**RESEARCH ARTICLE**

# Predicting Foliar Nutrient Concentrations across Geologic Materials and Tree Genera in the Northeastern United States Using Spectral Reflectance and Partial Least Squares Regression Models

**Wenxiu Teng[1], Qian Yu[1]\*, Ivan C. Mischenko[1], Alexandrea M. Rice[2], and Justin B. Richardson[2]**

[1]Department of Earth, Geographic, and Climate Sciences, University of Massachusetts, Amherst, MA, USA. [2]Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA.

*Address correspondence to: qyu@umass.edu

Spectral data can potentially offer a rapid assessment of nutrients in leaves and reveal information about the geologic history of the soil. This study evaluated the capability of the partial least squares regression (PLSR) for estimating foliar macro- and micronutrients (Ca, Mg, K, P, Mn, and Zn) using spectral data (400 to 2,450 nm). First, filter-based wavelength selection was conducted to reduce the independent variables. PLSR performance was then assessed across 4 geologic materials (coarse glacial till, glaciofluvial, melt-out till, and outwash) and 4 dominant tree genera (*Acer*, *Betula*, *Fagus*, and *Quercus*) in the northeastern United States. The spectral ranges 400 to 500 nm and 1,800 to 2,450 nm were found to be the most important spectral regions for estimating foliar nutrient concentrations. The developed PLSR model predicted 6 foliar nutrients with moderate to high accuracy (adjusted $R^2$ from 0.60 to 0.75). Foliar macronutrient concentrations were estimated with higher accuracy (mean adj. $R^2 = 0.69$) than micronutrient concentrations (mean adj. $R^2 = 0.635$). The prediction for the individual tree genera group and the individual geologic materials group outperformed the combined group; for instance, the adj. $R^2$ for estimating Ca and P was 39% higher for American beech (*Fagus grandifolia*) than all tree genera combined. Spectral measurements combined with wavelength selection and PLSR models can potentially be used to quantify foliar macro- and micronutrients at regional scales, and taking into account geologic materials and tree genera will improve this prediction.

## Introduction

Assessment of forest nutrients is an important task for determining forest health for agricultural productivity, timber and biofuel production, and ecosystem services such as habitat, pollutant sequestration, and water resource protection [1,2]. These important services are under threat across the northeastern United States from several large-scale co-occurring phenomena, such as climate change and species loss from diseases, and other factors that have unclear impacts on the growth of native northern hardwoods [3,4]. Traditional methods for assessing forest mineral nutrition of soil and leaf analyses to quantify nutrient uptake are costly and labor- and time-intensive [5,6], limiting our overall understanding of the spatial heterogeneity and processes governing forest nutrition; therefore, we need new, efficient tools that can be scaled up to entire forests.

Foliar nutrients have several major functions in plant metabolism in terms of establishing foliar structure, pigment synthesis, and metabolism as well as the electron transport chain, among others [7]. Various wavelengths of foliar spectral reflectance have been associated with foliar water content, pigment level, and nutrient concentration [8]. On this basis, spectral data in the form of hyperspectral reflectance provide an alternative source of information for monitoring plant leaf nutrients. Research that has focused on estimating nitrogen (N) content has significantly advanced model prediction accuracy and precision [9,10], but the evaluation of other nutrients and trace elements (e.g., Ca, Mg, P, K, Mn, and Zn) in plant leaves has been understudied and remains a challenge [11]. Estimating foliar nutrients using spectral reflectance is not a simple procedure due to difficulties in spectral autocorrelation and collinearity [12,13]. If the high dimensionality of spectral data cannot be properly reduced, the spectral features exhibit significant collinearity and the developed method can easily be overfit during the calibration process [14]. Overfitting by spectral models results in poor performance when they are validated or upscaled using independent test datasets. Therefore, methods for wavelength selection and prediction are essential for accurate estimations of foliar nutrient concentrations [15,16].

Some studies have explored using spectral data to predict macro- and micronutrients in leaves. Methodologically, partial

least squares regression (PLSR) and machine learning have emerged as the dominant methods for predicting foliar nutrients, with PLSR particularly noted for its ability to address overfitting issues and reduce spectral data's dimensionality into fewer uncorrelated components. Lu et al. [17] highlighted the shortwave-infrared (SWIR) region's (1,300 to 2,000 nm) effectiveness in predicting potassium (K) levels in rice leaves, establishing a significant correlation with leaf K content. Ramoelo et al. [18] combined spectral data with environmental variables using PLSR to estimate grass N and P concentrations, showcasing the method's ability to integrate climatic, edaphic, and topographic data for nutrient prediction. In another study, Malmir et al. [19] successfully applied visible–near-infrared spectral data (400 to 1,000 nm) and PLSR to estimate foliar calcium (Ca), potassium (K), phosphorus (P), and nitrogen (N) in cacao trees, though the prediction for K was less precise. Furthermore, Osco et al. [20] evaluated macro- and micro-nutrient content (N, P, K, Mg, S, Cu, Fe, Mn, and Zn) in Valencia Orange leaves using machine learning methods (e.g., Random Forest) combined with spectral data (380 to 1,020 nm). Gao et al. [21] applied hyperspectral remote sensing and a multi-factorial approach, incorporating topography, soil, vegetation, and meteorology, with machine learning algorithms to estimate forage P in alpine grassland. The larger datasets typically required by machine learning pose a challenge for our dataset, which comprises only 189 foliar samples across 4 geological materials (coarse glacial till, glaciofluvial, melt-out till, and outwash) and 4 dominant tree genera (*Acer*, *Betula*, *Fagus*, and *Quercus*), along with a high dimensionality of input variables featuring 2,151 wavelengths ranging from 350 nm to 2,500 nm. In contrast, PLSR has been shown to overcome the overfitting problem and reduce the dimensionality of the spectral data by transforming the high dimensionality of spectral data to a smaller number of uncorrelated components [22,23]. It performs well on relatively few samples with many predictor variables [24]. The PLSR method attempts to maximize covariance between the independent and dependent variables and keeps factors derived from the input spectral data orthogonal [13]. It is especially capable of handling highly correlated independent variables such as reflectance over a continuous spectrum [24,25]. Therefore, PLSR is the more suitable method due to the favorable conditions.

In this study, we assessed the relationship between foliar nutrients and spectral reflectance under potential genera-specific effects and impacts of geologic materials in temperate forests across the New England region (northeastern) of the United States. The 48 sites studied were situated in a geographic grid that captured variations across geologic materials, and focused on the dominant tree genera that span the region: *Acer* (*A. rubrum* and *A. saccharum*), *Betula* (*B. alleghaniensis*, *B. lenta*, and *B. papyrifera*), *Fagus grandifolia*, and *Quercus* (*Q. rubra* and *Q. alba*). Thereafter, spectral data (visible, near-infrared, and shortwave-infrared [VIS–NIR–SWIR], 400 to 2,500 nm) from fresh foliage and foliar macro- and micronutrient (Ca, Mg, K, P, Mn, Zn) concentrations were measured in the laboratory. Methods to predict foliar nutrient concentrations based on foliar spectra were developed using wavelength selection and the PLSR model. We hypothesized that the PLSR model would capture generalized relationships between foliar nutrient concentrations and foliar spectra. Further, we expected that grouping by tree genera and geologic materials would improve model accuracy due to genera-specific physiological

responses and mineral nutrition dependence on geologic materials underlying forest stands.

## Materials and Methods

### Study area
We studied 48 forested sites in a grid 180 km east–west by 240 km north–south across New England, with ~30-km segments between each site (Fig. 1), which varied by 1 to 3 km to ensure (a) a northern hardwood forest was present, (b) the forest was on slopes < 10° and physically accessible (at least 25 m away from any roads or human-made structures), and (c) the forest is at least moderately well-drained and shows no signs of seasonal flooding. Sites were moved if human disturbances were present or if they were >25% coniferous vegetation such as eastern hemlock (*Tsuga canadensis*) or white pine (*Pinus strobus*) to decrease conifer masking effects [26]. Geologic material at each site was identified using the USGS 1:5,000,000 Surficial Geology Map [27] and further confirmed through soil analyses (texture and rock fragments). Predominant parent geologic settings were coarse glacial till (primarily subglacial lodgement materials), melt-out till (primarily supraglacial materials), glaciofluvial deposits (from postglacial rivers and lakes), and outwash (plains and fans).

### Foliar sampling and chemical analysis
Mature trees, exhibiting no evidence of defoliation or disease and ranging from 15 to 25 cm DBH (diameter at breast height), were sampled at mid-canopy in late June and early July of 2019 across 48 sites within a 14-day timeframe. Foliage was collected from branches of 3 to 5 dominant trees, situated 4 to 25 m above the ground, using either a stainless-steel pole saw or an arborist throw-ball. In the throw-ball technique, a 0.4-kg throw-ball was lobbed over the upper canopy branches, and the branches were forcibly removed at their connection to the main trunk [28]. For shorter trees, a branch was collected from the main trunk using an extendible stainless steel pole saw. Foliage samples were collected from American beech (*Fagus grandifolia*), black birch (*Betula lenta*), red maple (*Acer rubrum*), white oak (*Quercus alba*), red oak (*Quercus rubra*), white birch (*Betula papyrifera*), and white ash (*Fraxinus americana*) in each forest stand and transported on ice in a cooler to the laboratory for spectral analysis within 5 h of collection. Overall, we collected 189 samples from 48 distinct sites. For each individual tree sampled, we ensured the collection of 2 to 3 replicate samples to bolster the reliability of our data. Specifically, we sampled 70 leaf samples for *Acer*, 23 for *Betula*, 27 for *Fagus*, 29 for *Quercus*, and 40 for other mixed species. The distribution of tree samples across different geological materials and genera is presented in Table 1.

