Original Research Article

JNIRS

Journal of Near Infrared Spectroscopy
2023, Vol. 0(0) 1–8

DOI: 10.1177/09670335221148594
journals.sagepub.com/home/jns
SAGE

Analysis of alkaloids and reducing sugars in processed and unprocessed tobacco leaves using a handheld near infrared spectrometer

Miguel S Castillo[1], Juan J Acosta[2], Gary R Hodge[2], Matthew C Vann[1] and Ramsey S Lewis[1]

Abstract

Near infrared (NIR) spectroscopy calibration models were developed to predict chemical properties of flue-cured tobacco (*Nicotiana tabacum* L.) leaf samples using a microPHAZIR™ handheld NIR spectrometer. The sample data set consisted of 348 leaf-bundled samples of upper-stalk flue-cured tobacco leaves collected from an array of cultivars evaluated in multiple locations. Unprocessed leaf samples were intact whole unground leaves collected from curing barns. Processed leaf samples were further dried and ground before scanning. The NIR prediction models for percent reducing sugars, percent total alkaloids, and percent nicotine were very good for processed leaves [r^2 (SE_p in %) values = 0.98 (0.82), 0.92 (0.17), and 0.92 (0.14), respectively]. The models for the same three variables for unprocessed leaves were also very good, with only slightly lower fit statistics [r^2 (SE_p) = 0.93 (1.58), 0.87 (0.22), and 0.88 (0.18), respectively). Fit statistics for anabasine NIR models were intermediate with r^2 (SE_p in %) values ranging from 0.73 (0.003) to 0.76 (0.003), while the lowest fit statistics were observed for anatabine and norticotine with r^2 (SE_p in %) ranging from 0.49 (0.005) to 0.55 (0.017), respectively, for both unprocessed and processed leaves. Hence, use of a handheld NIR spectrometer would be of more limited value for these variables. The chemical composition of flue-cured tobacco leaf samples for some chemical traits can be directly assessed at the point when the leaves exit the curing barns, thus minimizing the need to dry and grind samples for colorimetric and chromatographic analyses.

Keywords

Near infrared spectroscopy, handheld, alkaloids, tobacco, reducing sugars, unprocessed, intact

Received 27 June 2022; accepted 10 October 2022

Introduction

Low-cost, rapid, and reliable techniques for determination of chemical constituents of agricultural commodities are sought-after tools for use in crop plant research and at the end-user interface for selection of products.[1–4] The use of near infrared (NIR) spectrometers has potential to meet those needs in agricultural science.[5] The foundational relationship between NIR spectroscopy and biochemistry is that near infrared energy incident on an object is either absorbed or reflected based on the objects molecular composition.[6]

In tobacco research, NIR spectroscopy has proven useful for qualitative discrimination of intact tobacco leaves to complement manual grading and classification approaches,[7] for determination of filling capacity,[8] and for quantification of nicotine, nitrogen, reducing sugar and total sugar concentrations.[9–12] Nicotine is the predominant tobacco alkaloid in most commercial genotypes, typically constituting more than 90% of the total alkaloid pool. Anabasine, nornicotine, and anatabine are additional alkaloids that are usually present at much lower concentrations.[13] These tobacco alkaloids are identified as precursors of tobacco-specific nitrosamines, some of which are recognized potent carcinogens found in tobacco products.[14]

Previous literature for NIR spectroscopic applications in tobacco research has mostly described analyses of dried and ground leaf material scanned with benchtop-type NIR spectrometers. Recent technological advances have led to the commercial availability of a variety of portable spectroscopy devices[15] that are more affordable compared to traditional benchtop-type devices and that could potentially be of value to farmers, consultants, or researchers for in situ and non-destructive analysis of samples.[1,16] In addition, optimum application of NIR spectroscopy could result in resource-savings (i.e., drying, grinding, reagents, equipment) compared to the effort required to conduct wet chemistry analyses in the laboratory. Near infrared

[1]Crop and Soil Sciences Department, College of Agriculture and Life Sciences, North Carolina State University, Raleigh, NC, USA
[2]Camcore, Department of Forestry & Environmental Resources, College of Natural Resources, North Carolina State University, Raleigh, NC, USA

Corresponding author:
Miguel S Castillo, Crop and Soil Sciences Department, College of Agriculture and Life Sciences, North Carolina State University, 2413 Williams Hall, Raleigh, NC 27695-7001, USA.
Email: mscastil@ncsu.edu

spectroscopy has also been reported to outperform genomics for predicting feedstock quality traits.[4] The specific objectives of this study were to (1) develop and evaluate NIR spectroscopic models using reflectance spectra acquired from unprocessed (intact) and processed (dried and ground) flue-cured tobacco leaves with a handheld NIR spectrometer to predict percent reducing sugars, total alkaloids, nicotine, anabasine, anatabine, and nornicotine and (2) compare predictions between unprocessed and processed leaves.

