ISSN 1392-3196 / e-ISSN 2335-8947
Zemdirbyste-Agriculture, vol. 102, No. 1 (2015), p. 51−58
DOI 10.13080/z-a.2015.102.006

Determination of lettuce nitrogen content using spectroscopy with efficient wavelength selection and extreme learning machine

Hongyan GAO, Hanping MAO, Xiaodong ZHANG

Key Laboratory of Modern Agricultural Equipment and Technology,
Ministry of Education and Jiangsu Province, Jiangsu University
Zhenjiang Jiangsu, 212013, China
E-mail: maohp@ujs.edu.cn

Abstract

Measuring the nitrogen content of plants is useful for nitrogen fertilizer management. The aim of this work is to explore the use of spectroscopy for estimating lettuce leaf nitrogen content. Leaf reflectance spectra were measured using a spectroradiometer with a range of 350–2500 nm, and 160 fresh lettuce leaves given five different nitrogen treatments were used for spectra acquisition and total nitrogen determination. Interval partial least squares (iPLS), synergy interval partial least squares (siPLS), and siPLS combined with genetic algorithms (GA-siPLS) were used to select the optimal intervals or variables. Partial least squares (PLS) and extreme learning machine (ELM) methods were used to develop calibration models. One characteristic wavelength region (470–590 nm) using 120 variables was selected by iPLS; four characteristic wavelength regions (440–530, 530–620, 620–710 and 890–980 nm) using 362 variables were selected by siPLS, and 56 wavelength variables were selected by GA-siPLS. Six different regression models were established for nitrogen content by PLS and ELM based on optimal intervals or variables. The results imply that GA-siPLS is an efficient variable selection algorithm and ELM is a successful nonlinear regression tool. Furthermore, GA-siPLS combined with ELM is a feasible method for measuring nitrogen content in lettuce, as it performed better than other models. The optimal results achieved a root mean square error of prediction (RMSEP) = 0.2890% and correlation coefficient Rp = 0.9218.

Key words: extreme learning machine, nitrogen, partial least squares, spectroscopy, wavelength selection.

Introduction

In agriculture, optimal management and the efficient use of fertilizer are necessary to improve yield and quality as well as to reduce production costs and environmental pollution. Among the essential macronutrients, nitrogen has been described as the most important element for vegetative growth (Debouba et al., 2007; Kumagai et al., 2009; Sosa et al., 2012; Bavec et al., 2013). Because nitrogen is directly related to the photosynthetic process, it is an essential element throughout plant growth (Andrews et al., 2013). Many growers prefer to classify crop nitrogen status by leaf samples at an early stage, when nitrogen deficiency can still be remedied. By the time symptoms are clearly visible, irreparable damage to the crop may have already occurred (Abdel-Rahman et al., 2010). Early diagnosis plays a very important role in nutrition regulation. In general, several approaches to measure nutrient status in plants have been developed and evaluated. The most common method is chemical analysis, which is a laborious, lengthy and destructive measurement (Vigneau et al., 2011). To address this problem, spectroscopy is a fast, accurate and non-destructive technique that has been adopted as an important analytical method in agriculture (Jain et al., 2007; Ulissi et al., 2011). The use of spectral reflectance characteristics enables the detection of stress before the symptoms appear.

However, spectroscopy information is abundant and redundant. Not all regions or wavelengths are significantly associated with the lettuce nitrogen content, and the selection of specific spectral regions or variables is key to decreasing the complexity of the calibration model and increasing its prediction precision (Leardi, 2000; Nørgaard et al., 2000). Spectral variables are selected manually, which, unless guided by previous experience and knowledge, can reduce the performance of the model (Jiang et al., 2012). Principal component analysis (PCA) and successive projections algorithm (SPA) have frequently been used to perform wavelength selection. However, PCA does not take into account the relation of the spectral data to the specified object, and SPA depends on the number of samples (Otero et al., 2011). Interval partial least squares (iPLS) and synergy interval partial least squares (siPLS) can search for the optimal spectral intervals, which are the most important information to model the parameter of interest. iPLS divides the whole spectral region into different intervals and searches for the optimal interval for establishing a partial least squares (PLS) model; siPLS is another advanced algorithm that splits the whole spectral region into different intervals and combines two, three or four spectral intervals to establish the PLS model. The optimal intervals are selected by cross validation and determined

according to the lowest root mean square error of cross validation (RMSECV). In addition, a genetic algorithm (GA) is an adaptive heuristic search algorithm for selecting spectral variables. When combined with siPLS, the method is called GA-siPLS.

Among the different regression models for multivariate calibration, the extreme learning machine (ELM) has excellent performance with respect to regression problems and has some advantages compared to other conventional learning methods. The aim of this study is to select efficient intervals or variables using iPLS, siPLS, and GA-siPLS, and establish different ELM models based on the three algorithms to rapidly determine nitrogen content in lettuce.