Oven-dried foliage samples were digested to determine macro- and micronutrient concentrations using a modified EPA 3050B method [29], in which samples were combusted prior to strong acid digestion. To begin the process, plant material was dried to a constant mass at 45 °C in closed paper bags in a greenhouse. Mid-veins were removed, and the leaf blades were shredded and ground. A subsample of the ground material was transferred to an acid-washed ceramic vessel and combusted at 550 °C for 8 h. The ashes were transferred to 50-ml centrifuge tubes and digested with 5 ml of reverse aqua regia (9:1 $HNO_3$:HCl) and lightly capped to degas. The digest was diluted to 50 g using deionized water. Samples were further treated by diluting 3 g of the plant tissue
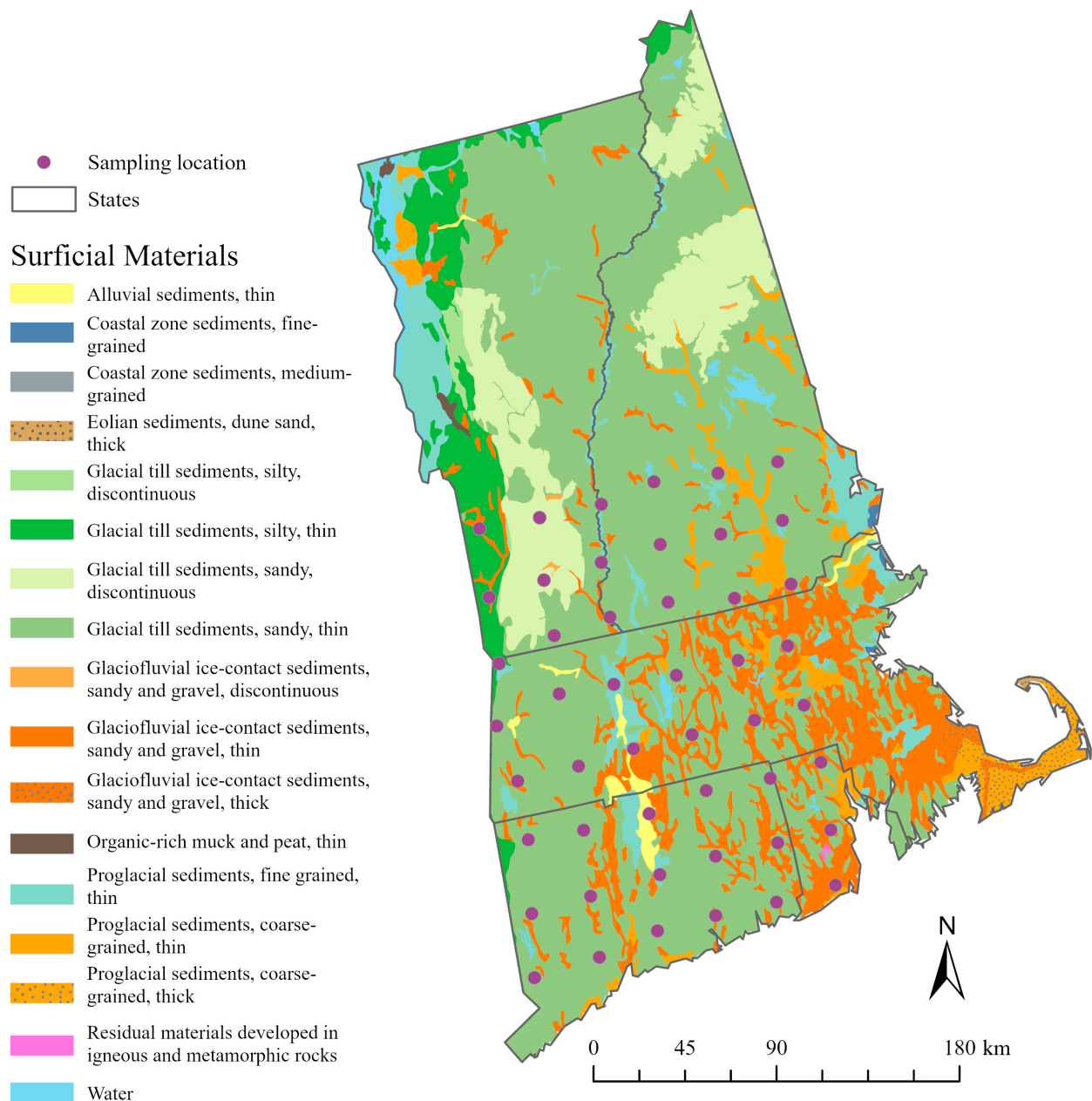
**Surficial Materials**

- Alluvial sediments, thin
- Coastal zone sediments, fine-grained
- Coastal zone sediments, medium-grained
- Eolian sediments, dune sand, thick
- Glacial till sediments, silty, discontinuous
- Glacial till sediments, silty, thin
- Glacial till sediments, sandy, discontinuous
- Glacial till sediments, sandy, thin
- Glaciofluvial ice-contact sediments, sandy and gravel, discontinuous
- Glaciofluvial ice-contact sediments, sandy and gravel, thin
- Glaciofluvial ice-contact sediments, sandy and gravel, thick
- Organic-rich muck and peat, thin
- Proglacial sediments, fine grained, thin
- Proglacial sediments, coarse-grained, thin
- Proglacial sediments, coarse-grained, thick
- Residual materials developed in igneous and metamorphic rocks
- Water

**Fig. 1.** Location of sampling sites and geologic surficial materials.

digest to 15 g using 2.5% w/v $HNO_3$ solution for analysis. Plant leaf digests were diluted with deionized water and analyzed for macro- (Ca, K, Mg, and P) and micronutrients (Mn and Zn) with an Agilent 5110 Inductively Coupled Plasma Optical Emission Spectrometer (Agilent Technology, Santa Clara, CA, USA).

## Spectral measurement

The foliar reflectance spectra at 350 to 2,500 nm were measured using a spectroradiometer, the ASD FieldSpec 3 full range [30]. Our ASD has a spectral resolution of 3 nm for VNIR. The leaves were collected in the field and transported to a dark lab to measure their spectra. The lamp is the light source for the measurement in the laboratory. Its brand is ProLamp and it was purchased as an ASD's standard accessory. The light intensity is 14.5 V/50 W. It is a full spectral lamp (not mono-color or

LED). Spectral reflectance of freshly collected leaves was measured within 5 h of field sampling. A full-spectrum ASD Pro Lamp was mounted on a tripod and pointed at the leaves in a darkened lab. A Spectralon white reference panel with nearly 100% reflectance was used to calibrate illumination. Both the leaves and white reference panel were positioned with a fixed geometry to the optical fiber cable tip and the lamp. The fiberoptic cable has a field of view of 25°. The laboratory measurement was set up with an incidence angle of 18.5° to avoid the shadow. The distance of fore optic to the leaves or white reference was 8 cm. The footprint has a diameter of 3.5 cm. The ASD recorded an average of 10 scans per spectrum. The integration time for ASD scanning was 17 ms. The dark current was automatically subtracted in RS3 software in the reflectance mode after the optimization. We also recorded 3 spectra for

**Table 1.** The distribution of tree samples across different geological materials and genera

| Geological material | Genus | Number of trees |
|---|---|---|
| Coarse glacial till | *Acer* | 36 |
| Coarse glacial till | *Betula* | 16 |
| Coarse glacial till | *Fagus* | 15 |
| Coarse glacial till | *Quercus* | 7 |
| Coarse glacial till | *Others* | 20 |
| Glaciofluvial | *Acer* | 18 |
| Glaciofluvial | *Betula* | 4 |
| Glaciofluvial | *Fagus* | 5 |
| Glaciofluvial | *Quercus* | 9 |
| Glaciofluvial | *Others* | 8 |
| Melt-out till | *Acer* | 12 |
| Melt-out till | *Betula* | 3 |
| Melt-out till | *Fagus* | 6 |
| Melt-out till | *Quercus* | 8 |
| Melt-out till | *Others* | 10 |
| Outwash | *Acer* | 4 |
| Outwash | *Betula* | 0 |
| Outwash | *Fagus* | 1 |
| Outwash | *Quercus* | 5 |
| Outwash | *Others* | 2 |

each foliage sample. The spectral curves showed a small fluctuation at 2 ends of the spectrum due to the low light intensity of the lamp. To eliminate the fluctuations, the 2 ends of the raw spectra were trimmed, and reference spectra between 400 and 2,450 nm were used for this study (Fig. 2A and B). The spectral curve was first transformed to first derivative, which enhances the detectability of absorption features that may not be accurately captured or even detected in spectral curves [20,31–33]. The advantage of derivative transformation over the spectral curve lies in its ability to eliminate background interferences, separate overlapping spectra, and minimize baseline drift in raw spectra [34]. Studies focusing on the application of derivative analysis to plant reflectance curves have identified strong correlations with nitrogen (N) [31,32] and cadmium (Cd) concentrations [33]. These findings indicate that derivative analysis effectively highlights components that are typically challenging to detect, demonstrating its utility in spectral analysis. Therefore, we used the first derivative of the reflectance (Fig. 2C and D) to build the PLSR model in this study.