Materials and methods

Database description, chemical analyses, and NIR spectroscopy

Plant material consisted of bundles (n = 348) of upper-stalk flue-cured tobacco leaves (*Nicotiana tabacum* L.). These samples were part of the North Carolina Official Flue-Cured Tobacco Variety Testing Program and were collected from field evaluations of 42 commercial tobacco cultivars carried out during 2018 at three North Carolina locations (Kingston, K; Rocky Mount, RM, and Oxford, O).

The six response variables of interest were percent anabasine, anatabine, nicotine, nornicotine, reducing sugars, and total alkaloids. The average moisture concentration of flue-cured leaves was ≤14%. For laboratory analyses, the flue-cured-leaf samples were further dried to 60 °C until constant weight, using an air-forced drier and then ground to ≤ 1-mm. Quantification of percent reducing sugars and tobacco alkaloids was performed according to published colorimetric and gas chromatographic analyses methods.[17] The concentration of reducing sugars ranged from 4.36 to 25.39%, 1.09–4.13% for total alkaloids, 0.86–3.32% for nicotine, 0.004–0.028% for anabasine, 0.017–0.167% for anatabine, and 0.004–0.0465% for nornicotine (Figure 1). Additional descriptive statistics of the laboratory results are presented in Figure 1.

A handheld microPHAZIR™ (Thermo Fisher Scientific, Waltham, MA, USA) NIR spectrometer was used to acquires spectral data from unprocessed and processed cured tobacco leaves (*Nicotiana tabacum* L.). The microPHAZIR™ spectrometer acquires spectral data from 1596 to 2396 nm at 8-nm intervals for a total of 100 data points per sample. Although this is lower precision than many laboratory-grade spectrometers, the microPHAZIR™ device produced comparable results in predicting nutritive value of forages when compared to a benchtop-type FOSS 6500 device.[18]

Collection of spectra and database creation

Spectra were acquired from unprocessed (intact) and processed (dried and ground) flue-cured tobacco leaves. Both samples were brought into the laboratory (room temperature and light-controlled environment) for acquisition of