Materials and methods

Sample preparation. Lettuce (*Lactuca sativa* L., Woshu Seeds Co. Ltd., China) seedlings of five true leaves were transplanted individually into pots filled with perlite in a greenhouse at Jiangsu University in China from 11 May to 26 June, 2012 in the spring. The average temperature and humidity of the greenhouse were respectively 23.14ºC and 54.46%. The Yamasaki lettuce recipe was used in the irrigation. The contents of the nutrient solution were: $Ca(NO_3)_2 \cdot 4H_2O$ 236 mg L^{-1}, KNO_3 404 mg L^{-1}, $NH_4H_2PO_4$ 57 mg L^{-1}, $MgSO_4 \cdot 7H_2O$ 123 mg L^{-1}, Fe-EDTA 16 mg L^{-1}, $MnCl_2 \cdot 4H_2O$ 1.2 mg L^{-1}, H_3BO_3 0.72 mg L^{-1}, $ZnSO_4 \cdot 4H_2O$ 0.09 mg L^{-1}, $CuSO_4 \cdot 5H_2O$ 0.04 mg L^{-1} and $(NO_4)_2Mo_7O_4$ 0.01 mg L^{-1} (Wang et al., 2000). Among them, NO_3^- was 6 mmol L^{-1} and NH_4^+ was 0.5 mmol L^{-1}. Nutrient solutions with five different levels of nitrogen requirements were prepared: 1) NO_3^- 1.5 mmol L^{-1} and NH_4^+ 0.125 mmol L^{-1}, 2) NO_3^- 3 mmol L^{-1} and NH_4^+ 0.25 mmol L^{-1}, 3) NO_3^- 4.5 mmol L^{-1} and NH_4^+ 0.375 mmol L^{-1}, 4) NO_3^- 6 mmol L^{-1} and NH_4^+ 0.5 mmol L^{-1}, and 5) NO_3^- 7.5 mmol L^{-1} and NH_4^+ 0.625 mmol L^{-1}. The roots of the lettuce were always kept in a fixed nutrient solution content by a custom-built timing irrigation and collection system. Imidacloprid was used to exterminate the greenhouse whitefly (*Trialeurodes vaporariorum*) at the end of growth period.

Spectra measurement. Thirty-two lettuces were randomly selected from each nitrogen treatment for reflectance measurement two weeks after the transplanting, yielding 160 samples. The leaves for measurement were taken from the lower part of the plants. The leaves' spectral reflectance was measured using a portable spectroradiometer FieldSpec® 3 (Analytical Spectral Devices Inc., USA) with a wavelength range of 350–2500 nm. The spectral readings were acquired in 1 nm increments. The leaf spectral data were collected by a fibre optic cable with a 25° field of view from about 50 mm above the leaves. We used a 50 W halogen light source (Analytical Spectral Devices Inc.) that was positioned 300 mm away from the leaf sample with a 45° zenith illumination angle to produce the radiation in a dark room. Five replicate spectral measurements were taken for each leaf in order to reduce the noise by averaging the spectra. The air temperature and humidity were kept at a steady level in the laboratory. Relative reflectance spectra were calculated by dividing the leaf radiance by the reference radiance from a spectralon white reference panel (Labsphere Inc., USA) for each wavelength.

Sample processing and chemical analysis. Spectral data collection and sample drying were carried out on the same day. The fresh leaves were oven-dried at 65°C for 24 h and then ground and oven-dried again at 65°C to a constant weight, resulting in leaf powder. The leaf powder was digested by 98% sulphuric acid (w/w) until the solution was transparent. The total nitrogen content in leaves was measured using the Kjeldahl method by continuous-flow analyzer Auto Analyzer 3-AA3 (SEAL Analytical Co., UK) and expressed as a percentage of dry weight.

Multivariate calibration methods. In order to develop an optimal calibration model for total nitrogen content, the analytical samples were reasonably divided into a calibration set and a prediction set, the former was utilized to construct model, and the latter was used to verify the robustness of the model. The performance of optimal model was evaluated using the root mean square error of cross validation (RMSECV) and the correlation coefficient R_c in the calibration set, they were calculated as equations (1) and (2):

$$RMSECV = \sqrt{\frac{1}{n_c}\sum_{i=1}^{n_c}(y_{ci} - \hat{y}_{ci})^2} \quad (1),$$

$$R_c = \sqrt{1 - \frac{\sum_{i=1}^{n_c}(y_{ci} - \hat{y}_{ci})^2}{\sum_{i=1}^{n_c}(y_{ci} - \bar{y}_c)^2}} \quad (2),$$

where n_c is the number of samples in the calibration set, y_{ci} – the reference measurement value of the *i*th sample, $\hat{y}_{ci}$ – the estimation value for the *i*th sample by the constructed model when the *i*th sample is left out, $\bar{y}_c$ – the average of all reference measurement values of samples in the calibration set.

