JI, Gorritxategi, E, Mabe, J, Arnaiz, A, Fernandez, LA (2017) Cider
727 fermentation process monitoring by Vis-NIR sensor system and chemometrics. Food
728 Chemistry 221, 100-106.
729 Vincini, M, Frazzi, E, D'Alessio, P (2008) A broad-band leaf chlorophyll vegetation index at the
730 canopy scale. Precision Agriculture 9, 303-319.
731 Vincini, M, Frazzi, E, D'Alessio, P (2006) 'Angular dependence of maize and sugar beet VIs from
732 directional CHRIS/Proba data, the 4th ESA CHRIS PROBA Workshop.'
733 Vogelmann, JE, Rock, BN, Moss, DM (1993) Red edge spectral measurements from sugar maple
734 leaves. International Journal of Remote Sensing 14, 1563-1575.
735 Wang, KB, Liu, F, Liu, ZH, Huang, JA, Xu, ZX, Li, YH, Chen, JH, Gong, YS, Yang, XH (2010)
736 Analysis of chemical components in oolong tea in relation to perceived quality.
737 International Journal of Food Science and Technology 45, 913-920.
738 Wang, LF, Park, SC, Chung, JO, Baik, LH, Park, SK (2004) The compounds contributing to the
739 greenness of green tea. Journal of Food Science 69, S301-S305.
740 Wellburn, AR (1994) The spectral determination of chlorophyll a and chlorophyll b, as well as
741 total carotenoids, using various solvents with spectrophotometers of different resolution.
742 Journal of Plant Physiology 144, 307-313.
743 Whetton, RL, Hassall, KL, Waine, TW, Mouazen, AM (2018) Hyperspectral measurements of
744 yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. Biosystems
745 Engineering 166, 101-115.
746 Williams, P (1987) Variables affecting near-infraredreflectance spectroscopic analysis. In 'Near-

747 Infrared Technology in the Agricultural and Food Industries.' (Eds P Williams, K Norris.)
748 pp. 143-167. (American Association of Cereal Chemists Inc.:
749 Wu, C, Niu, Z, Tang, Q, Huang, W (2008) Estimating chlorophyll content from hyperspectral
750 vegetation indices: Modeling and validation. Agricultural and Forest Meteorology 148,
751 1230-1241.
752 Wu, CY, Niu, Z, Tang, Q, Huang, WJ, Rivard, B, Feng, JL (2009) Remote estimation of gross
753 primary production in wheat using chlorophyll-related vegetation indices. Agricultural and
754 Forest Meteorology 149, 1015-1021.
755 Yamamoto, A, Nakamura, T, Adu-Gyamfi, JJ, Saigusa, M (2002) Relationship between chlorophyll
756 content in leaves of sorghum and pigeonpea determined by extraction method and by
757 chlorophyll meter (SPAD-502). Journal of Plant Nutrition 25, 2295-2301.
758 Yan, Y (2016) 'Bayesian Optimization of Hyperparameters.' Available at
759 Zarco-Tejada, PJ, Berni, JAJ, Suarez, L, Sepulcre-Canto, G, Morales, F, Miller, JR (2009) Imaging
760 chlorophyll fluorescence with an airborne narrow-band multispectral camera for
761 vegetation stress detection. Remote Sensing of Environment 113, 1262-1275.
762 Zarco-Tejada, PJ, Miller, JR, Haboudane, D, Tremblay, N, Apostol, S (2003a) Detection of
763 chlorophyll fluorescence in vegetation from airborne hyperspectral CASI imagery in the
764 red edge spectral region. IGARSS 2003: IEEE International Geoscience and Remote
765 Sensing Symposium, Vols I - Vii, Proceedings: Learning from Earth's Shapes and Sizes
766 598-600.
767 Zarco-Tejada, PJ, Miller, JR, Haboudane, D, Tremblay, N, Apostol, S (2003b) Detection of
768 chlorophyll fluorescence in vegetation from airborne hyperspectral CASI imagery in the

red edge spectral region. IGARSS 2003: IEEE International Geoscience and Remote Sensing Symposium, Vols I - Vii, Proceedings: Learning from Earth's Shapes and Sizes 598-600.

Zarco-Tejada, PJ, Miller, JR, Noland, TL, Mohammed, GH, Sampson, PH (2001) Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 39, 1491-1507.

Zarco-Tejada, PJ, Pushnik, JC, Dobrowski, S, Ustin, SL (2003c) Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. Remote Sensing of Environment 84, 283-294.