### Partial least squares regression

The PLSR model (Fig. 3) includes an $x$ model and a $y$ model and is generally described in the following form with equivalent matrix notations in parentheses:

$$x_{ik} = \sum_a t_{ia} p_{ak} + e_{ik} \qquad \left(X = TP' + E\right) \qquad (1)$$

$$y_{im} = \sum_a u_{ia} c_{am} + g_{im} \qquad \left(Y = UC' + G\right), \qquad (2)$$

where $x_{ik}$ is the spectral data matrix for foliar samples, $i$ is the number of foliar samples, $k$ is the number of spectral wavelengths from 400 to 2,450 nm, and $a$ is the number of components. In this study, $i$ and $k$ are 189 and 2,051, respectively. $p_{ak}$ is the loading matrix, and $e_{ik}$ are the $X$-residuals. $y_{im}$ is the nutrient data matrix for foliar samples, $u_{ia}$ is the $Y$-scores, $c_{am}$ is mass, $g_{im}$ are the residuals, and $m$ is the number of nutrients to be modeled. We fit a PLSR for each individual nutrient; thus, $m$ is equal to 1 in this study.

PLSR attempts to find a few "new" uncorrelated PLS components (also known as latent variables) to overcome overfitting [13]. These "new" spectral variables are also called $X$-scores and the linear combinations of the response variables are called $Y$-scores. The formulas for $X$-scores and $Y$-scores are shown below in both element and matrix form:

$$t_{ia} = \sum_k x_{ik} w^*_{ka} \qquad \left(T = XW^*\right) \qquad (3)$$

$$u_{ia} = \sum_k y_{im} c^*_{ma} \qquad \left(U = YC^*\right), \qquad (4)$$

where $w^*_{ka}$ and $c^*_{ma}$ are masses. Since $X$-scores are good predictors of $Y$ [13], the foliar nutrients can be estimated as:

$$y_{im} = \sum_a t_{ia} c^*_{am} + f_{im} \qquad \left(Y = TC' + F\right), \qquad (5)$$

where $f_{im}$ comprise the $Y$-residuals, the deviations between the observed and modeled responses. Equations 3 and 5 are merged to obtain a multiple regression model:

$$y_{im} = \sum_k x_{ik} w^*_{ka} \sum_a c^*_{am} + f_{im} = \sum_k x_{ik} b_{km} + f_{im} \\ \left(Y = XW^* C' + F = XB + F\right), \qquad (6)$$

where $b_{mk}$ comprise the PLS-regression coefficients ($\beta$) and can be written as:

$$b_{km} = \sum_a w^*_{ka} c^*_{am}; \qquad \left(B = W^* C'\right) \qquad (7)$$

A high absolute value of the regression coefficient $b_{km}$ indicates that the specific wavelength has a high correlation with the foliar nutrient concentration. The PLSR model aims to maximize the covariance between $T$ and $U$. The PLSR analysis and evaluation were coded using Python. Specifically, we used the PLSRegression class from the sklearn.cross_decomposition module, part of the scikit-learn library [35], version 0.24.2. The algorithm development and data analysis were conducted using Jupyter Notebook and PyCharm as integrated development environments (IDEs).

### Wavelength selection and PLSR predictive model

PLSR is a suitable method to overcome the overfitting problem and to reduce the dimensionality of the spectral data. However, very high dimensions and small sample size can still alter the PLSR results, and a large number of irrelevant variables may yield large variations on the prediction based on the test set
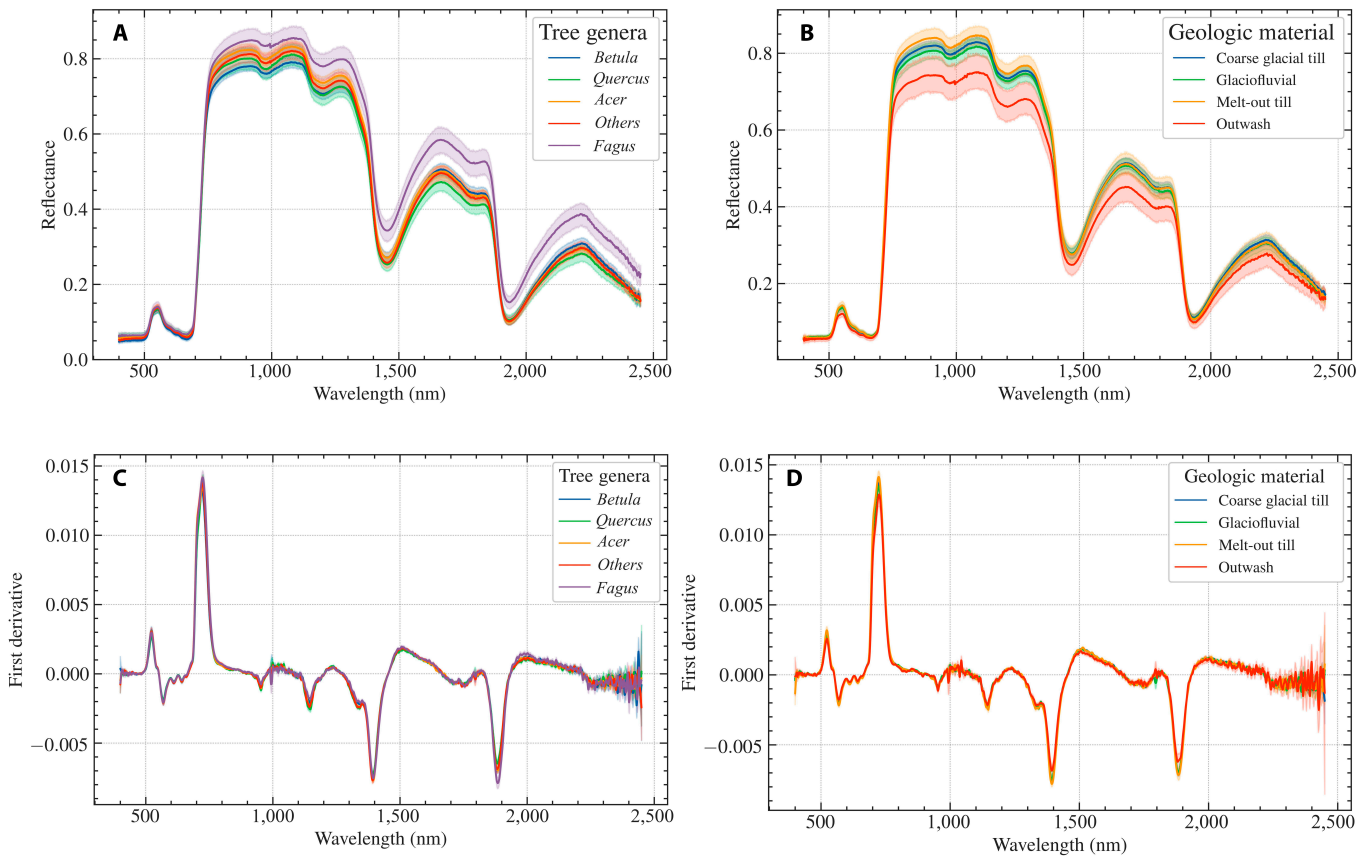
**Fig. 2.** Spectral reflectance curves. Mean (±SD) spectral reflectance for (A) tree genera and (B) geologic materials. First derivative of the reflectance for (C) tree genera and (D) geologic materials.
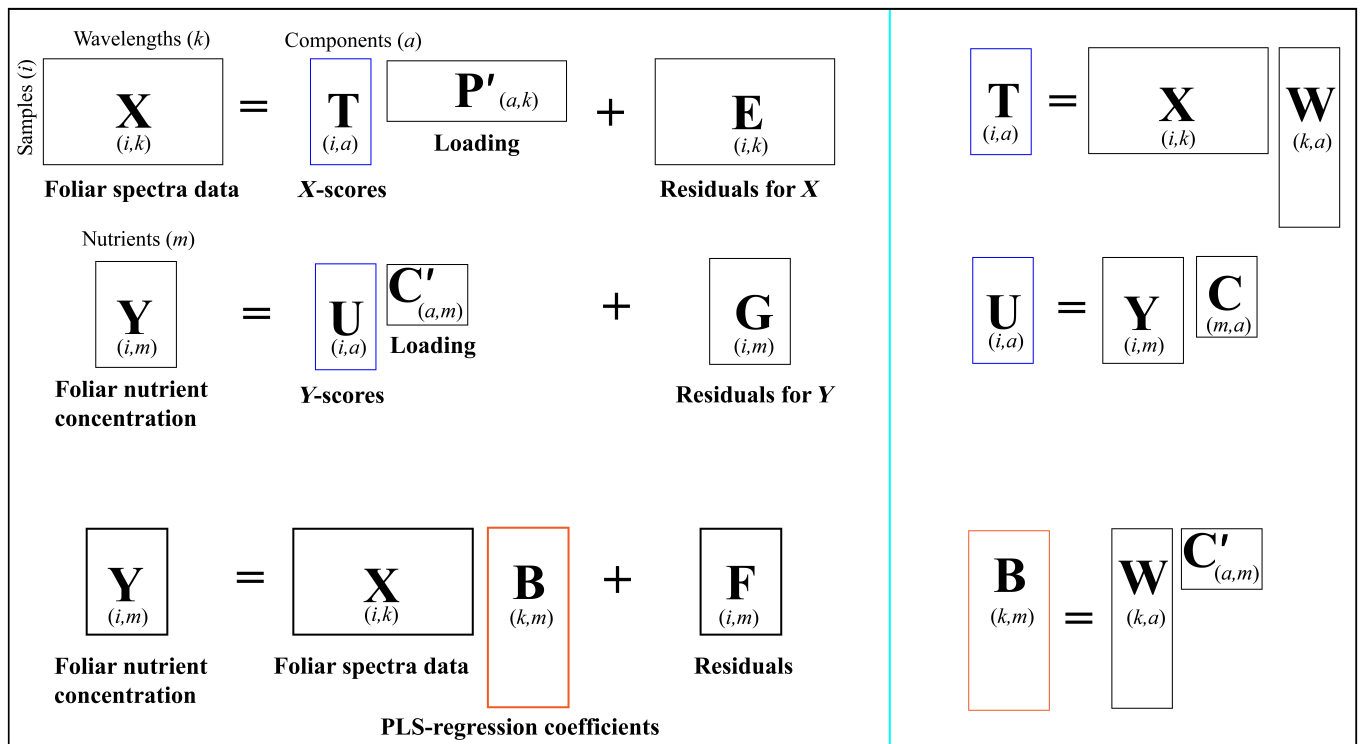