The optimal model was verified by the independent samples in the prediction set. The root mean square error of prediction (RMSEP) and the correlation coefficient R_p were calculated as equations (3) and (4):

$$RMSEP = \sqrt{\frac{1}{n_p}\sum_{i=1}^{n_p}(y_{pi} - \hat{y}_{pi})^2} \quad (3),$$

$$R_p = \sqrt{1 - \frac{\sum_{i=1}^{n_p}(y_{pi} - \hat{y}_{pi})^2}{\sum_{i=1}^{n_p}(y_{pi} - \bar{y}_p)^2}} \quad (4),$$

where n_p is the number of samples in the prediction set, y_{pi} – the reference measurement value of the *i*th sample in the prediction set, $\hat{y}_{pi}$ – the estimation value of the model of the *i*th sample in the prediction set, $\bar{y}_p$ – the average of all reference measurement values of samples in the prediction set.

Statistical analysis. All methods were implemented in software *MATLAB R2009a* (MathWorks, USA). The algorithms iPLS (interval partial least squares) and siPLS (synergy interval partial least squares) were developed by Nørgaard et al. (2000), and the algorithm GA-PLS (partial least squares combined with genetic algorithms) is the one proposed by Leardi (2000).

Results

Reference measurement of total nitrogen. In order to avoid bias in subset selection, the calibration set was comprised of 75% of the total samples. Therefore, the calibration and prediction sets were created with 120

and 40 separate samples, respectively. Table 1 shows the number of samples used and the statistics associated with the nitrogen content measured by the Kjeldahl method in the laboratory. The actual nitrogen content range was 2.23–5.54% in the calibration set and 2.27–5.51% in the prediction set. The mean value and standard deviation indicated that there were no significant differences in the calibration and prediction sets. This proved that the distribution of the samples in the calibration and prediction sets was appropriate.

Table 1. Results of nitrogen content analysis in the calibration and prediction sets

Data	Number of samples	Minimum %	Maximum %	Mean %	Standard deviation %
Calibration	120	2.23	5.54	3.89	0.89
Prediction	40	2.27	5.51	3.91	0.86

Spectral data preprocessing. Figure 1 presents the raw spectra of all samples in the calibration set. The raw spectra contains the background information and noise in addition to the sample information. A first-order derivative transformation can be a good noise filter because it eliminates baseline drifts, reduces the effects of multiple scattering of radiation caused by sample surface roughness, enhances small spectra differences, and locates the positions of absorption features on the spectra (Abdel-Rahman et al., 2010; Miphokasap et al., 2012). Figure 2 shows the spectra after applying a first-order derivative transform to the data of the original spectra.

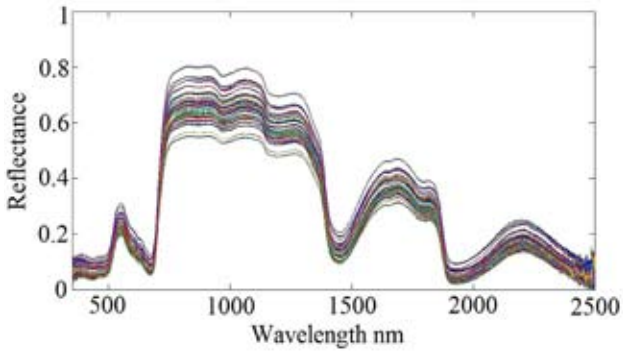

Figure 1. Raw spectra of lettuce leaves

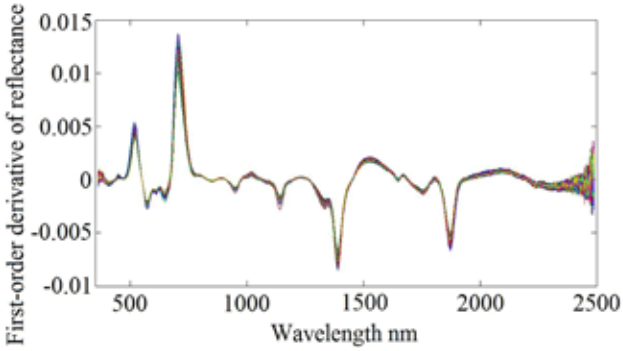

Figure 2. First-order derivative spectra of lettuce leaves

Selection of optimal spectra intervals by interval partial least squares (iPLS). Collinearity and information redundancy exists in spectral data. The algorithm iPLS focuses on important spectral intervals and removes interference in the full-spectral region (Nørgaard et al., 2000; Ferrao et al., 2011). The total spectral region (350–2500 nm) was divided into equidistant non-overlapping intervals, and each interval underwent a separate PLS modelling to determine the most useful variable range. First, the whole spectral region was divided into 16 cases of 15 to 25 equidistant intervals, respectively. For each case, the lowest RMSECV is shown in Figure 3. We can see that the iPLS model produced the best results (RMSECV = 0.4621%) when the spectral region was divided into 18 intervals. Next, 18 PLS models were established and the corresponding RMSECV were calculated. In Figure 4, bar plots of the RMSECV are shown for each range of variables and for the models developed with one to six principal components (PCs) in italics. The dotted line represents the RMSECV of the full-spectrum PLS model. It can be seen that the optimal range chosen was number 2, because its RMSECV, which included 120 variables, was below the others.

