

Contents lists available at ScienceDirect

Smart Agricultural Technology



journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Calibrating lab and field reflectance spectra for nutrient estimation in potato plants using local support vector regression models

Reem Abukmeil^a, Ahmad Al-Mallahi^{a,*}, Felipe Campelo^b

^a Department of Engineering, Faculty of Agriculture, Dalhousie University, Nova Scotia, Canada
^b Department of Computer Science, Aston University, Birmingham, UK

ARTICLE INFO

ABSTRACT

Keywords: Leaf reflectance Vis/NIR spectra Neighbor-based variable selection Support vector regression NPK estimation

This study presents a methodology based on multiple local support vector regression (SVR) to calibrate the spectra taken in the field in relative to lab-derived spectra. Laboratory based foliar spectral measurement is a common method to provide lab-derived spectra as a service where a grower sends sample leaves collected manually. The drawback of this method is being time-consuming when the samples are collected and analyzed. In contrast, in-field spectral measurements can be an alternative method capable of providing immediate readings. While both methods work based on the same principle, the insturmental differences as well as the conditional difference under which the instruments operate may cause differences in the spectral patterns of the same target. In this work, after developing the calibration method, we validated it by estimating NPK measurements in potato plants using in-field, lab, and field calibrated spectral measurements over two testing modes: dried and fresh. The results showed that the calibration using SVR models could minimize the percentage relative error (PRE) between lab and field spectra within the visible range by considering the influence of the neighboring wavebands up to 32 nm width which improved the alignment of the local maxima of the specral curves. Also, a substantial PRE reduction from 120 % to 20 % for some wavebands in the short-wave infrared (SWIR) region of the fresh mode was observed due to the influence of scaling within the SVR method. The calibration improved the alignment of NPK estimated values between lab and field calibrated spectra of both modes with an emphasis on its necessity to estimate nutrients in the fresh mode as the root mean square error was < 0.1 for the three elements.

1. Introduction

Potato nutrient use efficiency is key to plant growth, and for that purpose, potato growers undertake multiple assessment tools to assess nutrient deficiencies. Those tools include visual diagnosis, plant tissue and soil tests, and cropping history [5]. While tissue tests are acknowledged to be the most representative [12], they are considered destructive, laborious, time-consuming, and expensive [23]. These drawbacks of tissue testing have led growers to seek other nutrient assessment techniques.

The introduction of near-infrared spectroscopy (NIRS) for plant nutrient assessment is one technique that shows the potential to estimate crop nutrient status as a non-destructive, and more accessible tool if the readings are service-paid [14]. According to previous studies, significant spectral bands have been identified in forestry and crop applications within the visible and very near-infrared range (Vis-VNIR, 400 - 1100 nm) as well as the short-wave infrared range (SWIR, 1100 - 3000 nm) [19].

While lab spectral measurement is the common method for spectral analysis, its sample preparation is still destructive which means that significant time is needed to reach to, collect, and arrange the leaves. In addition, there is a gap between sample collection and analysis that may require special handling. For instance, at least 20 leaves are needed to fill the required size of a sample cup of a lab NIRS analyzer (FOSS DS2500, Hilleroed, Denmark) to ensure proper sample filling and minimize any potential noise or variability during the analysis process. Moreover, the credibility of lab spectral measurements immensely depends on the time gap between sample collection and testing [9]. In contrast, in-field spectral measurements can be a non-destructive method capable of giving spectral readings in a matter of seconds, enabling more measurements to be taken on the spot and eliminating the gap between sample collection and testing. Economic considerations can

* Corresponding author. E-mail address: ahmad.almallahi@dal.ca (A. Al-Mallahi).

https://doi.org/10.1016/j.atech.2024.100492

Received 23 March 2024; Received in revised form 7 June 2024; Accepted 16 June 2024 Available online 20 June 2024

^{2772-3755/© 2024} The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Table 1

Operating specifications of the two spectral instruments.