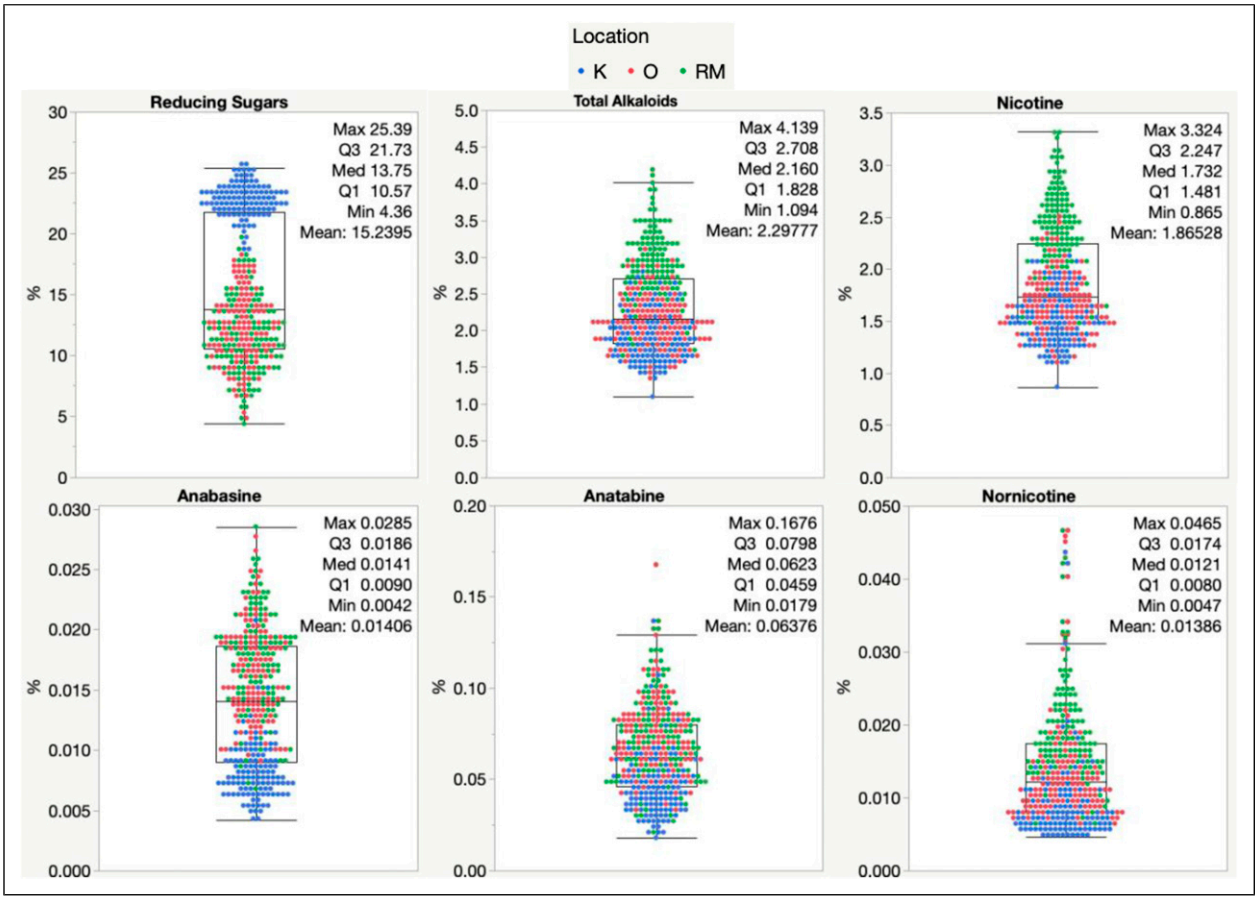

Figure 1. Descriptive statistics of laboratory measurements of flue-cured tobacco leaves. The letters K, O, and RM correspond to the three geographical locations of Kingston, Oxford, and Rocky Mount, respectively.

spectra. For unprocessed leaves, 3 leaves were stacked, ensuring that leaves were in the same direction, and two scans from the upper side (adaxial) and two from the lower side (abaxial) were taken in diagonal direction on each leaf for a total 4 scans per leaf (Figure 2). The handheld device was in direct contact with the tobacco leaves. A total of 12 spectral signatures were collected for the three-leaf bundle (2 per leaf x 2 sides x 3 leaves). Custom scripts in R (environment version 3.3.2)[19] were developed to calculate the mean reflectance spectrum for each three-leaf bundle.

For the processed leaves (dried and ground), the handheld device was not in direct contact with the ground sample; instead, scans were taken through the cover glass of FOSS 6500 module sampling cups containing the samples. The samples were scanned in four positions by rotating the sample cups 90° and taking measurements while in static position. This approach previously proved successful when comparing results from this handheld device to a benchtop-type NIR spectrometer.[18] Custom scripts were used to calculate the mean reflectance spectrum for each sample. Average spectra for the processed and unprocessed leaves are presented in Figure 3.

Figure 2. Unprocessed (intact) flue-cured tobacco leaf, microPHAZIR™ handheld NIR, and representation of the two sampling points (black dots on leaf).

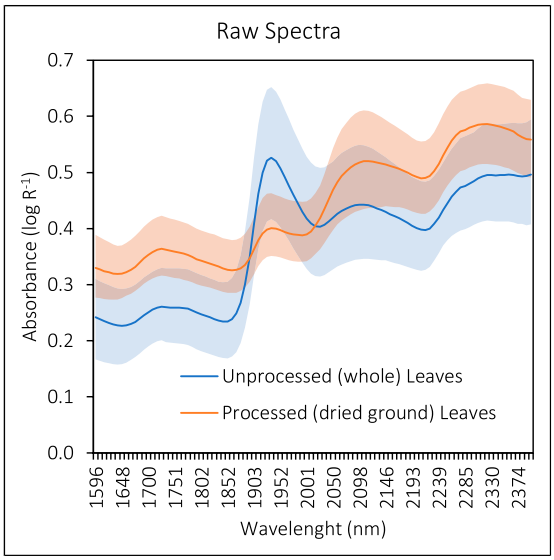

Figure 3. Raw NIR spectra measured using the microPHAZIR™ handheld NIRS device. The plotted spectra are mean values of all scanned samples and corresponding maximum and minimum values at measured wavelengths.

NIR calibration development

Calibration development was performed using a data analysis pipeline written for the R environment. The pipeline was previously used in the successful development of NIR models for several crops including trees and forages[20,21] and to compare the prediction performance of handheld- and benchtop-type NIR spectrometers in forage samples.[18] The pipeline has two separate phases: (a) transformations and outlier detections, and (b) model training, validation, cross-validation, and prediction of new observations.