Zhang, Y, Huang, JF, Wang, FM, Blackburn, GA, Zhang, HKK, Wang, XZ, Wei, CW, Zhang, KY, Wei, C (2017) An extended PROSPECT: Advance in the leaf optical properties model separating total chlorophylls into chlorophyll a and b. Scientific Reports 7, 10.

Zhao, J, Feng, M, Wang, C, Yang, W, Li, Z, Zhu, Z, Ren, P, Liu, T, Wang, H (2014) Simulating the Content of Chlorophyll in Winter Wheat Based on Spectral Vegetation Index. Journal of Shanxi Agricultural University (in Chinese) 34

787 **Tables**

788 Table 1 Vegetation indices evaluated in this study. dG: maximum of the first derivative of reflectance

789 in the green, dRE: maximum of the first derivative of reflectance in the red edge

Index	Formula	Reference
Chlorophyll absorption ratio index (CARI)	R_{700}*(SQRT((a*670+R_{670}+b)^2))/R_{670}*(a^2+1)$^{0.5}$ a = (R_{700}–R_{550})/150 and b = R_{550}–(a*550)	(Kim *et al.* 1994)
Chlorophyll absorption ratio index 2 (CARI2)	(|(a+1)*R_{670}+b|/(a^2+1)$^{0.5}$)*(R_{700}/R_{670}) a = (R_{700}– R_{550})/150 and b = R_{550}–(a*550)	
Reflectance value at 550 nm (Carter)	R_{550}	(Carter and Knapp 2001)
Chl_{green}	R_{NIR}/R_{GREEN}–1 R_{NIR}: mean reflectance value from 760 to 800 nm R_{GREEN}: mean reflectance value from 540 to 560 nm	(Gitelson *et al.* 2006)
$Chl_{red\ edge}1$	R_{NIR}/$R_{RED\ EDGE}$–1 R_{NIR}: mean reflectance value from 760 to 800 nm $R_{RED\ EDGE}$: mean reflectance value from 690 to725 nm	
$Chl_{red\ edge}2$	R_{750}/R_{710}–1	(Wu *et al.* 2009)
Red-edge chlorophyll index (CI)	R_{675}* R_{690}/R_{683}^2	(Zarco-Tejada *et al.* 2009)
Chlorophyll vegetation index (CVI)	R_{NIR}*R_{RED}/R_{GREEN}^2 R_{GREEN}: mean reflectance value from 490 to 570 nm R_{RED}: mean reflectance value from 640 to 760 nm R_{NIR}: mean reflectance value from 780 to 1400 nm	(Vincini *et al.* 2008; Hunt *et al.* 2011)
D_{703}	D_{703}	(Boochs *et al.* 1990)
D_{720}	D_{720}	
Datt1	R_{860}/(R_{550}*R_{708})	(Datt 1998)
Datt2	R_{672}/(R_{550}*R_{708})	
Datt3	R_{672}/R_{550}	
Datt4	D_{754}/D_{704}	(Datt 1999b)
Datt5	R_{850}/R_{710}	(Datt 1999a)

Datt6	$(R_{850}–R_{710})/(R_{850}–R_{680})$	
Double difference (DD)	$(R_{749}–R_{720})–(R_{701}–R_{672})$	(le Maire *et al.* 2008)
DDn	$2*R_{710}–R_{(710–50)}–R_{(710+50)}$	(le Maire *et al.* 2008)
dND(522, 728)	$(D_{522}–D_{728})/(D_{522}+D_{728})$	(Sonobe and Wang 2017b)
Double-peak index (DPI)	$(D_{688}*D_{710})/D_{697}^2$	(Zarco-Tejada *et al.* 2003b)
dSR1	D_{730}/D_{706}	
dSR2	D_{705}/D_{722}	
DR(800, 550)	$R_{800}–R_{550}$	(Buschmann and Nagel 1993)
DR(800, 680)	$R_{800}–R_{680}$	(Jordan 1969)
Edge-green first derivative normalized difference (EGFN)	(dRE–dG)/(dRE+dG)	(Peñuelas *et al.* 1994)
Edge-green first derivative ratio (EGFR)	dRE/dG	
Enhanced vegetation index (EVI)	$2.5*((R_{800}–R_{670})/(R_{800}–(6*R_{670})–(7.5*R_{475}) + 1))$	(Huete *et al.* 2002)
First derivative normalized difference vegetation index (FDNDVI)	$(D_{630}–D_{723})/(D_{630}+D_{723})$	(Zhao *et al.* 2014)
Greenness index (GI)	R_{554}/R_{677}	(Smith *et al.* 1995)
Gitel1	$1/R_{700}$	(Gitelson *et al.* 1999)
Gitel2	$(R_{750}–R_{800}/R_{695}–R_{740})–1$	(Gitelson *et al.* 2003)
Global imager vegetation index (GLI)	$(2*R_{GREEN}–R_{RED}–R_{BLUE})/ (2* R_{GREEN} + R_{RED} + R_{BLUE})$ R_{BLUE}: mean reflectance value from 420 to 480 nm R_{GREEN}: mean reflectance value from 490 to 570 nm R_{RED}: mean reflectance value from 640 to 760 nm	(Gobron *et al.* 2000)
Green normalized difference vegetation index	$(R_{800}–R_{550})/(R_{800}+R_{550})$	(Gitelson *et al.* 1996)