Item	Specifications				
	NIRS analyzer	Vis-NIR spectrophotometer			
Brand	FOSS DS2500	ASD FieldSpec 4			
Measurement Mode	Reflectance	Reflectance			
Wavelength Range	400 - 2500	350 – 2500 nm			
	nm				
Spectral resolution	0.5 nm	1 nm			
Wavelength Accuracy	\pm 0.05 nm	\pm 0.5 nm			
Dried leaves presentation to instrument	Samples cup	Sample turntable			
Fresh leaves presentation to instrument	Samples cup	Leaf clip			
Weight	27 kg	5.44 kg			
Mechanical environment	Stationary	Portable (leaf clip), Stationary (turntable)			
Voltage supply	100–240 V AC	90-240 VAC or external battery			
Format of outputs	Nir format	Export to ASCII text files is possible			

also play a role in the choice of measurement method. Laboratory measurements may be accessible by growers as a fee-for-service, whereas in the absence of similar in-field service, the farmers would need to buy or rent the equipment to perform instantaneous measurements.

Although both instruments are based on the same technology, each spectrophotometer may have different spectral characteristics or responses due to variances in their electronic components, detectors, accessories, lighting conditions, etc. Also, the different environmental conditions under which the instruments are used may lead to different measurements. Given the recent success of machine learning algorithms, detecting patterns in spectral data, and modeling their relationships with the reference concentrations is possible. However, as emphasized by Krinitsky et al. [10], calibration needs to be developed to minimize biases and noise that can emerge from using one model to estimate nutrients using two different instruments under different environmental conditions.

The Empirical Line Method (ELM) is the most widely used method to calibrate two sets of remote sensing data and surface reflectance assuming a linear relationship. A study by Xu et al. [21] used the ELM to calibrate multi-spectral images taken by unmanned aerial vehicles, introducing the constraint condition of spectral angle and the full use of the information of each band. Pompilio et al. [15] applied the ELM to calibrate the airborne Daedalus-CZCS scanner for compatibility with the ASD FieldSpec spectroradiometer of a simultaneous field survey to achieve absolute atmospheric correction and geometric calibration. This method was also assessed to measure reference targets of known reflectance to obtain reliable quantitative measurements from the remote sensors [13].

Alternatively, non-linear methods appeared to provide more accurate models for the cases of full spectrum calibration. For instance, Arai [2] and Coliban et al. [4] assessed the calibration between remote sensing data and surface reflectance based on a nonlinear mixing model and artificial neural networks. Support Vector Regression (SVR) is also widely used as a robust nonlinear method for this sort of calibration. SVR is a machine learning technique with a quantitative response by introducing an error measure as a margin of tolerance (ε -insensitive) to ignore errors of size less than ε . This makes the fitting less sensitive to outliers and focuses on significant deviations [7].

A previous study used an SVR model with a Gaussian radial basis kernel as a calibration model to predict sugar content in grape berries using spectral imaging measurements [20]. Their study reported good performance of the SVR over unseen data of different varieties. Another study mentioned the suitability of SVR for the calibration of remote sensing data of small samples with an excellent generalization ability [18]. Their study also addressed the problem of selecting the model parameters for SVR, and thus they assessed the importance of the parameter selection process prior to the SVR in estimating chlorophyll concentration and suspended sediment concentration in water bodies. Another study also assessed the importance of variable selection prior to the implementation of SVR for proper calibration to predict soil nutrients using Vis-NIR spectra [6]. Xu and Wang [22] utilized local information among samples to enhance the prediction accuracy of protein in wheat using neighbor-based weighted regression and SVR.

In this work, we present a method to calibrate field spectral data taken from intact potato leaves for compatibility with lab spectral data taken from potato leaves slotted inside a lab spectral instrument. The proposed calibration methodology is based on fitting SVR models to calibrate every waveband separately, using the local neighborhood of the target waveband in the field data as predictor variables. This approach has the advantage of not assuming homogeneity of the calibration function across the whole spectrum, but rather learning local transformation functions that consider potential nonlinearities in the mapping functions. The resulting calibration model was assessed by estimating NPK using nutrient estimation models developed using lab spectral data on an independent subset of the spectrum of potato leaves taken from the field and lab.

2. Materials and methods

2.1. Reflectance spectra collection in the lab and in the field

The lab spectral measurements, which acted as reference spectra, were collected using the NIRS analyzer for both fresh and dried leaves. On the other hand, the field spectral measurements were collected using a Vis-NIR spectrophotometer (FieldSpec 4 Standard-Res., Malvern Panalytical Technologies, Worcestershire, UK). For fresh leaves, the spectrum was taken using a leaf clip, whereas the dried leaves were scanned on a turntable after grinding. The decision to utilize lab spectra as reference measurements in this study was based on previous research that demonstrated a correlation between potato petiole and leaf reflectance using the lab NIRS analyzer [1]. The operating specifications of the two instruments are presented in Table 1.