In summary, a total of 13 mathematical transformations of the spectra were first applied to raw spectra ($\log R^{-1}$) to remove scattering of diffuse reflections associated with sample particle size and to improve subsequent regression analyses. Scatter-correction methods included multiplicative scatter correction (MSC), standard normal variate (SNV) and detrend (DT). Spectral derivative methods included Savitzky-Golay second derivative (second order polynomial fit) with two different window sizes of 5 and 7 points (SG5 and SG7). In addition, several pairs of transformations were also investigated (SNV + DT, MSC + DT, SNV + SG, MSC + SG, and DT + SG). Local outlier factors (LOF) were used to filter-out atypical spectral data.[22] In this analysis a LOF score for each observation was calculated based on its 20 nearest neighbors. Samples with LOF values greater than 2 were excluded from the analysis.

The second phase of the pipeline uses the outlier-free datasets to fit NIR calibrations between spectral data and wet chemistry laboratory values using partial least squares regressions (PLS). Model performance was evaluated using leave-one-out (LOO) cross validation. Desirable models were considered those that maximized the cross-validation coefficient of determination (r^2_{CV}) estimated using the one-sigma approach.[23] The one-sigma approach is a method to reduce risk of over-fitting the model by reducing the number of factors; the algorithm examines models using from 1 to 20 factors and identifies the model with the lowest number of factors where standard error of cross validation (SECV) is within one standard error of the minimum SECV.

NIR calibration comparison

Performance evaluation of the predictions was carried out using a test set. Outlier-free datasets of the cured and processed leaves were randomly split in two sets, a training set and a test set. The training set consisted of 75% of the observations and was used for developing calibration models. The test set, with the remaining 25% of

observations, was used for performance evaluation of predictions. Both sets, the calibration and test sets, had the same distribution (75:25 calibration: test set) of samples based on location. The number of samples contained in the outlier-free datasets is presented in Table 1. The predictive power of the PLS calibrations was assessed by plotting scatterplots of the wet chemistry laboratory values on the *x*-axis and NIR predicted values on the *y*-axis (Figure 4). In addition, the coefficient of determination (r^2), standard error of prediction (SE_p), ratio of performance to deviation (RPD), bias, intercept, and slope (Table 2) were determined.

Results and discussion

The data analysis pipeline generated summary tables with fit statistics and figures for all the possible PLS regression calibrations for each response variable. An example table with 14 models generated to predict nicotine from processed leaf samples in the calibration data set is presented in Table 3. In addition, Figure 5 illustrates an example output of SECV as a function of the number of factors (latent variables) for the selected SG7 model in Table 1 using the one-sigma approach.

Prediction performance of the PLS regression models was assessed using the test sets (Table 2; Figure 4). For unprocessed leaves, r^2 (SE_p in %) values were 0.93 (1.58) for percent reducing sugars, 0.88 (0.18) for percent total alkaloids, 0.87 (0.21) for percent nicotine, 0.73 (0.003) for percent anabasine, 0.55 (0.017) for percent anatabine, and 0.55 (0.005) for percent nornicotine. For reducing sugars, total alkaloids, and nicotine, fit statistics for the PLS regression calibration models were only slightly better for processed leaves compared to unprocessed leaves, whereas fit statistics for anabasine, anatabine, and nornicotine were essentially the same for processed and unprocessed leaves. The r^2 (SE_p in %) values for processed leaves were 0.98 (0.82) for percent reducing sugars, and 0.92 (0.17) for percent total alkaloids, 0.92 (0.14) for percent nicotine. In addition to r^2 and SE_p values, Table 2 also provides RPD, bias, intercept, and slope values. Desirable models are those with intercept and bias values closer to zero and slope values closer to 1. The RPD values ranged from 1.5 to 3.7 for unprocessed leaves and from 1.5 to 7.1 for processed leaves. Models with RPD values >2 have been suggested as models with good prediction ability;[24] however, different and higher threshold values for RPD have also been suggested in the literature.[25] Reporting RPD values may be redundant when r^2 values are already reported for normally distributed variables and large datasets.[26] The RPD values for reducing sugars, total alkaloids, and nicotine were consistently ≥2.8 for PLS regression models for both processed and unprocessed leaf models.

The average raw spectrum absorbance was lower for unprocessed leaves but followed a similar trend in terms of valleys and peaks compared to the processed leaves (Figure 3). An exception occurred in the region between 1894 and 2026 nm, where the spectra for unprocessed leaves exhibited greater absorbance. Five prominent water absorption bands (regions) in the near infrared spectrum are 760, 970, 1190, 1450, and 1940 nm.[26] Higher absorption for the unprocessed leaves in the 1894–2026 nm regions can possibly be attributed to higher moisture content of unprocessed leaf samples.