(GNDVI)		
Mac	$(R_{780}–R_{710})/(R_{780}–R_{680})$	(Maccioni *et al.* 2001)
Modified chlorophyll absorption in reflectance index (MCARI)	$((R_{700}–R_{670})–0.2*(R_{700}–R_{550}))*(R_{700}/R_{670})$	(Daughtry *et al.* 2000)
MCARI/OSAVI	MCARI/OSAVI	
Modified chlorophyll absorption in reflectance index 1 (MCARI1)	$1.5[1.2(R_{712}–R_{670})–0.5(R_{712}–R_{550})](R_{712}/R_{670})$	(Guan and Liu 2009)
MCARI1/MSAVI	MCARI1/MSAVI	
Modified chlorophyll absorption in reflectance index 2 (MCARI2)	$((R_{750}–R_{705})–0.2*(R_{750}–R_{550}))*(R_{750}/R_{705})$	(Wu *et al.* 2008)
MCARI2/OSAVI2	MCARI2/OSAVI2	(Wu *et al.* 2008)
mND705	$(R_{750}–R_{705})/(R_{750}+ R_{705}–2R_{445})$	(Sims and Gamon 2002)
mNDVI	$(R_{800}–R_{680})/(R_{800}+R_{680}–2R_{445})$	
mREIP	The index based on the Gaussian fit of the red edge derivative	(Miller *et al.* 1990)
Modified soil adjusted vegetation index (MSAVI)	$0.5*(2*R_{800}+1–SQRT((2*R_{800}+1)^2–8*(R_{800}–R_{670})))$	(Qi *et al.* 1994)
mSR1	$(R_{750}–R_{445})/(R_{705}–R_{445})$	(Sims and Gamon 2002)
mSR2	$(R_{800}–R_{445})/(R_{680}–R_{445})$	
mSR3	$(R_{750}/R_{705}–1)/SQRT((R_{750}/R_{705})+1)$	(Chen 1996)
MERIS terrestrial chlorophyll index (MTCI)	$(R_{754}–R_{709})/(R_{709}–R_{681})$	(Dash and Curran 2004)
Modified triangular vegetation index (MTVI)	$1.5[1.2(R_{712}–R_{550})–2.1(R_{670}–R_{550})]$	(Guan and Liu 2009)
MTVI/MSAVI	MTVI/MSAVI	
ND	$(R_{565}–R_{735})/(R_{565}+R_{735})$	(Gong *et al.* 2014)
Normalized difference	$(R_{800}–R_{670})/(R_{800}+R_{670})$	(Tucker 1979)

vegetation index 1 (NDVI1)		
Normalized difference vegetation index 2 (NDVI2)	$(R_{750}–R_{705})/(R_{750}+R_{705})$	(Gitelson and Merzlyak 1994; Gamon and Surfus 1999)
Normalized difference vegetation index 3 (NDVI3)	$(R_{682}–R_{553})/(R_{682}+R_{553})$	(Gandia *et al.* 2005)
Normalized pigments reflectance index (NPCI)	$(R_{680}–R_{460})/(R_{680}+R_{460})$	(Blackburn 1998a, 1998b)
Normalized pigments reflectance index 2 (NPCI2)	$(R_{680}–R_{430})/(R_{680}+R_{430})$	(Peñuelas *et al.* 1994)
Optimized soil adjusted vegetation index (OSAVI)	$(1+0.16)*(R_{800}–R_{670})/(R_{800}+R_{670}+0.16)$	(Rondeaux *et al.* 1996)
Optimized soil adjusted vegetation index 2 (OSAVI2)	$(1+0.16)*(R_{750}–R_{705})/(R_{750}+R_{705}+0.16)$	(Wu *et al.* 2008)
Pigment specific normalized difference for chlorophyll a (PSNDa)	$(R_{800}–R_{680})/(R_{800}+R_{680})$	
Pigment specific normalized difference for chlorophyll b (PSNDb)	$(R_{800}–R_{635})/(R_{800}+R_{635})$	(Blackburn 1998a, 1998b)
Pigment specific simple ratio for chlorophyll a (PSSRa)	R_{800}/R_{680}	
Pigment specific simple ratio for chlorophyll b (PSSRb)	R_{800}/R_{635}	