Sample collection included a total of 45 data points of the Russet Burbank potato variety. The data points were obtained from six subplots located across four potato farms in Florenceville-Bristol, NB, Canada. Six or nine data points were taken from each subplot depending on the initial date of data collection of the subplot. The sampling process commenced in early July, approximately 55 days after planting, and continued until mid-August 2021, under a rule that spaced data collection from the same subplot by a minimum of two weeks. Each data point was comprised of observations from 60 randomly selected leaves within the subplot. The 60 leaves were divided into two groups, Group 1 for lab testing contained 40 leaves separated into fresh and dried modes (20 leaves in each), and Group 2 contained 20 fresh leaves scanned fresh first, before drying them for the dried mode data collection.

The leaves of Group 1 were immediately vacuum-packed after peeling off the petioles, placed into sampling bags, and refrigerated before shipment. Each sample was packed with an ice bag and the shipment took 24 h until they arrived at the DairyLand Lab Inc. (Arcadia, Wisconsin, USA) for fresh and dried spectral analysis using the NIRS analyzer. The leaves were placed in a paper envelope before putting them in an oven at 55 - 60 °C for 16 - 24 h. The spectral observations of the dried and fresh leaves were taken in a black cup to reduce the impact of stray light. The leaves were trimmed symmetrically for all samples to fit the size of the cups.

The leaves of Group 2 were removed from the petioles for in-field spectral analysis using the leaf clip of the Vis-NIR spectrophotometer. The black background of the leaf clip was assessed to give the best modeling accuracy in comparison to the white background for field



Fig. 1. Taking in-situ spectral measurements. (a) In-field spectral instrument, (b) Leaf clip of the in-field spectral instrument with probe and light to read spectral data, (c) Computer device for showing and saving spectral data.

spectral measurement [23], which was used in all in-field spectral measurements. Fig. 1 shows an image of the in-situ spectral measurement where; (a) shows the Vis-NIR in-field spectrophotometer, (b) displays the collection of the spectral measurements using the leaf clip at the black background, and (c) shows the computer device where spectral measurements were shown and saved. Also, a white reference was taken every ten readings to ensure normalized in-field spectral measurements. Directly after measuring the spectrum of the fresh leaves, they were dried and ground down to a 1 mm powder prior to scanning on the turntable. The ground powder was placed in a glass petri dish (100 x 15 mm) (Pyrex, Corning Brand) which was positioned on a petri dial where a white reference panel (Spectralon panel, LabSphere Inc., North Sutton, USA) was placed over the rotating turntable. The thickness of the ground powder in the petri dish was around 3 cm to ensure complete diffuse reflectance which follows the recommendation given by Williams [25] for wheat and seeds of comparable size. The petioles of the removed leaves of both Groups 1 and 2 were sent for the chemical testing of N, P, and K following the Association of Official Analytical Chemists (AOAC) methods.

2.2. Dataset development

The data generated by the WinISI software of the NIRS analyzer are displayed as absorbance which was converted into reflectance values using the relationship of $10^{-absorbance}$ at 0.5 nm intervals resulting in 4200 readings. Rather than using the entire 4200 readings, readings were down-sampled to an 8 nm resolution (i.e., 8 nm resolution accounts for every 16 readings because the spectral resolution is 0.5 nm), resulting in a total of 262 readings with centers ranging from 404 to 2492 nm. Two datasets were developed representing the lab spectral measurements in both testing modes, fresh and dried leaves, hereafter called *lab datasets*. On the other hand, the data generated by the Indico

Pro software of the Vis-NIR spectrophotometer represent reflectance values every 1 nm from 350 nm to 2500 nm giving a total of 2151 readings. To mitigate the impact of noise, the spectral range of 350–399 nm was excluded due to their comparatively lower signal-to-noise ratio, and then readings were also down-sampled at 8 nm resolution resulting in a total of 262 readings matching the size of the lab datasets. The choice of 8 nm resolution was found fine enough to ensure that the down-sampled data retains significant information for analysis and processing to reduce computational complexity.

The in-field spectral measurements of the fresh leaves were averaged for 20 leaves for each data point. The in-field spectrum of the dried leaves was averaged every three readings. Two datasets were then developed representing the in-field spectral measurement for fresh and dried leaves, hereafter called *field datasets*.