It is clear for all PLS regression calibrations, and especially those with r^2 values ≥0.88 corresponding to percent reducing sugars, total alkaloids, and nicotine, that NIR-predicted values are very close to the laboratory measured values (Figure 4), notwithstanding data segregation of the response variables by geographical location upon visual inspection (Figure 1). There were no apparent indications of heteroskedasticity and over- or under-estimation. Dispersion of points around reference lines (linear model and line with slope 1) were greater for percent anabasine, anatabine, and

Table 1. Number of samples identified as outliers and number of sample s in the resulting outlier-free databases used for development of PLS regression calibrations for anabasine, anatabine, nicotine, nornicotine, total alkaloids, and reducing sugar of unprocessed (whole) and processed (dried ground) flue-cured tobacco leaves.

Leaf Sample	Model[a]	Variable	# outliers	# samples in outlier-free calibration database	# samples in outlier-free Validation database
Unprocessed (intact) Leaves	DT_SG5 (5)	Reducing Sugars	0	261	87
	SG5 (5)	Total Alkaloids	0	261	87
	SG5 (5)	Nicotine	0	261	87
	NIR (8)	Anabasine	4	259	85
	SG5 (4)	Anatabine	0	261	87
	SG5 (2)	Nornicotine	0	261	87
Processed (ground dried) Leaves	MSC (7)	Reducing Sugars	2	258	87
	SG7 (7)	Total Alkaloids	0	260	87
	SG7 (7)	Nicotine	0	260	87
	SNV (4)	Anabasine	1	259	87
	SNV_SG5 (4)	Anatabine	0	260	87
	DT (2)	Nornicotine	2	258	87

[a]NIR, raw spectra (log R^{-1}); SG5, Savitzky-Golay 2nd derivative spectra using five points; DT, detrend; SNV, standard normal variate; MSC, multiplicative scatter correction; SG7, Savitzky-Golay 2nd derivative spectra using seven points. Numbers within parenthesis represent the number of loading factors (latent variables) in the partial least squares regression models. Fit statistics for the models are presented in Table 2.