Renormalized difference vegetation index (RDVI)	$(R_{800}–R_{670})/(SQRT(R_{800}+R_{670}))$	(Roujean and Breon 1995)
Wavelength of the red edge (RE)	Amplitude of the main peak in the first derivative of the reflectance spectra	(Filella *et al.* 1996)
Red-Edge Inflection Point (REIP)	The position of the red edge inflection point	(Collins 1978; Horler *et al.* 1983)
Red-edge position liner extrapolation method1 (REP1)	$700+40*((R_{re}–R_{700})/(R_{740}–R_{700}))$ $R_{re}=(R_{670}–R_{780})/2$	(Cho and Skidmore 2006)
Red-edge position liner extrapolation method2 (REP2)	$700+40*((R_{670}+R_{780})/2–R_{700})/(R_{740}–R_{700}))$	(Guyot and Baret 1988)
Structure intensive pigment index (SIPI)	$(R_{800}–R_{445})/(R_{800}–R_{680})$	(Peñuelas *et al.* 1995b)
Spectral polygon vegetation index (SPVI)	$0.4*3.7*(R_{800}–R_{670})–1.2*SQRT((R_{530}–R_{670})^2)$	(Vincini *et al.* 2006; Main *et al.* 2011)
SR1	R_{605}/R_{760}	
SR2	R_{695}/R_{760}	
SR3	R_{710}/R_{760}	(Carter 1994)
SR4	R_{695}/R_{420}	
SR5	R_{695}/R_{670}	
SR6	R_{675}/R_{700}	(Chappelle *et al.* 1992)
SR7	R_{750}/R_{550}	(Gitelson and Merzlyak 1996)
SR8	R_{750}/R_{700}	
SR9	R_{752}/R_{690}	
SR10	R_{440}/R_{690}	(Lichtenthaler *et al.* 1996)
SR11	R_{700}/R_{670}	(McMurtrey *et al.* 1994)
SR12	R_{430}/R_{680}	(Peñuelas *et al.* 1995a)
SR13	R_{740}/R_{720}	(Vogelmann *et al.* 1993)
SR14	R_{750}/R_{710}	(Zarco-Tejada *et al.* 2003c)
SR15	R_{565}/R_{740}	(Gong *et al.* 2014)
SR16	R_{1250}/R_{1050}	(Delalieux *et al.* 2009)
SRC	R_{800}/R_{680}	(Jordan 1969)
Sum of the	The sum of the amplitudes between 680 and	(Filella *et al.* 1995)

amplitudes between 680 and 780 nm in the first derivative of the reflectance spectra (Sum1)	780 nm in the first derivative of the reflectance spectra	
Sum of derivative values between 626 nm and 795 nm (Sum2)	The sum of derivative values between 626 nm and 795 nm.	(Elvidge and Chen 1995)
Transformed chlorophyll absorption ratio (TCARI)	$3*((R_{700}–R_{670})–0.2*(R_{700}–R_{550})*(R_{700}/R_{670}))$	(Haboudane *et al.* 2002)
TCARI/OSAVI	TCARI/OSAVI	
Transformed chlorophyll absorption ratio 2 (TCARI2)	$3*((R_{750}–R_{705})–0.2*(R_{750}–R_{550})*(R_{750}/R_{705}))$	(Wu *et al.* 2008)
TCARI2/OSAVI2	TCARI2/OSAVI2	
Triangular chlorophyll index (TCI)	$1.2*(R_{700}–R_{550})–1.5*(R_{670}–R_{550})*SQRT(R_{700}/R_{670})$	(Hunt *et al.* 2011)
Transformed vegetation index (TVI)	$0.5*(120*(R_{750}–R_{550})–200*(R_{670}–R_{550}))$	(Broge and Leblanc 2001)
Voge1	D_{715}/D_{705}	(Vogelmann *et al.* 1993)
Voge2	$(R_{734}–R_{747})/(R_{715}+R_{726})$	