2.3. Calibration method

By employing lab spectra as reference data, the lab spectra were considered the target variables (*y*), and the field spectra were taken as response variables (*x*). The spectral data were first split into six subsets based on the source plot of the leaves so that data from 5 plots were used exclusively for training and validation and 1 plot exclusively for testing the developed model. Validation was done using a grouped 5-fold cross validation used to obtain performance estimates as part of the model development and evaluation process before testing using the independent plot.

The methodology developed for mapping field measurements onto lab-compatible values was based on fitting individual nonlinear regression models for each target waveband. We first addressed the set of response variables by following Eq. (1):

$$Y = \{y_{ki}; i \in \{404, 412, ..., 2492\}, k = 1, ..., n_{samples}\},$$
(1)



Fig. 2. General structure of the proposed calibration modeling structure. (a) Matched spectra obtained from lab and field measurements. (b) Extracting a target variable (intensity at a specific waveband) and associated predictors (intensities at the local neighborhood of the target waveband). (c) Building of predictive models using the target and predictor variables values across all training data points. (d) Resulting SVR models whose implementation produces the calibrated spectrum shown in Figure (4).



Fig. 3. Flow of the calibration method towards developing calibrated field spectra involving the neighbor-based selection of variables followed by SVR.

where y_{ki} is the reference reflectance (i.e., obtained from lab measurement) value at waveband *i* for sample *k*. Correspondingly to each y_{ki} , we extracted a set X_{ki} , based on the predictor (i.e., raw field measurement) reflectance values and their squared values, according to Eq. (2):

$$X_{ki} = \left\{ x_{k(i+j)}, \ x_{k(i+j)}^2 \right\}; j \in \{0, \ \pm 8, \pm 16\}; k = 1, \ \dots, \ n_{samples},$$
(2)

subject to $400 \le (i + j) \le 2500$. This resulted in a maximum of 10 predictive features - the reflectance at the matching waveband (x_{ki}) plus up to two lower and two higher wavebands and their squared values, truncated at both ends of the measurement range. The inclusion of

neighbor variables in addition to the central waveband is aimed at capturing the local influence of the adjacent variables for possible relevant valuable information carried by the neighbor variables. This can be particularly helpful where the spectral data exhibits local variations. The addition of squared values allows to model some degree of nonlinearity in the relationship between the regressor and target wavebands, and the local scope not only helps in constraining the dimensionality of the feature space but also focuses the regression modeling on the regions in the spectrum that most likely carrying relevant information to learn the mapping function of each specific waveband. All feature isolation and data splitting were performed using package *dplyr* (Wickham et al., [24]) of the R statistical language and



(b)

Fig. 4. Lab, field (raw), and field (calibrated) reflectance measurements of the primary axis for a representative sample of the holdout split data in the (a) dried and (b) fresh modes. The yellow (), purple () and green () spikes show the locations of the peaks (local maxima) of the lab, field, and field calibrated spectra, respectively.

environment (R version 4.0.2; R [16]). Fig. 2 illustrates the extraction of target and predictor variables for a given sample k.

After arranging the pairs of target variables and their corresponding predictive features by following Eq. (1) and (2), SVR model was trained for each target waveband, aiming at estimating the reflectance value obtained in the lab, based on the observed values from the field measurements. The predicted value provided by these models (\hat{y}) is called here the estimated *field calibrated spectra* at each waveband. We chose SVR model for its capability to capture the complex relationships between the target and regressor variables based on the most significant deviation within the neighbouring wavebands through a nonlinear projection of the input features onto a higher-dimensional space, which does not require linearity assumptions about the relationship between the target and predictor variables. The SVR regression was implemented using R package *e1071* (Meyer et al., [11]).

The 262 resulting local models were each applied to the nine holdout data points of the testing dataset, resulting in the estimation of the field calibrated spectra (\hat{y}). The steps taken for applying the calibration method using lab and in-field spectra are illustrated in Fig. 3.