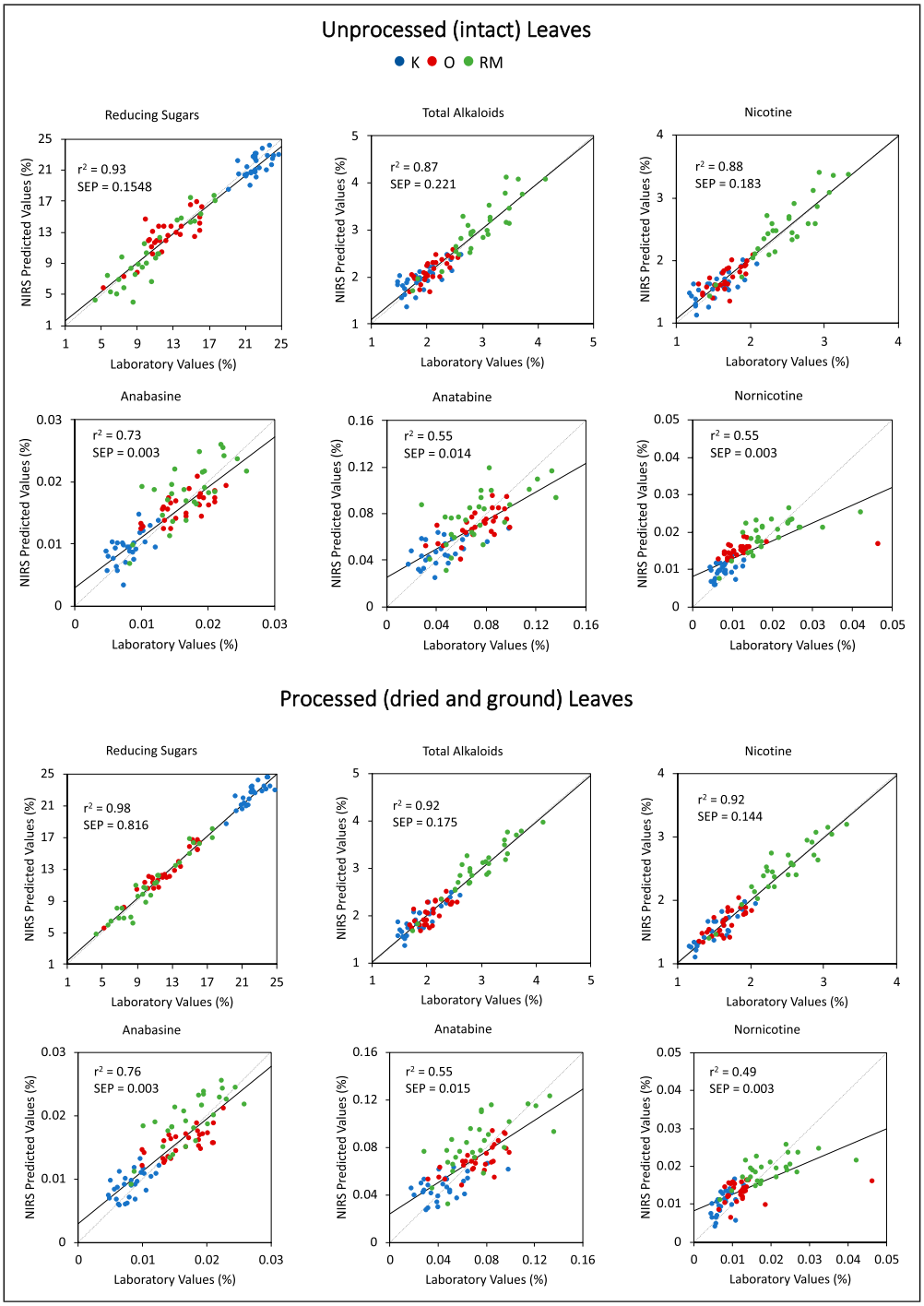

Figure 4. Validation scatterplots of the test-set samples for NIR-predicted vs reference values for six chemical traits of tobacco. Dotted line has slope = 1 and solid line is the linear regression for validation models in Table 1. The letters K, O, and RM correspond to the three geographical locations of Kingston, Oxford, and Rocky Mount, respectively.

Table 2. Fit statistics for performance evaluation of predictions of PLS regression models developed using 75% of the observations and evaluated on the remaining 25% of the observations. The fit statistics correspond to the scatter plots in Figure 4.

Sample	Spectral transformation (# factors)[a]	Variable	Calibration set				Test set					
			R^2_C	r^2_{CV}	SEC	SECV	r^2	SEP	RPD	Bias	Intercept	Slope
								%		%		
Unprocessed (intact) leaves	DT_SG5 (5)	Reducing Sugars	0.94	0.93	1.3520	1.4640	0.93	1.548	3.7	−0.2773	0.623	0.9391
	SG5 (5)	Total Alkaloids	0.88	0.87	0.2042	0.2186	0.87	0.221	2.7	0.0537	0.124	0.9692
	SG5 (5)	Nicotine	0.90	0.88	0.1634	0.1753	0.88	0.183	2.8	0.0462	0.100	0.9707
	NIR (8)	Anabasine	0.78	0.75	0.0026	0.0028	0.73	0.003	1.9	0.0006	0.003	0.8110
	SG5 (4)	Anatabine	0.64	0.58	0.0149	0.0159	0.55	0.014	1.8	0.0008	0.025	0.6122
	SG5 (2)	Nornicotine	0.33	0.30	0.0066	0.0068	0.55	0.003	2.5	0.0015	0.008	0.4808
Processed (dried and ground) leaves	MSC (7)	Reducing Sugars	0.98	0.97	0.8325	0.9102	0.98	0.816	7.1	−0.1452	0.4016	0.9827
	SG7 (7)	Total Alkaloids	0.95	0.94	0.1367	0.1517	0.92	0.175	3.4	−0.0020	0.0426	0.9824
	SG7 (7)	Nicotine	0.96	0.95	0.1056	0.1169	0.92	0.144	3.5	0.0260	0.0325	0.9840
	SNV (4)	Anabasine	0.80	0.78	0.0025	0.0026	0.76	0.003	2.0	0.0006	0.0029	0.8306
	SNV_SG5 (4)	Anatabine	0.68	0.62	0.0141	0.0153	0.55	0.015	1.7	0.0019	0.0242	0.6547
	DT (2)	Nornicotine	0.35	0.33	0.0066	0.0066	0.49	0.003	2.5	0.0007	0.0082	0.4325

Note. R^2_C, coefficient of determination, calibration; r^2_{CV}: coefficient of determination, cross-validation; SEC, standard error of calibration; SECV, standard error of cross-validation; r^2, coefficient of determination, prediction; SEP, standard error of prediction; RPD, ratio of performance to deviation (SD/SEP; SD = standard deviation of reference samples in validation test-set); Intercept, intercept of regression line in Figure 4; Slope, slope of the regression line in Figure 4.

[a]NIR, raw spectra (log R^{-1}; R = reflectance); SG7, Savitzky-Golay 2nd derivative spectra using five points; DT, detrend; SNV, standard normal variate; SG7, Savitzky-Golay 2nd derivative spectra using seven points; DT, detrend; numbers within parenthesis represent the number of loading factors (latent variables) in the partial least squares regression models.

Table 3. Example summary table of fit statistics of several PLS regression models developed to predict nicotine in processed (dried and ground) flue-cured tobacco leaves. The model SG7 was selected as the best model and further evaluated using the test set and results are presented in Table 1 and Figure 3.