790
791

792 Table 2 Main characteristics of the measurements in this study.

Shading	0%		35%		75%		90%	
Date	01 May	11 May	01 May	11 May	01 May	11 May	01 May	11 May
Number of samples	8	15	12	15	12	15	14	15
Minimum	9.24	24.58	16.84	36.20	19.62	36.41	17.03	29.63
1st Quartile	13.06	31.42	19.03	40.78	24.11	49.39	21.88	46.73
Median	16.01	33.64	22.42	44.28	26.16	51.27	27.02	49.30
Mean	15.74	35.10	23.37	44.79	27.40	51.05	26.77	49.39
3rd Quartile	18.72	38.35	27.21	46.93	30.46	55.01	32.61	54.99
Maximum	21.63	46.25	31.75	55.80	38.51	57.89	38.04	60.60

793

794

795 Table 3. Selected wavelengths (nm) based on GA.

Round	Wavelength (nm)
1	481, 494, 541, 544, 564, 582, 618, 630, 634, 664, 730, 770, 771
2	409, 426, 513, 520, 546, 553, 574, 576, 586, 606, 657, 681, 689, 702, 745, 753
3	535, 556, 565, 571, 609, 612, 620, 643, 650, 736, 755, 766, 768
4	410, 520, 526, 534, 572, 573, 614, 618, 623, 639, 642, 713, 770, 772
5	434, 448, 528, 591, 599, 636, 677, 695, 697, 742, 749, 752, 763
6	401, 419, 432, 492, 625, 648, 672, 677, 706, 736, 737, 749, 777
7	443, 450, 453, 495, 497, 569, 587, 646, 648, 677, 689, 739
8	415, 423, 449, 485, 511, 526, 541, 545, 591, 605, 654, 655, 688, 704, 728, 776
9	427, 472, 485, 493, 518, 527, 546, 596, 602, 619, 659, 672, 675, 711, 728
10	441, 458, 461, 597, 679, 697, 703, 735, 747, 760

796
797

Table 4. Selected indices based on GA.

Round	Index
1	SR2, SR4, MCARI2
2	Datt1, EGFN, NPCI2, mSR2, TCI
3	Mac, D1, MSAVI, RDVI.1
4	PSSRb, SR6, SR14
5	EGFR, MSAVI, OSAVI2, GLI
6	EGFN, mSR2, SR3, GI, MTCI
7	mREIP, PSNDb, SR11, MCARI1, TCI
8	PSNDb, SR8, CI, DDn, DPI, RDVI
9	Datt4, Gitel1, MSAVI, D800_550
10	EGFR, NPCI, TVI