**Fig. 3.** Diagram of the PLSR model.

[36]. Therefore, a PLS-based filter method was used to select significant wavelengths to improve the estimation [36]. The filter-based methods select wavelengths in 2 steps. First, the PLSR model was fitted to the spectral and foliar nutrient data. Then, regression coefficients ($\beta$), a single measure of association between each wavelength and the foliar nutrient concentration, are used to select wavelengths. The wavelength with the lowest correlation was discarded according to the regression coefficients ($\beta$). The above 2 procedures were iterated until the root mean square error (RMSE, Eq. 10) decreases. Finally, the remaining wavelengths were used to establish a new PLSR between spectral and foliar nutrient data. For the new PLSR model, the optimized PLS components with the lowest RMSE were selected by searching from PLS components, with the number of components capped at 30 since we observed the decrease of RMSE becoming insignificant between 7 and 24 components for individual nutrients.

## Method evaluation metrics

The coefficient of determination ($R^2$; Eq. 8), adjusted $R^2$ ($R^2_{Adj.}$, Eq. 9), and RMSE (Eq. 10) were used to evaluate the accuracy of predicted foliar nutrient concentrations. Residual prediction deviation (RPD) [37] was then used to further evaluate the reliability of the predictions (Eq. 11). The prediction ability of the model was considered to be good when RPD values were greater than 1.4 and to be excellent when RPD values were greater than 2 [38]. Generally, larger values of $R^2$ and RPD and a smaller value of RMSE indicate a method with good predictability.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{8}$$

$$R^2_{Adj.} = 1 - \frac{n-1}{n-p-1}\left(1-R^2\right) \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{n}} \tag{10}$$

$$RPD = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n}}}{RMSE} \tag{11}$$

where $\widehat{y}_i$ is the predicted value, $y_i$ is the measured value, $\overline{y}$ is the mean of the measured values, $n$ is the number of samples, and $p$ is the number of independent variables. The metrics were calculated using a 10-fold cross-validation procedure for each PLSR model.

## Results and Discussion

### Foliar nutrient variations across tree genera and geologic materials

Foliar nutrient concentrations by geologic material and plant genera are shown in Fig. 4. Foliar nutrient concentrations were

comparable with previous studies in the region [26,28,39]. Among geologic materials, foliar Ca, Mg, P, K, and Zn nutrient concentrations were highest on coarse glacial till and glaciofluvial materials. Higher foliar nutrient concentrations are most certainly due to the abundance of less weathered, nutrient-bearing minerals in glacial till compared to highly weathered minerals present in glaciofluvial soil parent material [28,40]. Comparing among tree genera, foliar Ca, Mg, K, and Zn nutrient concentrations were higher for *Betula* but similar for *Acer*, *Fagus*, and *Quercus*, despite co-occurring among the same surficial deposits. These results suggest that foliar response to soil nutrient uptake rates were genera-dependent, with *Betula* individuals being high accumulators and potentially less sensitive to differences in nutrient availability across soil and surficial geological materials. The bioaccumulation of elements by *Betula* has been reported in several previous studies, whereas other tree species such as *Fagus* tend to have lower elemental concentrations [26,41].

The descriptive statistics of 6 nutrient concentrations are presented in Table 2. The mean concentrations reveal that Ca has the highest average level among the nutrients studied, with a mean of 7,066 mg/kg, indicating its abundant presence in the foliar samples. Mg and P also show significant mean concentrations of 2,295 mg/kg and 1,836 mg/kg, respectively, while K has a lower mean value of 804 mg/kg. The variability in nutrient concentrations is captured by the standard deviation, with Ca showing the highest variation (4,434 mg/kg) among the macronutrients, which suggests a wide range of Ca levels across the samples. Similarly, Mn exhibits considerable variability among the micronutrients, with a standard deviation nearly equal to its mean (748 mg/kg), pointing to diverse Mn concentrations within the foliar samples. These descriptive statistics underscore the diverse nutrient profiles within the foliar samples, reflecting the complex interplay of environmental, genetic, and soil factors influencing nutrient uptake and distribution.

### Significant wavelengths for predicting foliar nutrient concentrations

Spectral reflectance, selected wavelengths, and PLS coefficients associated with the PLSR model for predicting 6 different foliar nutrients are shown in Fig. 5. The high absolute value of a PLS coefficient indicates a significant wavelength. Overall, the selected significant wavelengths were similar among 6 foliar nutrients, which focused on 400 to 500 nm, around 1,000 nm, and 1,800 to 2,450 nm. These results are partly in line with previous findings that the green region around 470 to 520 nm is important for predicting Ca concentrations [19], and agree with the physiological function of Ca promoting greater chlorophyll or chloroplasts [42,43], likely through indirect enhancement via photoprotection and regulating photosynthetic electron transfers [44]. The higher abundance of Ca allows for greater green pigmentation for photosystem II. Although the selected wavelengths shift slightly between different foliar nutrients, most of the important wavelengths remained at a similar region, suggesting that these wavelengths are important for estimating the concentrations of important foliar nutrients. For Mg and Zn, the red edge (~710 nm) is also important, unlike for the other 4 foliar nutrients, which agrees with the finding that greater Mg is associated with higher anthocyanin levels [45], seen as greater red pigmentation in leaves.

The value of the coefficient (negative or positive) indicates the importance of the wavelength in terms of explaining the
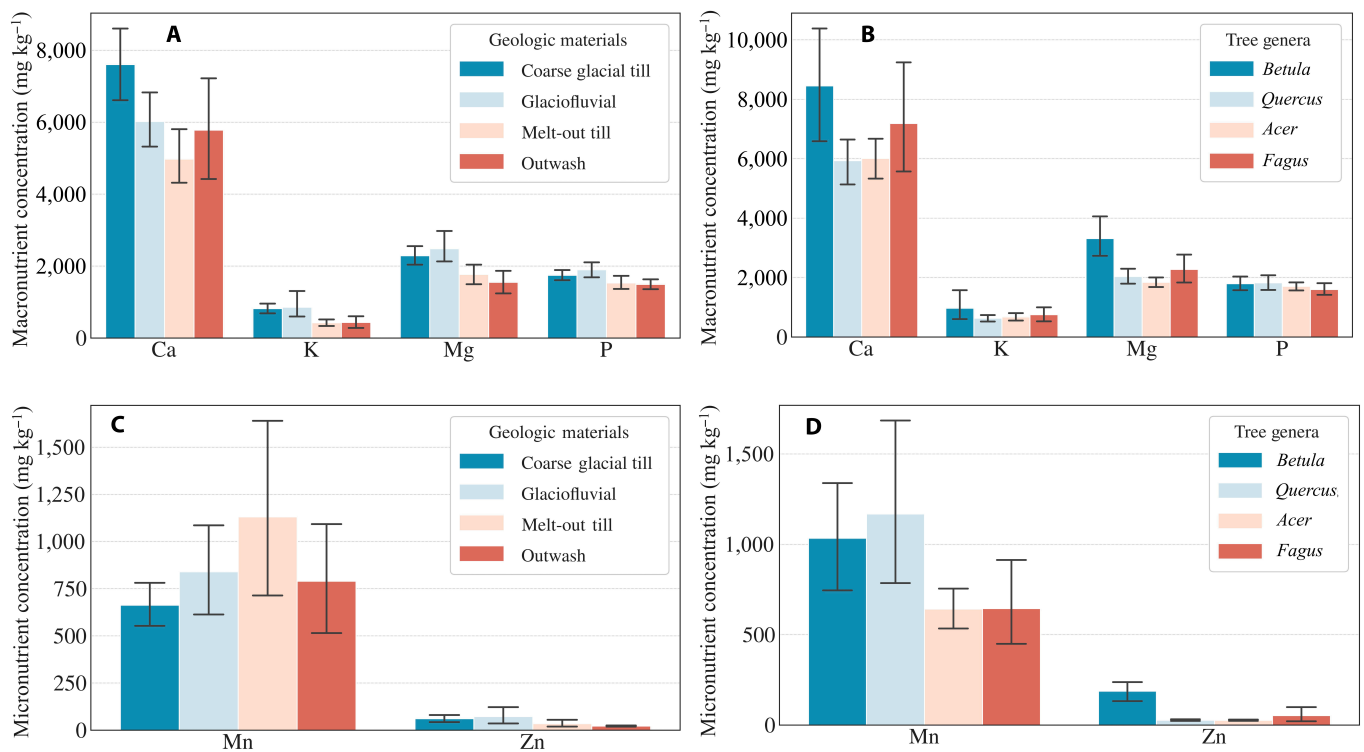
**Fig. 4.** Foliar nutrient concentrations by geologic material and tree genus. Macronutrient concentrations (Ca, K, Mg, and P) across (A) geologic materials and (B) tree genera. Micronutrient concentrations (Mn and Zn) across (C) geologic materials and (D) tree genera. The whiskers show 5th and 95th percentile.