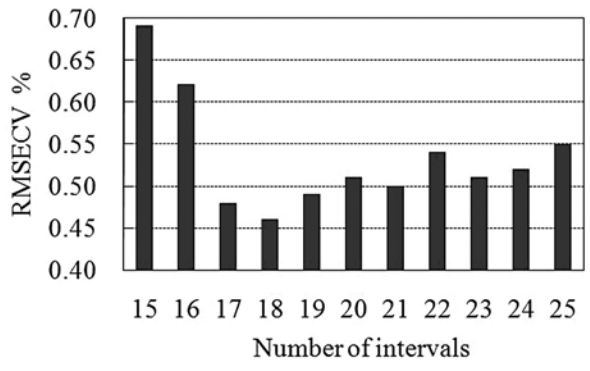

Figure 3. Number of intervals optimized according to the root mean square error of cross validation (RMSECV) for the iPLS model

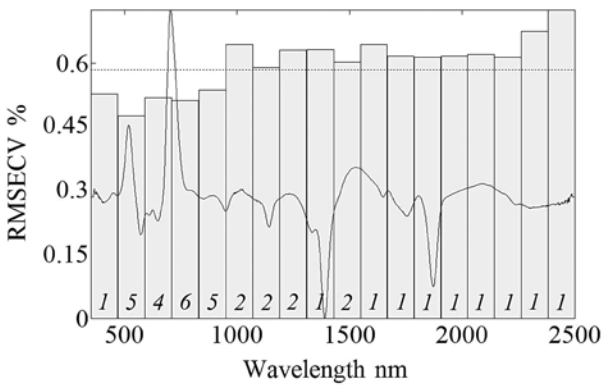

Figure 4. Performance of the root mean square error of cross validation (RMSECV) at 18 intervals

Selection of optimal spectra intervals using synergy interval partial least squares (siPLS). The algorithm siPLS is the expansion of the algorithm iPLS (McGoverin et al., 2010; Muller et al., 2012). The whole spectral region was also divided into 16 cases of 15 to 25 equidistant intervals, respectively, and PLS regression models were constructed by all permutations and combinations of two, three, or four intervals. For each case, the lowest RMSECV is listed in Table 2. It can be seen that the optimal siPLS model with the lowest RMSECV of 0.3764% is achieved using a combination of four intervals chosen from 24 intervals. The performance parameters of this optimal model are shown in bold. The best number of PCs is 10, and the optimal combination of intervals selected is [2 3 4 7], as seen in Figure 5, which correspond to 440–530, 530–620, 620–710 and 890–980 nm, using 362 variables.

Table 2. Calibration results using the siPLS model with different spectral interval selections

Number of intervals	Number of models	Selected intervals	PCs	Calibration set		Prediction set	
				Rc	RMSECV %	*Rp*	RMSEP %
15	455	[2 3 4]	7	0.8425	0.4697	0.8216	0.5105
16	560	[2 3 4]	10	0.8307	0.4750	0.8198	0.5146
17	680	[2 3 5]	10	0.8696	0.4142	0.8511	0.4738
18	816	[2 3 5]	11	0.8888	0.3839	0.8588	0.4632
19	3876	[2 3 5 6]	10	0.8538	0.4534	0.8247	0.5017
20	4845	[2 3 4 6]	10	0.8725	0.4100	0.8541	0.4709
21	5985	[2 3 4 6]	9	0.8626	0.4244	0.8369	0.4961
22	7315	[2 3 4 6]	11	0.8728	0.4098	0.8443	0.4814
23	8855	[3 4 6 7]	8	0.8607	0.4361	0.8127	0.5222
24	*10626*	*[2 3 4 7]*	*10*	*0.8912*	*0.3764*	*0.8597*	*0.4565*
25	2300	[2 4 7]	9	0.8631	0.4232	0.8493	0.4801

PCs – principal components, RMSECV – root mean square error of cross validation

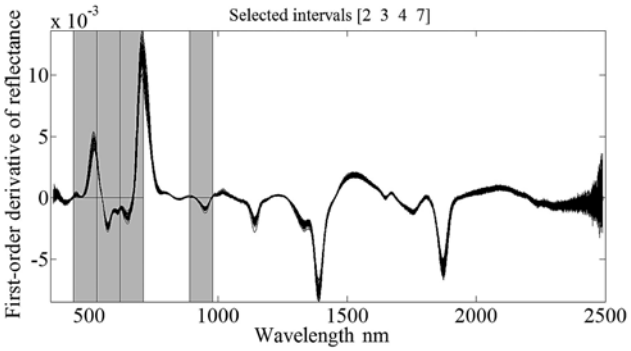

Figure 5. Optimal spectral intervals selected by siPLS model

Selection of optimal spectra variables by algorithm partial least squares combined with genetic algorithms (GA-PLS). It can be seen that the algorithm siPLS performed better than the algorithm iPLS with respect to the measurement of nitrogen content of lettuce leaves (Table 3). Although the intervals selected by iPLS and siPLS removed numerous useless spectral regions, there were still many collinear wavelengths in the spectral interval, and two variables from neighbouring wavelengths could also be collinear. These collinear wavelengths reduce the prediction precision and increase the complexity of the PLS model. Therefore, GA-PLS was used to further eliminate the collinear variables based on