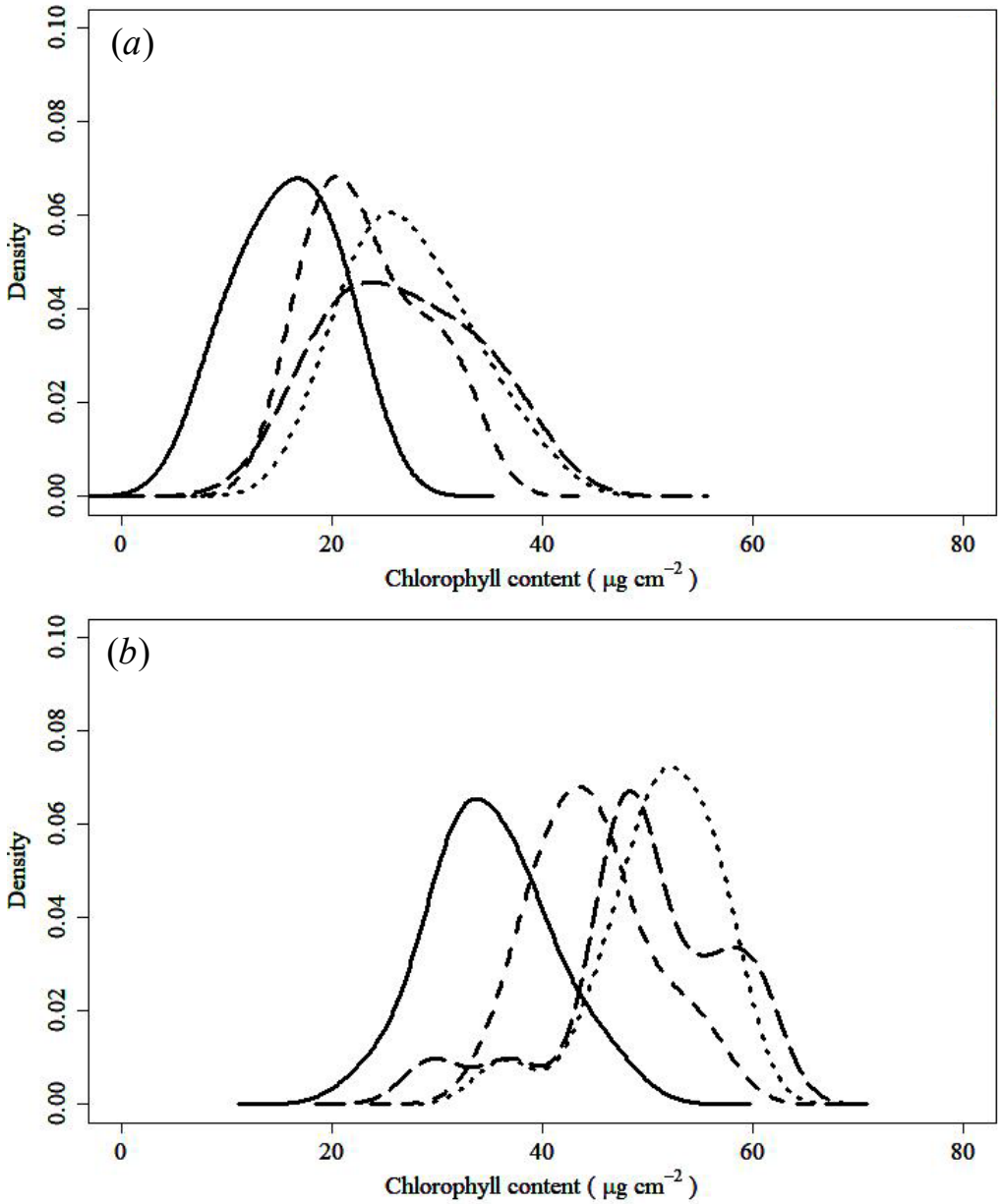
801

802 Table 5. Effective methods for quantifying chlorophyll content.

	Round 1			Round 2			Round 3			Round 4			Round 5		
	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
Machine learning															
Reflectance															
SVM	0.77	4.66	2.46	0.94	3.64	3.87	0.87	4.32	2.83	0.82	5.65	2.36	0.96	2.84	4.32
RF	0.87	4.04	2.84	0.94	3.39	4.15	0.93	3.63	3.37	0.85	4.85	2.75	0.95	3.29	3.73
KELM	0.95	2.57	4.47	0.94	2.83	4.96	0.95	2.85	4.29	0.83	3.58	3.73	0.96	2.59	4.73
Vegetation index															
SVM	0.69	7.87	1.46	0.93	3.60	3.90	0.82	6.16	1.98	0.39	11.03	1.21	0.89	4.10	3.00
RF	0.76	5.51	2.08	0.81	6.08	2.31	0.88	4.24	2.88	0.68	7.66	1.74	0.77	5.99	2.05
KELM	0.92	3.22	3.56	0.95	3.25	4.32	0.74	6.32	1.94	0.67	7.60	1.76	0.83	4.97	2.47
Model inversion															
PROSPECT-D	0.69	6.71	1.71	0.79	6.87	2.05	0.77	6.30	1.94	0.72	7.45	1.79	0.81	6.41	1.91
	Round 6			Round 7			Round 8			Round 9			Round 10		
	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
Machine learning															
Reflectance															
SVM	0.96	3.09	5.06	0.92	4.19	3.29	0.95	3.57	4.20	0.74	6.00	2.28	0.84	4.87	2.57
RF	0.96	3.49	4.48	0.95	3.12	4.42	0.90	4.86	3.08	0.88	4.70	2.91	0.94	3.97	3.15
KELM	0.97	2.65	5.92	0.93	2.82	4.89	0.94	3.75	4.00	0.92	4.05	3.38	0.96	2.66	4.70
Vegetation index															
SVM	0.69	14.50	1.08	0.62	8.70	1.58	0.52	10.33	1.45	0.84	5.71	2.40	0.51	11.80	1.06
RF	0.83	7.03	2.23	0.71	7.75	1.78	0.78	7.08	2.12	0.85	5.47	2.50	0.54	8.85	1.41
KELM	0.25	13.36	1.17	0.76	7.25	1.90	0.85	6.07	2.47	0.80	6.48	2.11	0.58	8.27	1.51
Model inversion															
PROSPECT-D	0.92	6.78	2.31	0.81	7.93	1.74	0.76	8.38	1.79	0.83	5.92	2.31	0.77	6.62	1.89

803

804 **Figures**



805

806 Figure 1. Histograms of chlorophyll content on (a) 1 May and (b) 11 May. Continuous, dashed, dotted

807 and long dashed lines represent distributions of chlorophyll content after 0 %, 35 %, 75 % and 90 %

808 shading, respectively.