**Table 2.** Descriptive statistics for 6 foliar nutrient concentrations

| Summary | Macronutrients (mg/kg) | | | | Micronutrients (mg/kg) | |
|---|---|---|---|---|---|---|
| | Mg | P | Ca | K | Mn | Zn |
| Mean | 2,295 | 1,836 | 7,066 | 804 | 748 | 58 |
| SD | 1,182 | 698 | 4,434 | 743 | 745 | 90 |
| Min | 390 | 762 | 1,591 | 9.4 | 19.7 | 6.4 |
| Max | 9,747 | 5,159 | 25,260 | 3,617 | 5,860 | 588 |

variance of foliar nutrient concentrations [46,47]. Among the significant wavelength regions mentioned above, the PLS coefficients for 6 foliar nutrients are particularly high at 400 to 500 nm and 2,200 to 2,400 nm. These 2 spectral regions were the most important wavelengths for the retrieval of different foliar nutrients. The high correlation in the 2,200 to 2,400 nm region could be attributed to the SWIR region's ability to penetrate deeper into leaf tissues compared to other spectral regions. This penetration allows for the detection of specific chemical bonds associated with foliar nutrients. For instance, absorption features in this region are often related to water content, cellulose, lignin, and other biochemical components [48–50], which can indirectly correlate with nutrient levels due to their influence on plant physiological status and health.

## PLSR estimates for 6 plant nutrients

The PLSR estimates were validated using measured concentrations for 6 foliar nutrients (Fig. 6), and the descriptive regression statistics are presented in Table 3. Overall, the PLSR algorithms predict foliar nutrient concentrations with moderate to high accuracy (adj. $R^2$ from 0.60 to 0.75). Among all macronutrients studied, Ca was the most accurately estimated by PLSR through cross-validation (adj. $R^2 = 0.75$). We attribute 2 potential mechanisms to the accuracy of the PLSR model for estimating foliar Ca concentrations. First, Ca is commonly a limiting nutrient in temperate forests of New England, and Ca deficiencies may be more pronounced than other macronutrients due to historical losses from acid rain. Second, Ca is a complex macronutrient due to its many roles in structures, chemical signaling, and as an enzyme cofactor [51] and thus may express many controls on plant health via pigmented compounds such as chlorophylls or anthocyanins [52]. Similar modes of actions are hypothesized for K and Mg, with a clear linkage between Mg availability and chlorophyll $a$ and $b$ in foliage. Both micronutrients, Mn and Zn, had lower accuracy (adj. $R^2 = 0.60$ for Mn; adj. $R^2 = 0.61$ for Zn) compared to the 4 macronutrients for predicting foliar concentrations. We hypothesize that direct linkage between foliar Mn and Zn concentrations with spectra reflectance could be governed by 3 mechanisms. First, Mn and Zn are required as enzyme cofactors and their deficiencies could generate subtle changes to leaf pigments. Second, high Mn and Zn could negatively impact leaf nutrition [53,54], decreasing the health of the foliage. Lastly, Mn and Zn concentrations may not directly affect leaf pigment compounds and rather simply coincide with lower Ca or Mg concentrations. Our results demonstrate that macronutrients and micronutrients are quantifiable from spectral
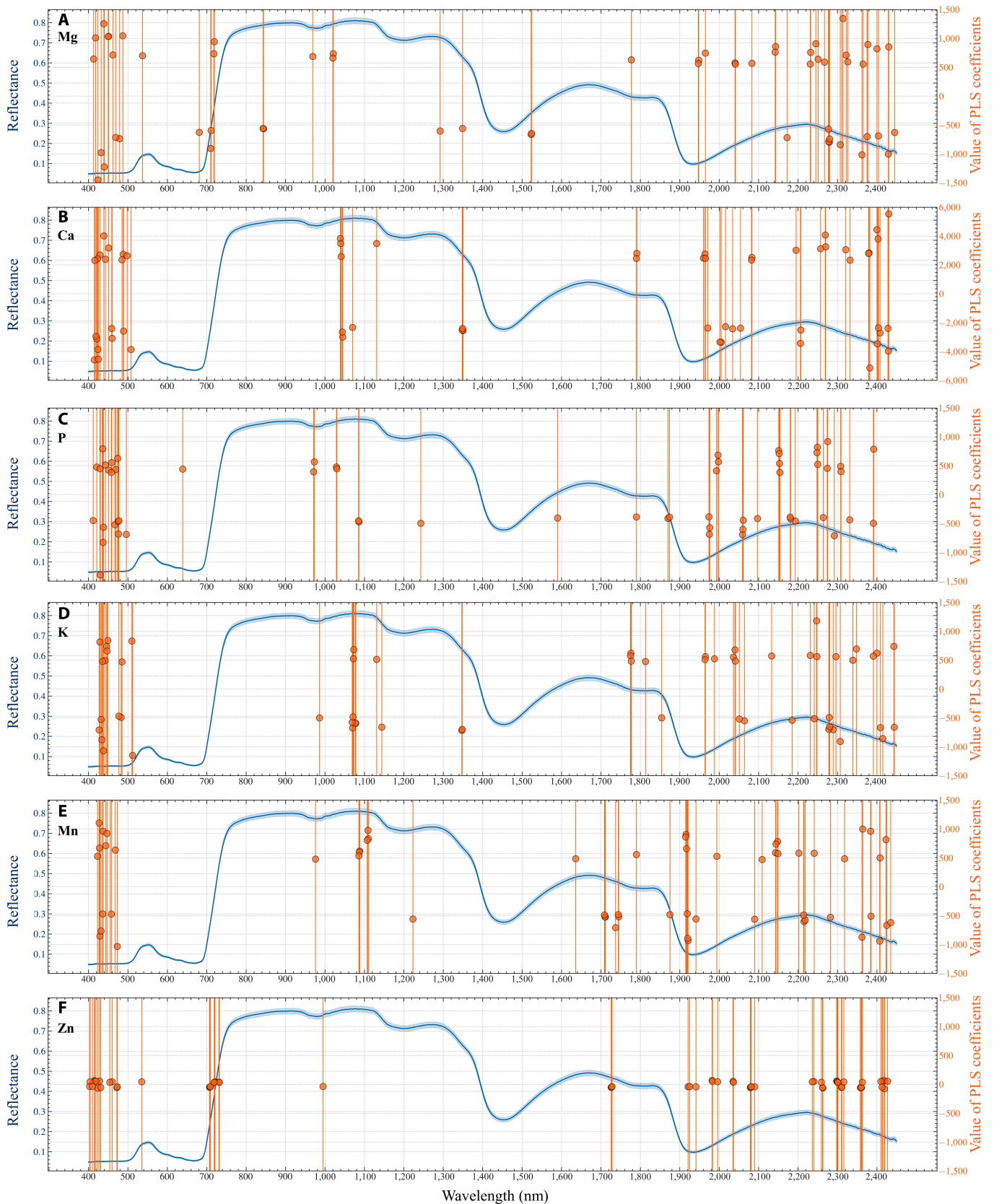
**Fig. 5.** Spectral reflectance (blue line), selected wavelengths (orange line), and PLS coefficients (orange points) associated with the PLSR method for predicting 6 different foliar nutrients. Only the top 60 significant wavelengths are shown for easier visualization. (A) Mg, (B) Ca, (C) P, (D) K, (E) Mn, and (F) Zn. The gray lines represent gridlines that correspond with the tick marks on the left Y-axis or right Y-axis.
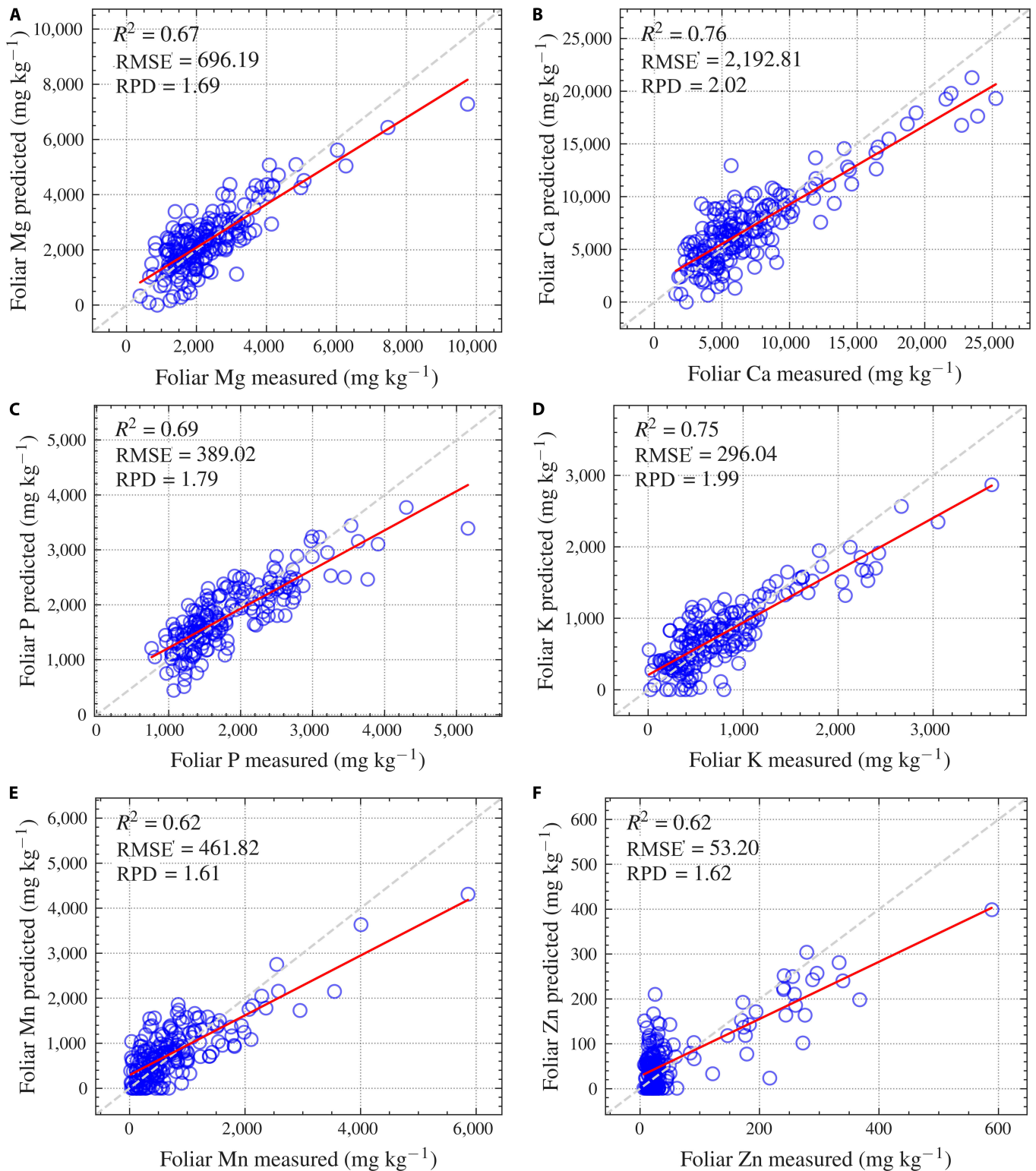
**Fig. 6.** PLSR prediction for foliar (A) Mg, (B) Ca, (C) P, (D) K, (E) Mn, and (F) Zn.