Table 3. Comparison of results based on different regression models

Models	Variables	PCs	Calibration set		Prediction set	
			Rc	RMSECV %	*Rp*	RMSEP %
iPLS (interval partial least squares)	120	5	0.8671	0.4621	0.8249	0.5016
siPLS (synergy interval partial least squares)	362	10	0.8912	0.3764	0.8597	0.4565
GA-siPLS (siPLS combined with genetic algorithms)	56	7	0.9249	0.2947	0.8915	0.3671
iELM (interval extreme learning machine)	120	5	0.8812	0.3873	0.8624	0.4138
siELM (synergy interval extreme learning machine)	362	9	0.9284	0.2914	0.9091	0.3539
GA-siELM (siELM combined with genetic algorithms)	56	8	0.9479	0.2573	0.9218	0.2890

PCs – principal components, RMSECV – root mean square error of cross validation, RMSEP – root mean square error of prediction

some intervals selected by siPLS. The genetic algorithm (GA) control values were set to the following: the initial population size was 30 chromosomes, probability of mutation was 0.01, probability of cross-over was 0.5, number of stopping iterations was 100, and number of "replications" was 10.

In some cases, GA cannot be used because it models the noise instead of the information. To verify whether this is the case, a randomization test can be performed (Leardi, Lupiáñez González, 1998). In this study, the result of the randomization test was 3.1621, showing that GA could safely be applied and the stop criterion should be 178 evaluations. Figure 6 shows the histogram of frequencies for which each variable selected, indicating which variables of the spectra are selected more often and which variables are seldom selected. The variables are used for establishing the PLS model if the frequency of selection reaches or exceeds the black horizontal line. The optimum model was developed when 56 variables and seven PCs were used.

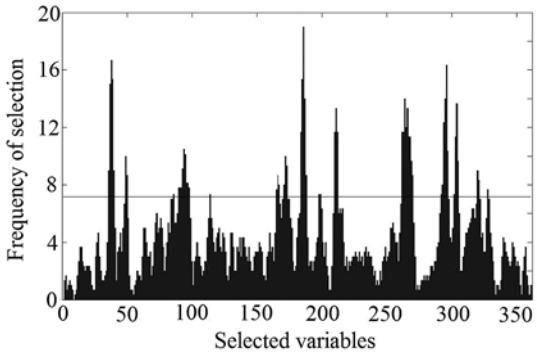

Figure 6. Selection frequency of each variable by genetic algorithm (GA)

Results of extreme learning machine (ELM) models based on optimal intervals and variables. The ELM is one of the most popular and important learning algorithms, proposed by Huang et al. (2006). It has excellent performance on regression problems (Yuan et al., 2011). In this study, ELM models were established based on the optimal intervals or variables selected by iPLS, siPLS and GA-siPLS. In the modelling, the PCs of the optimal intervals or variables were used as the input and the optimum number of PCs was determined according to the lowest RMSECV. Furthermore, the nitrogen content of the lettuce leaves was used as the output, the input weights and hidden biases were randomly generated, and the activation function was set to "sigmoidal". The only parameter that needed to be determined was the number of hidden nodes, which was obtained by trial and error. Because the number of samples in the calibration set was 120, the number of hidden nodes was varied from 10 to 120 in intervals of five. The optimal GA-siELM model for nitrogen content was obtained using 35 hidden nodes; the lowest RMSECV could be obtained when using eight PCs. The optimal results were obtained with R_c = 0.9479 and RMSECV = 0.2573% in the calibration set, R_p = 0.9218, RMSEP = 0.2890% in the prediction set. Figure 7 shows the correlation between the prediction values and reference measurements in the prediction set. Each number represents one sample. The optimal results of the iELM and siELM models are given in Table 3.