Database	Factors	R^2_C	SEC	r^2_{CV}	SECV
SNV	9	0.95	0.1106	0.94	0.1227
MSC	9	0.95	0.1108	0.94	0.1240
DT	9	0.96	0.1053	0.94	0.1218
SG5	8	0.96	0.1004	0.94	0.1196
SG7*	7	0.96	0.1056	0.95	0.1169
SNV_DT	9	0.96	0.1053	0.94	0.1216
MSC_DT	8	0.95	0.1104	0.94	0.1208
SNV_SG5	8	0.96	0.0979	0.95	0.1180
SNV_SG7	6	0.95	0.1103	0.94	0.1202
MSC_SG5	8	0.96	0.0981	0.94	0.1190
MSC_SG7	6	0.95	0.1106	0.94	0.1232
DT_SG5	7	0.96	0.1035	0.94	0.1204
DT_SG7	7	0.96	0.1028	0.95	0.1165
NIR	13	0.96	0.1022	0.94	0.1239

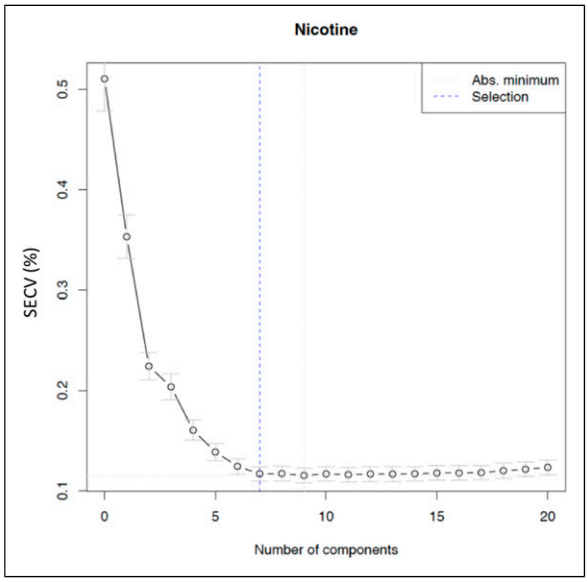

Figure 5. Example fit statistics of the one-sigma method for the selecting the number of model components. This example corresponds to the SG7 model selected in Table 3. The minimum SECV was found for the model with 10 factors, the selected model with seven factors has fewer factors and SECV not significantly different than the 10-factor model.

nornicotine. Anabasine models had intermediate r^2 values at 0.76 while for anatabine and nornicotine the values ranged between 0.49 to 0.55 (Figure 4; Table 2). These results report on the utility of NIR spectroscopy for determination of alkaloids and reducing sugars in tobacco, and most notably the utility of a handheld device to analyze unprocessed leaf samples. These findings can potentially represent significant savings in resources used for sample preparation and analysis (e.g., drying, grinding, labor) of traditionally wet chemistry analyses.

Conclusions and implications

The PLS regression models for percent reducing sugars, percent total alkaloids, and percent nicotine developed using NIR spectra were consistently better for processed tobacco leaves. Nonetheless, the models for unprocessed leaves had only slightly lower fit statistics, thus would be nearly as accurate and precise as models with processed leaves. Since acquisition of NIR spectra on intact leaves is faster and easier (requiring no additional time or expense of drying and grinding leaves), larger numbers of samples could be taken with unprocessed leaves to better characterize a particular sample set. Use of a handheld NIR spectrometer was found to provide great utility for prediction of levels of reducing sugars, total alkaloids, and nicotine concentrations for processed and unprocessed flue-cured tobacco leaves. Fit statistics for anabasine PLS regression models were intermediate. The lowest fit statistics were for anatabine and norticotine, suggesting that the use of a handheld NIR spectrometer device would be less useful for these alkaloids. These results demonstrate the utility of the handheld device to accurately predict concentrations of reducing sugars, total alkaloids, nicotine, and anatabine in tobacco leaf samples and potentially speed up the screening and analysis process by directly assessing the concentration of the chemical constituents anytime after the point when the leaves exit the curing barns.