809

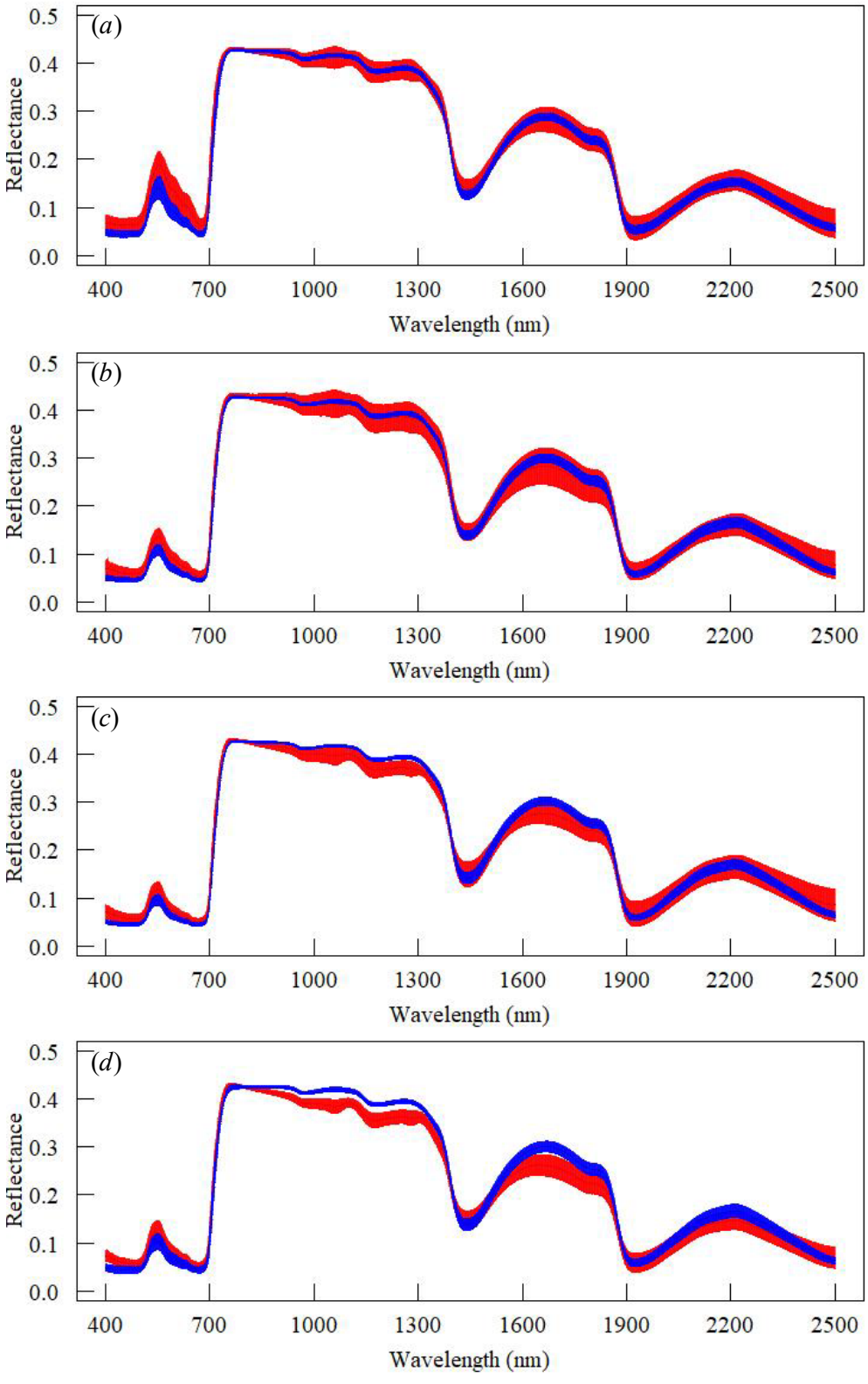


Figure 2. Mean reflectance spectra and standard deviations for (a) 0% shading, (b) 35% shading, (c)75% shading and (d) 90% shading. Red and blue represent measurements on 1 May and 11 May, respectively.