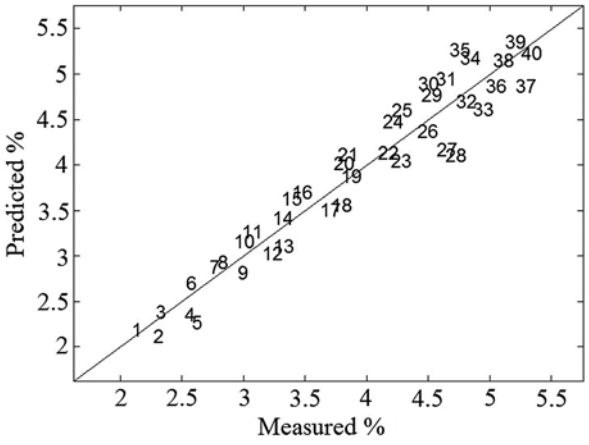

Figure 7. Reference measurement versus prediction for nitrogen content in the prediction set by GA-siELM model

Discussion

Comparison and discussion of the results. Table 3 shows the results from six different regression models: iPLS, siPLS, GA-siPLS, iELM, siELM and GA-siELM. The results imply that GA-siPLS is an efficient variable selection algorithm and ELM is a successful nonlinear regression tool. Further, GA-siPLS combined with ELM is a feasible means by which to measure nitrogen content in lettuce. The main reasons for this can be summarized as follows.

In the range 350–2500 nm, there are 2151 wavelengths. The large number of spectral variables often renders the predictions of a dependent variable as unreliable (Balabin, Smirnov, 2011). However, most of these wavelengths are collinear and redundant variables with respect to nitrogen in lettuce leaves, and selecting which and how many of variables to use are key factors to enhancing prediction model precision and adaptability. For the algorithm iPLS, only one interval was selected to establish the PLS model, and most of uncorrelated and redundant variables were removed. However, some useful variables were also deleted, resulting in the poor performance of the iPLS (Jiang et al., 2012). In the algorithm siPLS, the whole spectral region was divided into 24 intervals. Four intervals were then used to establish the model PLS, and the results on the calibration and prediction sets indicate that accuracy of the model siPLS is better than model iPLS. However, there still exist two bottlenecks to using siPLS: the run time is too long because 10,626 models had to be calculated and the complexity of the model siPLS is higher than the model iPLS because of the large number of variables. With respect to the second problem, we can conclude that some collinear variables still exist in the four intervals. Thus, GA-siPLS was presented to reduce the number of

variables and enhance the accuracy. Finally, 56 variables were selected from the four intervals. The performance of the GA-siPLS was better than siPLS and the complexity of the model was greatly simplified.

Chlorophylls and proteins are the main nitrogen sources in green leaves. In this work, most wavebands selected for nitrogen content estimation are within the visible region and are known features of pigment absorption. The literature (Min et al., 2006; Jain et al., 2007; Ustin et al., 2009) on the strong correlation between nitrogen content and chlorophylls has been critically reviewed. Chlorophylls exhibit strong absorption arising from conjugated C-C single and double bonds of the porphyrin ring in the visible region (Katz et al., 1966). The infrared spectra of chlorophylls show strong absorption because of the C-H bonds in the phytol tail of the molecule (Katz et al., 1966). Proteins are the other major primary nitrogenous constitute in leaves. They may have an effect on nitrogen prediction, but the moisture content of lettuce is very high (93–95%), and hence water and cellular structure may obscure protein information. Therefore, there is little useful information contained in the middle infrared region.

PLS, as a linear algorithm, achieved good performance. The increase and decrease in nitrogen fertilizer causes a complex dynamic process in lettuce that is accompanied by changes in chemical components. Therefore, the relationship between the changes of chemical components and the nitrogen content is very complicated and tends to exhibit a nonlinear correlation. ELM is a typically nonlinear algorithm, and is stronger than PLS with respect to the level of self-learning and self-adjustment (Chen et al., 2012). Hence, it was useful for improving the prediction performance of the model.

Conclusions

Characteristic variables selection combined with extreme learning machine (ELM) was used to build a model for the non-destructive detection of nitrogen content in lettuce leaves.

1. The four characteristic wavelength regions (440–530, 530–620, 620–710 and 890–980 nm) including 362 variables were selected by model siPLS, and then genetic algorithm (GA) were further used to reduce the number of variables. Finally, model GA-siPLS reduced the number of wavelengths from 2151 to 56.

2. ELM was superior to linear algorithm and showed great potential in this quantitative study. The optimal model GA-siELM was achieved with root mean square error of prediction RMSEP = 0.2890% and Rp = 0.9218. It can be concluded the model GA-siELM has high potential as a method to determine the nitrogen content in lettuce leaves and provide evidence for nutrition regulation to growers.

Acknowledgements

The authors gratefully acknowledge financial support by the Key Program of National Natural Science Foundation of China (Grant No. 61233006), National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant No. 2014BAD08B03), Scientific Innovation Research of College Graduate in Jiangsu Province (Grant No. CXZZ13_0690) and the Project Funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions. We are grateful to reviewers for valuable comments and suggestions on the manuscript.