Acknowledgements

The authors would like to thank Stephanie Sosinski, Research Technician of the Forage and Grassland Program at NC State University, and William Woodbridge, Camcore's database administrator, for their excellent technical support.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Miguel S. Castillo https://orcid.org/0000-0003-1066-5906

References

1. Pérez-Marín D, Paz P, Guerrero J-E, et al. Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. *J Food Eng* 2010; 99: 294–302.
2. Acharya UK, Subedi PP and Walsh KB. Robustness of tomato quality evaluation using a portable Vis-SWNIRS for dry matter and colour. *Int J Anal Chem* 2017; 2017: 1–8.
3. Kirchler CG, Pezzei CK, Beć KB, et al. Critical evaluation of spectral information of benchtop vs. portable near-infrared spectrometers: quantum chemistry and two-dimensional correlation spectroscopy for a better understanding of PLS regression models of the rosmarinic acid content in Rosmarini folium. *Analyst* 2017; 142: 455–464.
4. Gonçalves MTV, Morota G, Costa PM deA, et al. Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PLoS One* 2021; 16: e0236853.
5. Batten GD. An appreciation of the contribution of NIR to agriculture. *J Near Infrared Spectrosc* 1998; 6: 105–114. Epub ahead of print 2 February 2017. DOI: 10.1255/jnirs.127
6. Shenk JS, Workman JJ Jr and Westerhaus MO. Application of NIR Spectroscopy to Agricultural Products. In: *Handbook of Near-Infrared Analysis*. Boca Raton: CRC Press, 2007.
7. Lu M, Zhou Q, Chen T, et al. Qualitative discrimination of intact tobacco leaves based on Near-Infrared technology. *J Spectrosc* 2021; 2021: 1–9.
8. Watanabe N, Katsuyama N and Yuasa H. *Measurement of the filling capacity of tobacco using near-infrared spectroscopy analysis*. CORESTA. https://www.coresta.org/abstracts/measurement-filling-capacity-tobacco-using-near-infrared-spectroscopy-analysis-28902.html (2012, accessed 3 May 2022).
9. Ma B and Ji H. Particle swarm optimization algorithm establish the model of tobacco ingredients in near infrared spectroscopy quantitative analysis. In: D Li and Y Chen (eds) *Computer and computing technologies in agriculture VI*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 92–98.
10. Zhang Y, Cong Q, Xie Y, et al. Quantitative analysis of routine chemical constituents in tobacco by near-infrared spectroscopy and support vector machine. *Spectro Act Part A: Molec Biomole Spectrosc* 2008; 71: 1408–1413.
11. Tan C, Wang J, Wu T, et al. Determination of nicotine in tobacco samples by near-infrared spectroscopy and boosting partial least squares. *Vibr Spectrosc* 2010; 54: 35–41.
12. Borges Miranda A, Pérez Martínez C, Jiménez Chacón J, et al. Near infrared spectroscopic analysis of total alkaloids as nicotine, total nitrogen and total ash in Cuban cigar tobacco. *J Near Infrared Spectrosc* 2019; 27: 123–133.
13. Sisson V and Severson R. Alkaloid composition of the nicotiana species. *Contr Tobacco & Nicotine Res* 1990; 14: 327–339.
14. Hecht SS. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem Res Toxicol* 1998; 11: 559–603.
15. Chenxi Z, Xiaping F, Jianyi Z, et al. Review of portable near infrared spectrometers: Current status and new techniques. *J Near Infrared Spectrosc* 2022; 30(2): 51–66.
16. Warburton P, Brawner J and Meder R. Technical note: handheld near infared spectroscopy for the prediction of leaf physiological status in tree seedlings. *J Near Infrared Spectrosc* 2014; 22: 433–438.
17. Lewis RS, Lopez HO, Bowen SW, et al. Transgenic and mutation-based suppression of a Berberine Bridge Enzyme-Like (BBL) gene family reduces alkaloid content in field-grown tobacco. *Plos One* 2015; 10: e0117273.
18. Acosta JJ, Castillo MS and Hodge GR. Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value. *Crop Sci* 2020; 60: 3410–3422.

19. Hodge GR, Acosta JJ, Unda F, et al. Global near infrared spectroscopy models to predict wood chemical properties of Eucalyptus. *J Near Infrared Spectrosc* 2018; 26: 117–132.
20. Castillo MS, Tiezzi F and Franzluebbers AJ. Tree species effects on understory forage productivity and microclimate in a silvopasture of the Southeastern USA. *Agric, Ecosys & Envir* 2020; 295: 106917.
21. Breunig MM, Kriegel H-P, Ng RT, et al. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, New York, NY, USA, 2000. Association for Computing Machinery, pp. 93–104.
22. Hastie T, Tibshirani R and Friedman J. *The elements of statistical learning*. New York, NY: Springer, 2009. Epub ahead of print. DOI: 10.1007/978-0-387-84858-7
23. Chang C-W, Laird DA, Mausbach MJ, et al. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci Soc Am J* 2001; 65: 480–490.
24. Williams P. The RPD Statistic: a tutorial note. *NIR News* 2014; 25: 22–26.
25. Minasny B and McBratney A. Why you don't need to use RPD. *Pedometron* August 2013; 33: 14.
26. Curcio JA and Petty CC. The near infrared absorption spectrum of liquid water. *J Opt Soc Am* 1951; 41: 302.