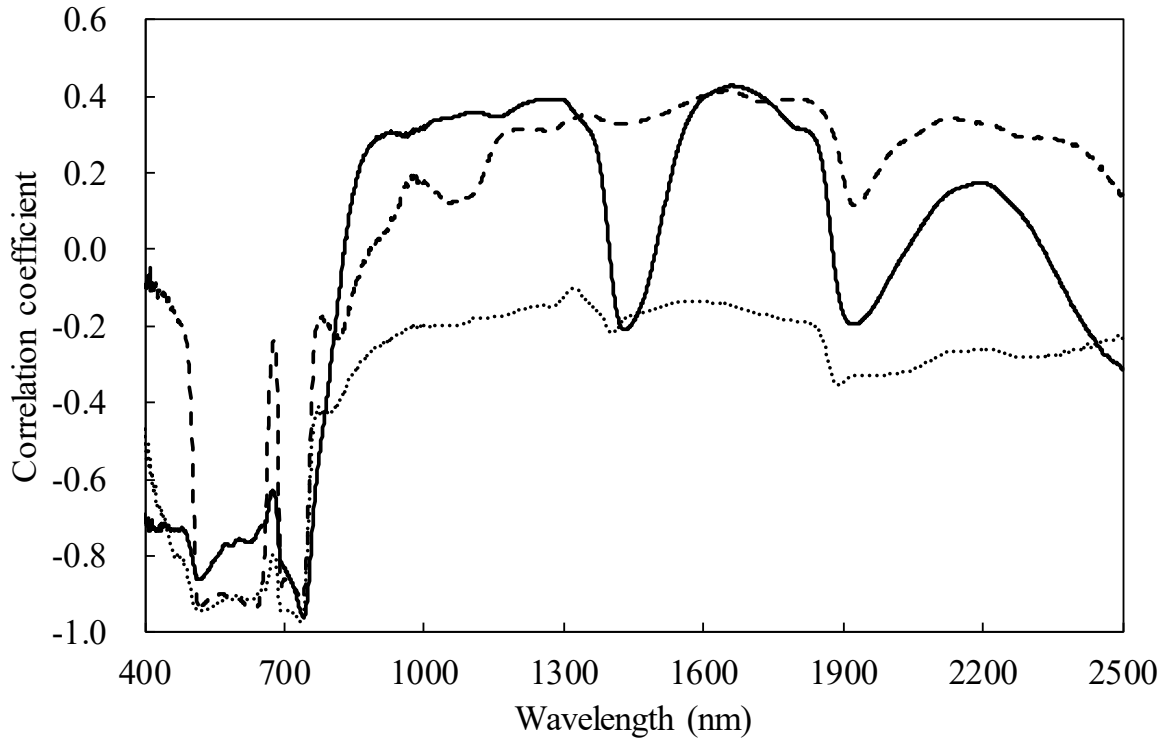
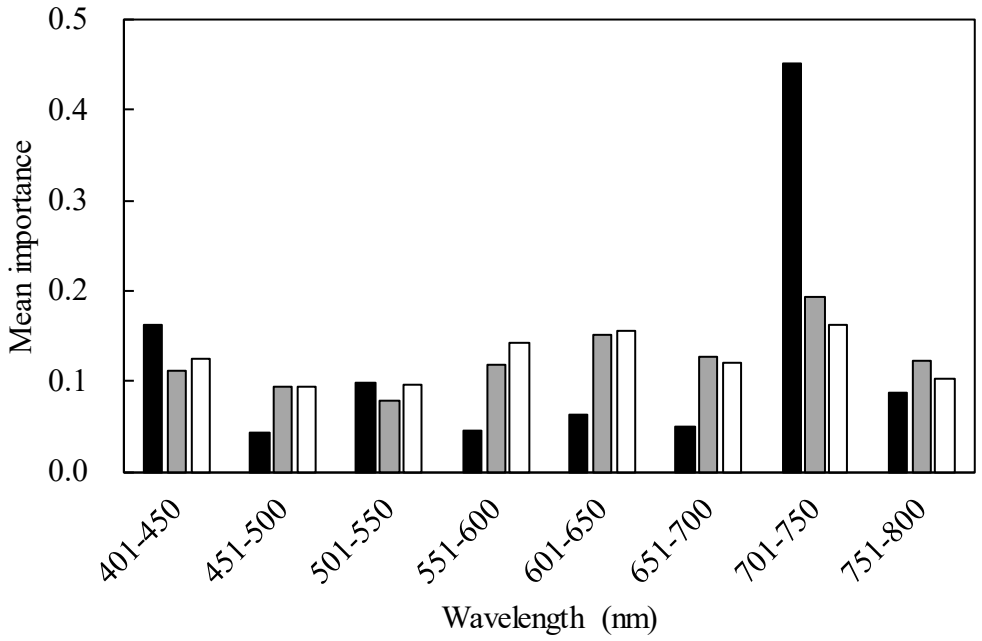


Figure 3. Correlations between reflectance and chlorophyll content. Continuous, dotted and broken lines represent correlation coefficients for all, on 1 May and on 11 May, respectively.



820

821 Figure 4. DSA results for RF (black), SVM (grey) and KELM(white). Importance values were
822 averaged for ten repetitions.