Received 14 07 2014
Accepted 19 12 2014

References

Abdel-Rahman E. M., Ahmed F. B., van den Berg M. 2010. Estimation of sugarcane leaf nitrogen concentration using *in situ* spectroscopy. International Journal of Applied Earth Observation and Geoinformation, 12: S52–S57 http://dx.doi.org/10.1016/j.jag.2009.11.003

Andrews M., Raven J. A., Lea P. J. 2013. Do plants need nitrate? The mechanisms by which nitrogen form affects plants. Annals of Applied Biology, 163 (2): 174–199 http://dx.doi.org/10.1111/aab.12045

Balabin R. M., Smirnov S. V. 2011. Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. Analytica Chimica Acta, 692 (1): 63–72 http://dx.doi.org/10.1016/j.aca.2011.03.006

Bavec F., Bavec M., Fekonja M. 2013. Organic and mineral nitrogen fertilizers in sweet maize (*Zea mays* L. *saccharata* Sturt.) production under temperate climate. Zemdirbyste-Agriculture, 100 (3): 243–250 http://dx.doi.org/10.13080/z-a.2013.100.031

Chen Q. S., Ding J., Cai J. R., Zhao J. W. 2012. Rapid measurement of total acid content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. Food Chemistry, 135 (2): 590–595 http://dx.doi.org/10.1016/j.foodchem.2012.05.011

Debouba M., Ghorbel M. H., Gouia H. 2007. Effects of NaCl and nitrogen deficiency en azote on chlorophyll fluorescence of PSII in tomato (*Solanum lycopersicon*, Chibli F1). Acta Botanica Gallica, 154 (4): 635–642 (in French) http://dx.doi.org/10.1080/12538078.2007.10516084

Ferrao M. F., Viera M. D., Panta Pazos R. E., Fachini D., Gerbase A. E., Marder L. 2011. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. Fuel, 90 (2): 701–706 http://dx.doi.org/10.1016/j.fuel.2010.09.016

Huang G. B., Zhu Q. Y., Siew C. K. 2006. Extreme learning machine: theory and applications. Neurocomputing, 70: 489–501 http://dx.doi.org/10.1016/j.neucom.2005.12.126

Jain N., Ray S. S., Singh J. P., Panigrahy S. 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. Precision Agriculture, 8 (4–5): 225–239 http://dx.doi.org/10.1007/s11119-007-9042-0

Jiang H., Liu G. H., Mei C. L., Yu S., Xiao A. H., Ding Y. H. 2012. Measurement of process variables in solid-state fermentation of wheat straw using FT-NIR spectroscopy and synergy interval PLS algorithm. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, 97: 277–283 http://dx.doi.org/10.1016/j.saa.2012.06.024

Katz J. J., Dougherty R. C., Boucher L. J. 1966. Infrared and nuclear magnetic resonance spectroscopy of chlorophyll. The Chlorophylls, p. 185–251

Kumagai E., Araki T., Ueno O. 2009. Effect of nitrogen-deficiency on midday photoinhibition in flag leaves of different rice (*Oryza sativa* L.) cultivars. Photosynthetica, 47 (2): 241–246 http://dx.doi.org/10.1007/s11099-009-0038-z

Leardi R. 2000. Application of genetic algorithm-PLS for feature selection in spectral data sets. Journal of Chemometrics, 14 (5–6): 643–655 http://dx.doi.org/10.1002/1099-128X(200009/12)14:5/6<643::AID-CEM621>3.0.CO;2-E

Leardi R., Lupiáñez González A. 1998. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemometrics and Intelligent Laboratory Systems, 41 (2): 195–207 http://dx.doi.org/10.1016/S0169-7439(98)00051-3

McGoverin C. M., Weeranantanaphan J., Downey G., Manley M. 2010. The application of near infrared spectroscopy to the measurement of bioactive compounds in food commodities. Journal of Near Infrared Spectroscopy, 18 (2): 87–111 http://dx.doi.org/10.1255/jnirs.874

Min M., Lee W. S., Kim Y. H., Bucklin R. A. 2006. Nondestructive detection of nitrogen in Chinese cabbage leaves using VIS-NIR spectroscopy. HortScience, 41 (1): 162–166

Miphokasap P., Honda K., Vaiphasa C., Souris M., Nagai M. 2012. Estimating canopy nitrogen concentration in sugarcane using field imaging spectroscopy. Remote Sensing, 4 (6): 1651–1670 http://dx.doi.org/10.3390/rs4061651

Muller A. L. H., Picoloto R. S., Mello P. D., Ferrao M. F., dos Santos M. D. P., Guimaraes R. C. L., Muller E. I., Flores E. M. M. 2012. Total sulfur determination in residues of crude oil distillation using FT-IR/ATR and variable selection methods. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, 89: 82–87 http://dx.doi.org/10.1016/j.saa.2011.12.001

Nørgaard L., Saudland A., Wagner J., Nielsen J. P., Munck L., Engelsen S. B. 2000. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Apply Spectroscopy, 54 (3): 413–419 http://dx.doi.org/10.1366/0003702001949500

Otero R. L. S., Galvao R. K. H., Araujo M. C. U., Cavalheiro E. T. G. 2011. Thermogravimetric determination of L-ascorbic acid in non-effervescent formulations using multiple linear regression with temperature selection by the successive projections algorithm. Thermochimica Acta, 526: 200–204 http://dx.doi.org/10.1016/j.tca.2011.09.020