To further assess how closely the calibrated reflectance match the lab reflectance values of the developed calibration method, the percentage relative error (PRE) was calculated at each waveband following Eq. (3). Lower percentage relative errors indicate a better agreement between the calibrated and lab reflectance values vs. field and lab reflectance values, suggesting higher accuracy of the developed calibration method.

$$PRE = \left| \frac{X_{i(f,cf)} - Y_i}{Y_i} \right| * 100\%$$
(3)

where $X_{i(f, cf)}$, denotes either the field reflectance or the field calibrated reflectance measurement at each wavelength of *i* spanning from 404 nm to 2492 nm., and Y_i represents the reference reflectance of the lab measurements at each corresponding wavelength.

2.4. Assessing the performance of the calibration method and its verification by estimating NPK in potato plants

To verify the performance of the proposed calibration method, we estimated the NPK content in potato petioles using the lab and the field measurements before and after calibration. For each type, we built a dataset where the chemical composition of NPK acted as responses and the spectra acted as predictors. We used the NPK models generated by Abukmeil et al. [1] in their Eq. (1) where responses were predicted from predictors. Their models were built by Lasso multiple linear regression algorithms using a complexity parameter to shrink irrelevant coefficients to zero. Such algorithms could handle the large number of predictors compared to the small number of responses, and their results showed a reasonable estimation performance for NPK on data collected over an entire season. We applied those models on both in cross validation and the holdout splits following Eq. (4).



Fig. 5. PRE at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the dried mode of the (a) cross-validation, and (b) holdout split. The thick middle line represents the mean relative error.

 Table 2

 RMSE of estimating NPK using the lab vs. field spectra, and lab vs field calibrated spectra.

Estimated values using spectra	Dried n	Dried mode			Fresh mode		
	N	Р	К	N	Р	К	
Lab vs. field Lab vs. field calibrated	0.057 0.026	0.005 0.002	0.172 0.059	0.069 0.026	0.009 0.003	0.308 0.076	

$$\mathbf{y} = \beta_0 + \sum_{Z_i}^{i=1} \mathbf{x}_i \beta_i \tag{4}$$

where, β_i and β_0 are the regression coefficients of the i th waveband and intercept, respectively, and *Z* is a vector of spectrum inputs from 1 to 262. We identified the significant wavebands along with the coefficients to estimate NPK in this paper. The Root Mean Square Error (RMSE) between the actual concentrations and the estimated ones given by lab, field, and field calibrated spectra were used to assess the calibration performance. A successful calibration process should result in a stronger agreement between the estimated values obtained in the lab and the ones obtained by field calibrated spectra than the ones by raw field spectra.

3. Results and discussions

3.1. Raw spectra of in-field datasets in comparison to lab datasets

The visual comparison of the spectrum between in-field and lab measurements for dried leaves revealed similar patterns within the VNIR-SWIR ranges but divergent pattern within the Vis range. This is demonstrated by the locations of the peaks (local maxima) of the spectral curves as shown in Fig. 4. The peaks were computed by getting the first derivative of the curve and identifying the zeros of the function which are the wavebands at which the first derivative transitions from a positive to a negative value. In the dried mode (Fig. 4(a)), despite the similarity observed in the visual comparison between lab and field spectra, the results of the first derivative analysis still indicated few distinct peaks at various wavebands within the VNIR and SWIR as well as in the Vis range, as illustrated by the yellow and purple curves.

For the fresh leaves (Fig. 4(b)), the SWIR reflectance spectra visually exhibited similar spectral patterns, which is also exhibited with the locations of the peaks, whereas distinct differences were observed in the Vis and VNIR ranges. Across all data points, Vis reflectance was generally lower in lab measurements than in-field ones for the fresh mode. As noticed from the figure, there were discrepancies between the raw lab and field spectra which would be referred, in part, to sample handling.



Fig. 6. PRE at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the fresh mode of the (a) cross-validation, and (b) holdout split (bottom). The thick middle line represents the mean relative error.

Firstly, there was time gap of 3 days between scanning the leaves in field before reaching the lab and getting scanned there that could potentially result in a vegetation stress and a change in the proportion of lightabsorbing pigments [3,17]. Secondly, the position of the leaves in front of the sensor was different as the leaves are scanned in the field as shown in Fig. 1 whereas they are placed inside a chamber facing the probe in the lab instrument. Nevertheless, one reason why the reflectance spectra within SWIR may be more sensitive to such sample variability is that the SWIR region generally exhibits lower reflectance values compared to the Vis and VNIR, meaning that the signal-to-noise ratio within this region is lower [25].