reflectance and PLSR model, with better prediction for macronutrients than micronutrients. Macronutrients like Ca and K are present in higher concentrations in foliar and are involved in major structural components and physiological processes. This higher concentration may result in more pronounced spectral signatures that can be more easily detected and quantified

through spectral reflectance. On the other hand, micronutrients like Mn and Zn are required in much smaller amounts and may not significantly alter the spectral reflectance to the same extent, making their detection and quantification more challenging.

The number of significant wavelengths in estimating 6 foliar nutrient concentrations ranged between 405 and 576.

**Table 3.** PLSR model results for 6 foliar nutrients

| Nutrient | $R^2$ | $R^2_{Adj.}$ | RMSE | RPD | No. of wavelengths | No. of PLS components |
|----------|-------|--------------|------|-----|--------------------|-----------------------|
| Mg | 0.67 | 0.64 | 696.19 | 1.69 | 474 | 17 |
| P | 0.69 | 0.66 | 389.02 | 1.79 | 576 | 16 |
| Ca | 0.76 | 0.75 | 2,192.81 | 2.02 | 543 | 10 |
| K | 0.75 | 0.71 | 296.04 | 1.99 | 576 | 24 |
| Mn | 0.62 | 0.60 | 461.82 | 1.61 | 405 | 9 |
| Zn | 0.62 | 0.61 | 53.20 | 1.62 | 436 | 7 |

The resulting PLSR models contain 7 to 24 PLS components (Table 3). For all macronutrients, more than 450 significant wavelengths were selected at the wavelength selection step, and more than 10 uncorrelated PLS components were used by the final PLSR model. In contrast, micronutrients have fewer significant wavelengths and PLS components, and their concentrations appear to have a weaker relationship with spectral reflectance than those of the macronutrients. These results confirm again that spectral reflectance and the PLSR model more accurately predict macronutrient than micronutrient concentrations.

## PLSR model results across geologic materials

Successful PLSR models were developed for each of the 4 geologic materials for Ca and P. For each group, the predicted foliar Ca and P concentrations were validated using measured concentrations shown in Fig. 7, and the descriptive regression statistics are presented in Table 4. The PLSRs fitted for the different geologic materials significantly improved foliar P prediction while maintaining similar accuracy for Ca. The PLSRs for individual geologic materials for foliar P showed a higher level of accuracy (average adj. $R^2 = 0.86$) compared to the combined group (adj. $R^2 = 0.66$). This increase in model prediction accuracy suggests that geologic material is an important factor affecting foliar P. In contrast, PLSR models for geologic materials for predicting foliar Ca concentrations (average adj. $R^2 = 0.78$) did not significantly improve over the combined group (adj. $R^2 = 0.75$). Geologic materials did not exhibit a strong influence on foliar Ca, unlike the case for foliar P. Among the 4 geologic materials, outwash yielded the highest predictive accuracy for both foliar Ca and P concentrations. Foliar Ca concentration predictions for glaciofluvial materials were hampered by outliers with a high concentration. The number of selected wavelengths used to estimate foliar Ca and P concentrations among geologic materials ranged between 121 and 345, and the number of PLS components ranged between 2 and 6 (Table 4), which is relatively lower likely due to the smaller sample size.

There are several possible explanations for geologic materials affecting nutrients and nutrient quantification using PLSR. First, geologic materials control the sourcing of nutrients to the soils and trees, and less chemically weathered geologic materials, such as glacial till, can supply more nutrients to trees than extensively chemically weathered geologic materials such as glaciofluvial deposits. Second, the physical nature of the geologic material can improve or diminish tree growth by affecting water movement and storage. Coarse particles of glacial till hold less water

**Table 4.** PLSR model results for Ca and P by groups of geologic materials

| | | Coarse glacial till | Glaciofluvial | Melt-out till | Outwash |
|---|---|---------------------|---------------|---------------|---------|
| $R^2$ | Ca | 0.77 | 0.67 | 0.82 | 0.98 |
| | P | 0.80 | 0.89 | 0.90 | 0.91 |
| $R^2_{Adj.}$ | Ca | 0.76 | 0.65 | 0.79 | 0.93 |
| | P | 0.79 | 0.86 | 0.88 | 0.89 |
| RMSE | Ca | 2,350.72 | 2,689.30 | 1,320.31 | 692.62 |
| | P | 345.62 | 249.37 | 207.06 | 64.18 |
| RPD | Ca | 2.06 | 1.71 | 2.32 | 3.11 |
| | P | 2.22 | 2.62 | 2.79 | 3.28 |
| No. of wavelengths | Ca | 316 | 328 | 135 | 112 |
| | P | 326 | 345 | 218 | 121 |
| No. of PLS components | Ca | 5 | 3 | 6 | 3 |
| | P | 6 | 5 | 4 | 2 |
| No. of samples | Ca | 94 | 44 | 39 | 12 |
| | P | 94 | 44 | 39 | 12 |

and have higher infiltration rates than the finer particles of glaciofluvial deposits [28]. Both of these influences of geologic materials on tree growth and nutrition across glacial till and glaciofluvial geologic materials were reported previously for the region [28]. Lastly, the improved PLSR relationship between nutrients from geologic materials and foliar nutrient concentrations may be attributed to reducing the across-group variability among the data.