Sosa A., Padilla J., Ortiz J., Etchevers J. D. 2012. Biomass accumulation and its relationship with the demand and concentration of nitrogen, phosphorus, and potassium in lettuce. Communications in Soil Science and Plant Analysis, 43 (1–2): 121–133 http://dx.doi.org/10.1080/00103624.2012.634695

Ulissi V., Antonucci F., Benincasa P., Farneselli M., Tosti G., Guiducci M., Tei F., Costa C., Pallottino F., Pari L., Menesatti P. 2011. Nitrogen concentration estimation in tomato leaves by VIS-NIR non-destructive spectroscopy. Sensors, 11 (6): 6411–6424 http://dx.doi.org/10.3390/s110606411

Ustin S. L., Gitelson A. A., Jacquemoud S., Schaepman M., Asner G. P., Gamon J. A., Zarco-Tejada P. 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing of Environment, 113: S67–S77 http://dx.doi.org/10.1016/j.rse.2008.10.019

Vigneau N., Ecarnot M., Rabatel G., Roumet P. 2011. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in wheat. Field Crops Research, 122 (1): 25–31 http://dx.doi.org/10.1016/j.fcr.2011.02.003

Wang J. X., Wang Z. H., He G. X. 2000. Vegetable soilless cultivation practical technology. China Agricultural University (in Chinese)

Yuan Y. B., Wang Y. G., Cao F. L. 2011. Optimization approximation solution for regression problem based on extreme learning machine. Neurocomputing, 74: 2475–2482 http://dx.doi.org/10.1016/j.neucom.2010.12.037

ISSN 1392-3196 / e-ISSN 2335-8947
Zemdirbyste-Agriculture, vol. 102, No. 1 (2015), p. 51−58
DOI 10.13080/z-a.2015.102.006

Azoto kiekio nustatymas salotose taikant spektroskopijos, efektyvaus bangos ilgio ir ekstremalaus mokymosi algoritmo metodus

H. Gao, H. Mao, X. Zhang

Kinijos Jiangsu universiteto Jiangsu provincijos Švietimo ministerijos
Šiuolaikinių žemės ūkio įrengimų ir technologijų pagrindinė laboratorija

Santrauka

Azoto kiekio augaluose matavimai yra naudingi, siekiant optimizuoti augalų tręšimą azotu. Tyrimo tikslas – ištirti spektroskopijos panaudojimo tikslingumą nustatant azoto kiekį salotų lapuose. Lapų atspindžio spektrai buvo matuoti spektroradiometru 350–2500 nm bangų intervalu. Atrinkta 160 šviežių salotų lapų iš 5 skirtingo tręšimo azotu variantų, nuskenuotas jų spektras ir nustatytas suminio azoto kiekis. Parenkant optimalius intervalus arba kintamuosius buvo naudoti daliniai mažiausių kvadratų intervalai, daliniai mažiausių kvadratų sinerginiai intervalai ir daliniai mažiausių kvadratų sinerginiai intervalai kartu su genetiniais algoritmais. Kalibraciniai modeliai buvo sukurti, taikant dalinius mažiausių kvadratų ir ekstremalaus mokymosi algoritmo metodus. Naudojant 120 kintamųjų, vienas būdingas bangų ilgis (470–590 nm) buvo parinktas pagal dalinius mažiausių kvadratų intervalus, naudojant 362 kintamuosius, keturi būdingi bangų ilgiai (440–530, 530–620, 620–710 ir 890–980 nm) buvo parinkti pagal dalinius mažiausių kvadratų sinerginius intervalus, o 56 bangų ilgio kintamieji buvo parinkti pagal dalinius mažiausių kvadratų sinerginius intervalus kartu su genetiniais algoritmais. Azoto kiekiui nustatyti buvo parinkti šeši skirtingi regresijos modeliai, taikant dalinius mažiausių kvadratų ir ekstremalaus mokymosi algoritmo metodus, remiantis optimaliais intervalais arba kintamaisiais. Tyrimo rezultatai leidžia teigti, kad daliniai mažiausių kvadratų sinerginiai intervalai kartu su genetiniais algoritmais yra tinkamas kintamųjų parinkimo algoritmas, o ekstremalaus mokymosi algoritmas – nelinijinės regresijos metodas. Be to, abu šie metodai gali būti taikomi azoto kiekiui salotų lapuose matuoti, nes gaunami tikslesni rezultatai nei taikant kitus metodus. Rezultatų palyginimas parodė vidutinę kvadratinės prognozės paklaidą (RMSEP) = 0,2890 % ir koreliacijos koeficientą *Rp* = 0,9218.

Reikšminiai žodžiai: azotas, bangos ilgio parinkimas, daliniai mažiausi kvadratai, ekstremalus mokymosi algoritmas, spektroskopija.