3.2. Calibrated spectra of in-field datasets in reference to lab datasets

The calibration method developed in this work resulted in the generation of field calibrated spectra of the testing dataset. The calibrated spectra presented in Fig. 4 pertain to a single typical data point for each mode, representing the reflectance spectra for both fresh and dried leaves. The calibrated Vis-NIR-SWIR reflectance spectra exhibited similar spectral patterns at both the fresh and dried modes for all nine testing data points.

In the dried mode, a comparison of the field and field calibrated spectra showed that the calibration method have either transitioned peaks, created, or canceled peaks at different regions depending on the significant deviation of the neighbours influence captured by SVR model (Fig. 4(a)). For instance, in the Vis range one peak before calibration (in purple) transitioned from 552 to 584 nm after calibration (in green), whereas a new peak appeared at 624 nm in proximity to the peak of the lab spectra (in yellow) at 632 nm.

The differences between the field and field calibrated spectra in reference to lab spectra were also visualized in Fig. 5 by displaying the PRE values between the field and field-calibrated values in relative to lab measurement at each waveband. Fig. 5 demonstrates a considerable PRE reduction induced by the calibration procedure in the dried mode, both in the 5-fold cross-validation as well as for the holdout split. This reduction is particularly pronounced within the Vis range where the PRE values reduced from around 60 % to 20 % (Fig. 5(a) and (b)). Although the application of the SVR model at each waveband could align the nonlinear patterns between the lab and the field spectrum, the calibration did not completely match the field calibrated spectra in relative to lab spectra (green curve) which appears in creating a new peak around 440 nm as shown in Fig. 4(a). The calibration was also successful within the SWIR with a consistent PRE reduction from around 20 % downed to 10 %. This reduction would refer to the reason that the lab and field spectra had originally similar plateau of reflectance values and similar peaks as shown in Fig. 4(a).



(c)

Fig. 7. Estimated error of (a) N, (b), P, and (c) K estimation values of the five cross-validation folds and the holdout split of the dried mode by field and field calibrated spectra.

Similarly, the calibration succeeded in matching the plateau of the spectrum within the Vis in the fresh mode along with transitioning the peak before calibration from 544 nm to 568 nm making it closer to the lab peak at 576 nm and creating a peak at 624 nm to match the one in the lab spectra. Another important change can be observed in the VNIR where the peaks after calibration aligned with lab spectra, mainly due to the effect of the preprocessing step of scaling. The influence of scaling,

which the SVR automatically implements before starting the non-linear regression process [8], is more evident in the SWIR region (Fig. 4(b)) where the peaks were matching before implementing SVR. In this case scaling ranged the reflectance of the local neighbours between 0 and 1 which aligned the values of the lab and field spectra which matched the plateau across the SWIR to drop the PRE values from 120 % to 20 %.



Spectra 🗭 Field 💼 Field calibrated

(c)

Fig. 8. Estimated error of (a) N, (b), P, and (c) K estimation values of the five cross-validation folds and the holdout split of the fresh mode by field and field calibrated spectra.

3.3. Verification of the calibration method by the estimation of NPK

The calibration of reflectance readings resulted in an improved agreement of nutrient value estimation with the values obtained using the lab spectra (Figs. 7-8). In the dried mode (Fig. 7(a), (b), and (c)), the mean differences in estimated nutrient levels between models of field and lab readings and the field calibrated values for the holdout set were

0.55 %, 0.82 %, and 0.76 % for N, P, and K, respectively in comparison with mean discrepancies of 1.24 %, 1.81 %, and 2.35 % when using the raw (uncalibrated) field values. In the fresh mode (Fig. 8(a), (b), and (c)), a greater improvement was observed for the three nutrients: 0.57 % % vs. 1.67 % for N, 1.12 % % vs. 3.15 % for P, and 1.03 vs. 4.45 % for K. In particular, it is noteworthy that the deviation in the estimation of K dropped to around 1 % after calibration, in comparison to

approximately 4.5 % when using the raw (uncalibrated) field data.

Table 2 reports the RMSE values observed for the holdout set on both the dried and fresh modes.