## PLSR model results across tree genera

As done for the geologic materials, PLSR models were developed for foliar Ca and P concentrations for each tree genus (Fig. 8), and the descriptive regression statistics are presented in Table 5. Predictions for the tree genera groups also improved the model accuracy over the combined model and even slightly better than geologic materials. This result implies that the tree genus
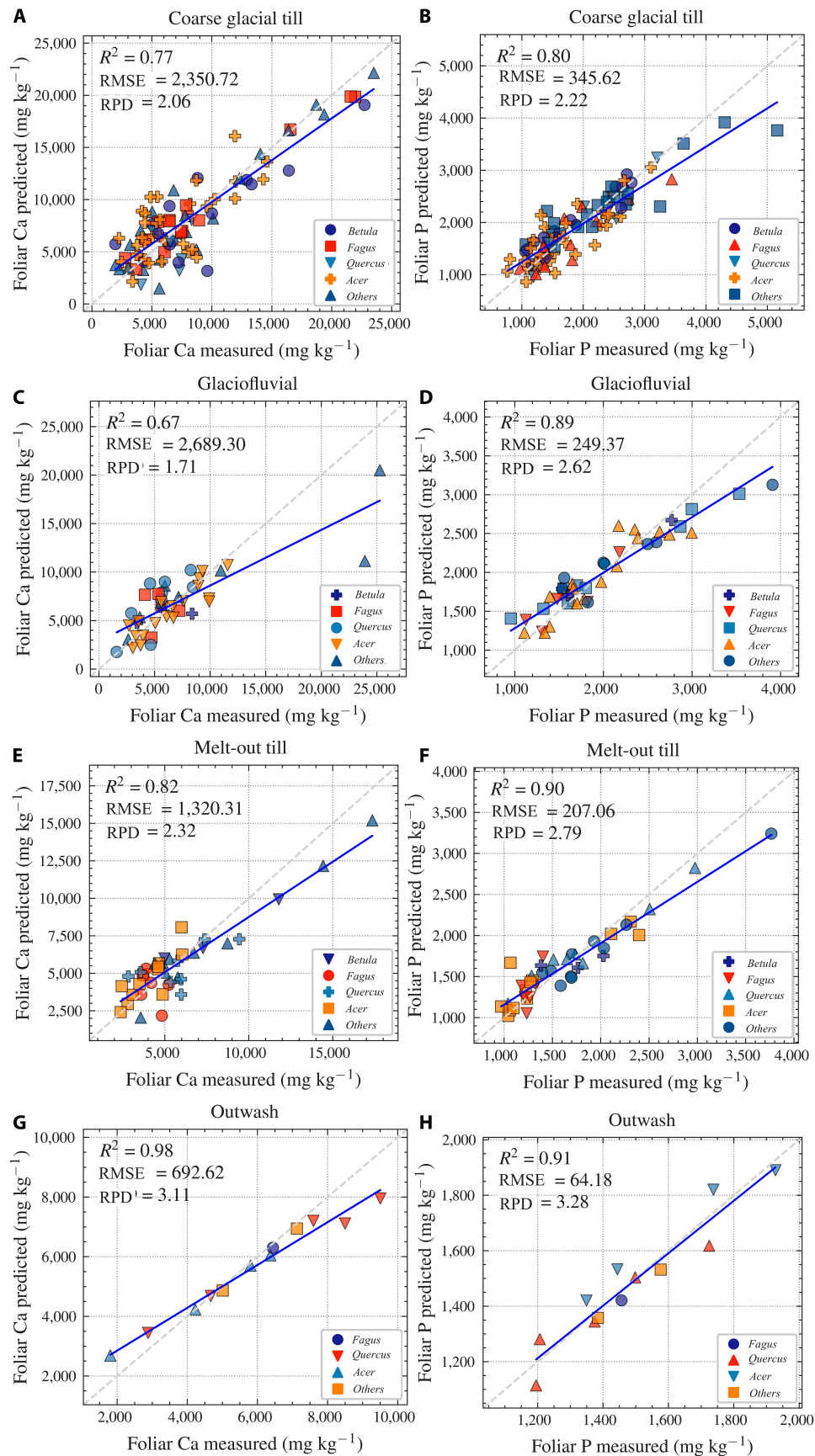
**Fig. 7.** PLSR model results for Ca and P by the categories of geologic material. (A) Coarse glacial till for Ca, (B) coarse glacial till for P, (C) glaciofluvial for Ca, (D) glaciofluvial for P, (E) melt-out till for Ca, (F) melt-out till for P, (G) outwash for Ca, and (H) outwash for P.
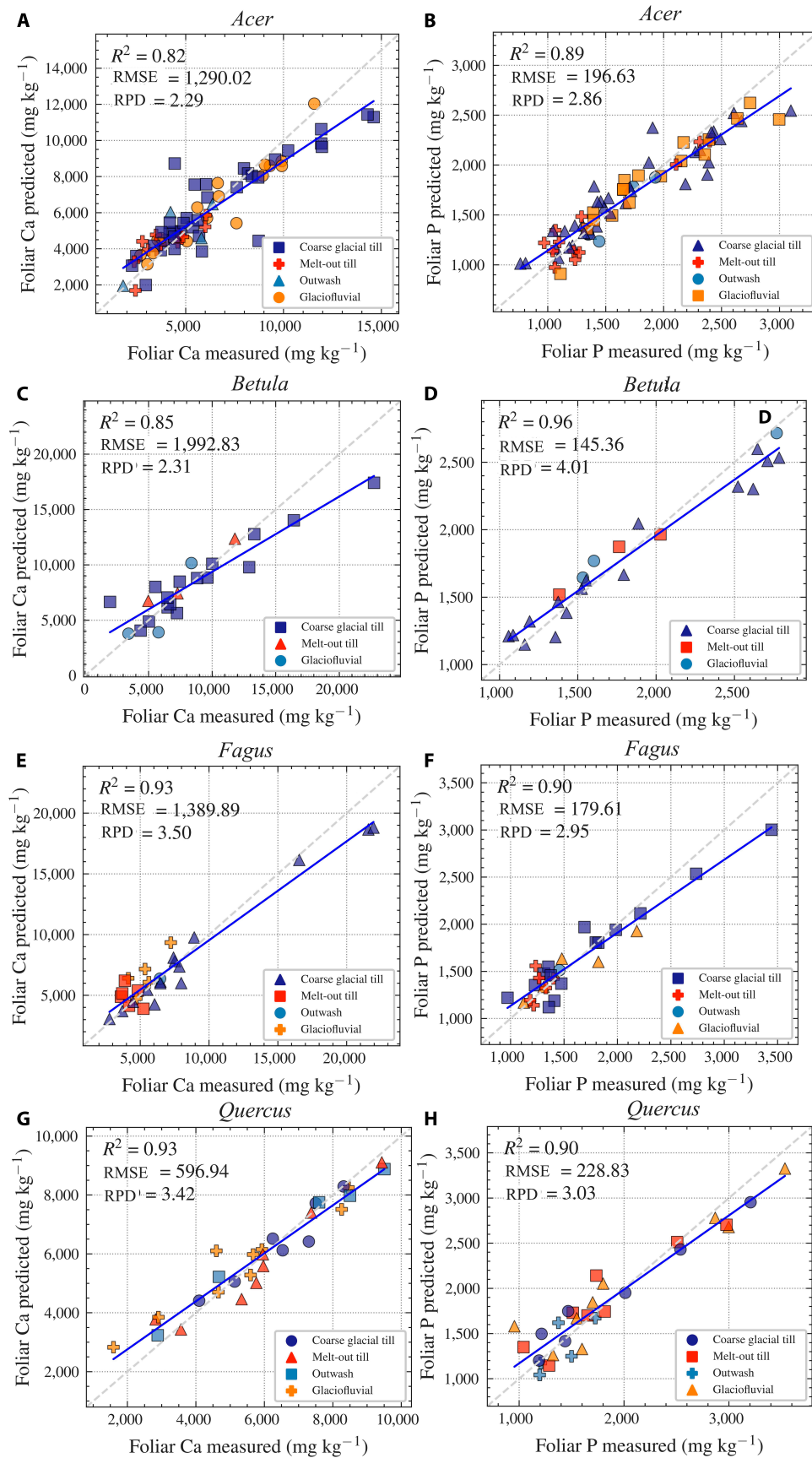
**Fig. 8.** PLSR model results for Ca and P by groups of tree genus (A) *Acer* for Ca, (B) *Acer* for P, (C) *Betula* for Ca, (D) *Betula* for P, (E) *Fagus* for Ca, (F) *Fagus* for P, (G) *Quercus* for Ca, and (H) *Quercus* for P.

**Table 5.** PLSR model results for Ca and P by groups of tree genera

|  |  | Acer | Betula | Fagus | Quercus |
|---|---|---|---|---|---|
| $R^2$ | Ca | 0.82 | 0.85 | 0.93 | 0.93 |
|  | P | 0.89 | 0.96 | 0.90 | 0.90 |
| $R^2_{Adj.}$ | Ca | 0.76 | 0.71 | 0.93 | 0.91 |
|  | P | 0.86 | 0.89 | 0.88 | 0.86 |
| RMSE | Ca | 1,290.02 | 1,992.83 | 1,389.89 | 596.94 |
|  | P | 196.63 | 139.82 | 179.61 | 228.83 |
| RPD | Ca | 2.29 | 2.31 | 3.50 | 3.42 |
|  | P | 2.86 | 4.17 | 2.95 | 3.03 |
| No. of wavelengths | Ca | 343 | 212 | 173 | 117 |
|  | P | 436 | 236 | 226 | 187 |
| No. of PLS components | Ca | 26 | 12 | 4 | 7 |
|  | P | 21 | 14 | 6 | 8 |
| No. of samples | Ca | 70 | 23 | 27 | 29 |
|  | P | 70 | 23 | 27 | 29 |

is a more important factor affecting foliar macronutrients than geologic material. The individual tree genus models for foliar P showed a much higher level of accuracy (average adj. $R^2 = 0.87$) compared to the combined group (adj. $R^2 = 0.66$). The individual genus models for Ca also improved over the combined group from average adj. $R^2$ 0.75 to 0.83. For PLSR foliar Ca models, *Fagus* and *Quercus* had higher accuracy (adj. $R^2 = 0.93$ for *Fagus*; adj. $R^2 = 0.91$ for *Quercus*) compared to *Acer* and *Betula* (adj. $R^2 = 0.76$ for *Acer*; adj. $R^2 = 0.71$ for *Betula*). For PLSR foliar P models, all individual genus predictions had similar accuracy. The number of selected wavelengths used to estimate Ca and P concentrations across tree genera ranged between 117 and 436 and the number of PLS components ranged between 6 and 26 (Table 5), which are relatively higher than separating by geologic materials (number of selected wavelengths: 121 to 345, PLS components: 2 to 6). These results suggest that tree genus is the more important factor affecting foliar nutrients than geologic material because more significant wavelengths correlated with foliar nutrient had been selected and more uncorrelated PLS components could be used by the PLSR model.