The results show consistent improvement in estimating the nutrients after calibration. The more modest, but still relevant, improvement in the dried mode could be attributed to specific spectral features in the raw field spectra where significant wavebands are transitioned around waveband regions (690 nm – 720 nm) that already closely resembled the corresponding features observed in the lab spectra (Fig. 5). On the other hand, the greater improvement in the fresh mode may refer to the significant wavebands that are located within the SWIR where the calibration succeeded in reducing the PRE from 120 % to 20 % under the influence of scaling (Fig. 6). The results suggest that the calibration model can suppress the differences between the two spectral measurement modes across the spectrum. This might be of greater benefit when estimating the concentration of other nutrients that rely on a spectrum of the SWIR, such as Zn, Fe, B, Cu, and Al [1].

4. Conclusion

The developed calibration method included scaling and a neighborbased selection of variables with their squared terms prior to the implementation of SVR at each waveband. The calibration could successfully align the local maxima between lab and field spectra within the Vis range by capturing the significant deviation within the neighbouring wavebands through a nonlinear projection. The calibration further contributed to minimizing the PRE between lab and field spectra within the SWIR range of the fresh mode from 120 % to 20 % by the effect of scaling and local maxima alignment. The new calibration method was verified by estimating NPK in potato petioles using the three datasets of spectra: lab, field, and field calibrated spectra. The general consistent improvement in estimating NPK using field calibrated spectra in reference to lab estimations validates the ability of the developed calibration method to eliminate discrepancies between the two spectral measurements. The results of NPK estimation especially in the fresh mode highlight the importance of having consistent spectral measurements for nutrients that have significant wavebands positioned within the SWIR as an RMSE < 0.1 was achieved for the three elements.

Ethics statement

Not applicable: This manuscript does not include human or animal research. If this manuscript involves research on animals or humans, it is imperative to disclose all approval details.

CRediT authorship contribution statement

Reem Abukmeil: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. Ahmad Al-Mallahi: Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. Felipe Campelo: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

Data will be made available on request.

Acknowledgement

This work is supported by; the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Collaborative Research and Development Grant –Project (CRDPJ 543912–19), McCain Foods Limited, and Potatoes New Brunswick (PNB); and the New Brunswick Enabling Agricultural Research and Innovation (EARI) program of the Canadian Agricultural Partnership (CAP), project number: C1920–0056; and Mitacs Globalink Research Award, project number: IT32288. The authors would like to especially thank McCain Foods Limited for facilitating access to the Farms of the Future in Riverbank NB, where all leaf sampling was conducted.

References

- R. Abukmeil, A.A. Al-Mallahi, F. Campelo, New approach to estimate macro and micronutrients in potato plants based on foliar spectral reflectance, Comput. Electron. Agric. 198 (2022) 107074, https://doi.org/10.1016/j. compag.2022.107074.
- [2] K. Arai, Nonlinear mixing model of mixed pixels in remote sensing satellite images taking into account landscape, Internat. J. Advanced Comput. Sci. Applicat. 4 (1) (2013), https://doi.org/10.14569/IJACSA.2013.040115.
- [3] G.A. Carter, R.L. Miller, Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands, Remote Sens Environ. 50 (3) (1994) 295–302, https://doi.org/10.1016/0034-4257(94)90079-5.
- [4] R.M. Coliban, M. Marincaş, C. Hatfaludi, M. Ivanovici, Linear and non-linear models for remotely-sensed hyperspectral image visualization, Remote Sens 12 (15) (2020) 2479, https://doi.org/10.3390/rs12152479.
- [5] N.K. Fageria, M.P. Barbosa Filho, A. Moreira, C.M. Guimarães, Foliar fertilization of crop plants, J. Plant Nutr. 32 (6) (2009) 1044–1064, https://doi.org/10.1080/ 01904160902872826.
- [6] P. Guo, T. Li, H. Gao, X. Chen, Y. Cui, Y. Huang, Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using Vis-NIR spectroscopy, Remote Sens 13 (19) (2021) 4000, https://doi.org/10.3390/ rs13194000.
- [7] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction (Second Edition), Springer, New York, 2008, pp. 1–758.
- [8] C.-H. Ho, C.J. Lin, Large-scale linear support vector regression, J. Mach. Learn. Res. 13 (2012) 3323–3348.
- [9] G.J. Hochmuth, D. Maynard, C. Vavrina, E. Hanlon, E Simonne, Plant tissue analysis and interpretation for vegetable crops in florida. horticultural sciences department, UF/IFAS extension, HS964 Series, 2018. https://edis.ifas.ufl.edu/p ublication/ep081. accessed on June 01, 2021.
- [10] M. Krinitskiy, M. Aleksandrova, P. Verezemskaya, S. Gulev, A. Sinitsyn, N. Kovaleva, A. Gavrikov, On the generalization ability of data-driven models in the problem of total cloud cover retrieval, Remote Sens 13 (2) (2021) 326, https:// doi.org/10.3390/rs13020326.
- D. Meyer, Support vector machines, the interface to LIBSVM in package e1071, 2015. Austria. Available, https://cran.r-project.org/web/packages/e1071/vig nettes/svmdoc.pdf.
- [12] M.R. Motsara, R.N. Roy, Guide to laboratory establishment for plant nutrient analysis, Food and Agriculture Organization of the United Nations, Rome, 2008.
- [13] J.D. Ortiz, D. Avouris, S. Schiller, J.C. Luvall, J.D. Lekki, R.P. Tokars, R. C. Anderson, R. Shuchman, M. Sayers, R. Becker, Intercomparison of approaches to the empirical line method for vicarious hyperspectral reflectance calibration, Front. Mar. Sci. 4 (2017) 296, https://doi.org/10.3389/fmars.2017.00296.
- [14] J.A. Prananto, B. Minasny, T. Weaver, Chapter one near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues, Adv. Agron. 164 (2020) 1–49, https://doi.org/10.1016/bs. agron.2020.06.001.
- [15] L. Pompilio, L. Marinangeli, L. Amitrano, G. Pacci, S. D'andrea, S. Iacullo, E Monaco, Application of the empirical line method (ELM) to calibrate the airborne daedalus-CZCS scanner, Eur. J. Remote Sens. 51 (1) (2018) 33–46, https://doi.org/ 10.1080/22797254.2017.1399087.
- [16] R Core Team, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.R-pro ject.org/.
- [17] S. Rei, Q Wang, Towards a universal hyperspectral index to assess chlorophyll content in deciduous forests, Remote Sens 9 (3) (2017) 191, https://doi.org/ 10.3390/rs9030191.
- [18] J. Ren, J. Cui, W. Dong, Y. Xiao, M. Xu, S. Liu, J. Wan, Z. Li, J. Zhang, Remote sensing inversion of typical offshore water quality parameter concentration based on improved SVR algorithm, Remote Sens 15 (8) (2023) 2104, https://doi.org/ 10.3390/rs15082104.
- [19] H. Saari, I. Pellikka, L. Pesonen, S. Tuominen, J. Heikkilä, C. Holmlund, J. Mäkynen, K. Ojala, T. Antila, Unmanned aerial vehicle (UAV) operated spectral camera system for forest and agriculture applications, in: Proceedings volume 8174, remote sensing for agriculture, ecosystems, and hydrology XIII; 81740H. Event: SPIE remote sensing, Prague, Czech Republic, 2011, https://doi.org/ 10.1117/12.897585.

R. Abukmeil et al.

- [20] R. Silva, V. Gomes, A. Mendes-Faia, P. Melo-Pinto, Using support vector regression and hyperspectral imaging for the prediction of oenological parameters on different vintages and varieties of wine grape berries, Remote Sens 10 (2) (2018)
- 312, https://doi.org/10.3390/rs10020312.
 [21] K. Xu, Y. Gong, S. Fang, K. Wang, Z. Lin, F Wang, Radiometric calibration of UAV remote sensing image with spectral angle constraint, Remote Sens 11 (11) (2019) 1291, https://doi.org/10.3390/rs11111291.
 [22] Y. Xu, L Wang, K-Nearest neighbor-based weighted twin support vector regression,
- Appl. Intellig. 41 (2014) 299-309, https://doi.org/10.1007/s10489-014-0518-0.
- [23] J. Wang, C. Shen, N. Liu, X. Jin, X. Fan, C. Dong, Y. Xu, Non-destructive evaluation of the leaf nitrogen concentration by in-field visible/near-infrared spectroscopy in
- pear orchards, Sensors 17 (3) (2017) 538, https://doi.org/10.3390/s17030538.
 [24] H. Wickham, R. François, L. Henry, K. Müller, dplyr: A grammar of data manipulation. R package version 0.4, 2019, p. p156.
- [25] P. Williams, J. Antoniszyn, M. Manley, Near infrared technology: getting the best out of light, AFRICAN SUN MeDIA, 2019. http://doi.org/10.18820/9781 928480310.