There are several possible explanations for genus-specifc models improving the accuracy of quantifying nutrients by PLSR. First, the control of nutrient acquisition and transport from roots to leaves are often specific to a genus, species, or variety and are physiologically and genetically controlled. The effects of absolute and relative variability in leaf pigment compounds are also diminished when each genus is analyzed separately, which follows other methods of forest mensuration, e.g., aboveground biomass estimation by allometric equations [55]. As shown in previous studies, trees in different genera acquire different nutrient concentrations, despite similar nutrients within the geologic materials present, with nutrient concentrations higher in *Acer* and *Betula* and lower in *Fagus* [26,56]. Second, nutrient-poor soils may limit the establishment of nutrient-

dependent trees such as *Acer* [57] and promote trees adapted to nutrient-poor soils such as *F. grandifolia* [56], which can diminish the variability of observable nutrient concentrations. Various drivers can affect the relationship between foliar nutrients and leaf reflectance. For example, the characteristics of the leaf surface, as well as the leaf's internal structure, vary across different genera. Leaf surface features such as waxes and hairs significantly influence reflectance by altering the leaf's optical properties [58]. Similarly, cell wall molecules like cellulose and lignin affect the absorption of SWIR radiation, which is crucial for determining leaf properties [49,50]. Consequently, these variations may cause discrepancies in reflectance measurements, which, in turn, impacts the accuracy of nutrient predictions across genera.

There are a few limitations to our current study that should be noted. Our region is geographically focused only on southern New England, utilizing common tree genera, and mid-growing season leaves. Future studies will need to consider differences throughout the growing season, link spectra with physiological and biochemical changes, and leverage species-specific differences to determine which are more reliable indicators of soil deficiencies. This study focused on predicting foliar nutrient concentrations based on spectral reflectance at the leaf level measured in the laboratory. The research design was crafted to identify the spectral characteristics of nutrients while minimizing other potential interference that could be introduced. When upscaling these results to the canopy scale, the model requires further development, considering factors such as leaf density, leaf structure and light propagation within the canopy.

## Conclusion

The developed PLSR model predicted plant nutrients with moderate to strong accuracy for macro- and micronutrients in temperate hardwood forests of New England. This method holds promise for its expanded use in adjacent and other forested regions and can decrease the costs to estimate nutrients across larger areas. Foliar macronutrient concentrations were more accurately estimated compared to micronutrient concentrations, most likely due to their greater importance for tree mineral nutrition and health. The relationship between foliar nutrients and foliar spectra varied over the geologic materials and tree genera and decreases the strength of the connection between foliar spectra and foliar nutrients. Geologic materials are not commonly used in assessments of foliar nutritional status, and our results suggest that it can be an important factor to consider. However, separating trees by genus proved to be more significant for model accuracy and thus will require large spectral assessments for each genus. Spectral measurements combined with wavelength selection and PLSR models can be used to quantify foliar macro- and micronutrients at regional scales and can be further improved by incorporating geologic materials and tree genera.

## Acknowledgments

**Author contributions:** W.T., Q.Y., and J.B.R. designed the study and led the writing of this paper. I.C.M. and W.T. collected field data, and conducted spectral measurement and sample laboratory measurement under the guidance of J.B.R.

and Q.Y. W.T. conducted the spectral data analysis and modeling under the guidance of Q.Y. and J.B.R. W.T., Q.Y., J.B.R., and A.M.R. interpreted the field and laboratory data and modeling results.

**Competing interests:** The authors declare that they have no competing interests.

## Data Availability

The data that support the findings of this study are available on Scholarworks at University of Massachusetts-Amherst at https://doi.org/10.7275/nw9t-vd56 [59].

## References

1. Alexander SA, Palmer CJ. Forest health monitoring in the United States: First four years. *Environ Monit Assess*. 1999;55(2):267.

2. Woodall CW, Amacher MC, Bechtold WA, Coulston JW, Jovan S, Perry CH, Randolph KDC, Schulz BK, Smith GC, Tkacz B, et al. Status and future of the forest health indicators program of the USA. *Environ Monit Assess*. 2011;177(1):419–436.

3. Foster DR, Motzkin G, Slater B. Land-use history as long-term broad-scale disturbance: Regional forest dynamics in Central New England. *Ecosystems*. 1998;1(1):96–119.

4. Duveneck MJ, Thompson JR, Gustafson EJ, Liang Y, de Bruijn AM. Recovery dynamics and climate change effects to future New England forests. *Landsc Ecol*. 2017;32(7):1385–1397.

5. Dauer JM, Chorover J, Chadwick OA, Oleksyn J, Tjoelker MG, Hobbie SE, Reich PB, Eissenstat DM. Controls over leaf and litter calcium concentrations among temperate trees. *Biogeochemistry*. 2007;86(2):175–187.

6. Sparks DL, Page AL, Helmke PA, Loeppert RH. *Methods of soil analysis, part 3: Chemical methods*. John Wiley & Sons; 2020.

7. George E, Marschner H, Jakobsen I. Role of arbuscular mycorrhizal fungi in uptake of phosphorus and nitrogen from soil. *Crit Rev Biotechnol*. 1995;15(3–4):257–270.

8. Chadwick KD, Brodrick PG, Grant K, Goulden T, Henderson A, Falco N, Wainwright H, Williams KH, Bill M, Breckheimer I, et al. Integrating airborne remote sensing and field campaigns for ecology and earth system science. *Methods Ecol Evol*. 2020;11(11):1492–1508.

9. Adjorlolo C, Mutanga O, Cho MA. Predicting C3 and C4 grass nutrient variability using in situ canopy reflectance and partial least squares regression. *Int J Remote Sens*. 2015;36(6):1743–1761.

10. Inoue Y, Sakaiya E, Zhu Y, Takahashi W. Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. *Remote Sens Environ*. 2012;126:210–221.

11. Ferwerda JG, Skidmore AK. Can nutrient status of four woody plant species be predicted using field spectrometry? *ISPRS J Photogramm Remote Sens*. 2007;62(6):406–414.

12. Blackburn GA. Hyperspectral remote sensing of plant pigments. *J Exp Bot*. 2007;58(4):855–867.

13. Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics. *Chemom Intell Lab Syst*. 2001;58(2):109–130.

14. Abdel-Rahman EM, Mutanga O, Adam E, Ismail R. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS J Photogramm Remote Sens*. 2014;88:48–59.

15. Mouazen A, Kuang B, De Baerdemaeker J, Ramon H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*. 2010;158(1–2):23–31.

16. Nawar S, Buddenbaum H, Hill J, Kozak J, Mouazen AM. Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res*. 2016;155:510–522.

17. Lu J, Yang T, Su X, Qi H, Yao X, Cheng T, Zhu Y, Cao W, Tian Y. Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves. *Precis Agric*. 2019;21(2):324–348.

18. Ramoelo A, Skidmore AK, Cho MA, Mathieu R, Heitkönig IMA, Dudeni-Tlhone N, Schlerf M, Prins HHT. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. *ISPRS J Photogramm Remote Sens*. 2013;82:27–40.

19. Malmir M, Tahmasbian I, Xu Z, Farrar MB, Bai SH. Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. *J Soils Sediments*. 2019;20(1):249–259.

20. Osco LP, Ramos APM, Pinheiro MMF, Moriya EAS, Imai NN, Estrabis N, Lanczyk F, de Araujo FF, Liesenberg V, de Castro Jorge LA, et al. A machine learning framework to predict nutrient content in Valencia-Orange leaf hyperspectral measurements. *Remote Sens*. 2020;12(6):906.

21. Gao J, Meng B, Liang T, Feng Q, Ge J, Yin J, Wu C, Cui X, Hou M, Liu J, et al. Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of Tibetan plateau China. *ISPRS J Photogramm Remote Sens*. 2019;147:104–117.

22. Wold S. Chemometrics; what do we mean with it, and what do we want from it? *Chemom Intell Lab Syst*. 1995;30(1):109–115.

23. Huemmrich KF, Campbell P, Landis D, Middleton E. Developing a common globally applicable method for optical remote sensing of ecosystem light use efficiency. *Remote Sens Environ*. 2019;230:Article 111190.

24. Boulesteix A-L, Strimmer K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform*. 2007;8(1):32–44.

25. Martens H, Martens M. *Multivariate analysis of quality: An introduction*. Chichester (England): John Wiley & Sons; 2001.

26. Richardson JB, Friedland AJ. Influence of coniferous and deciduous vegetation on major and trace metals in forests of northern New England USA. *Plant Soil*. 2016;402(1–2):363–378.

27. Soller DR, Reheis MC, Garrity CP, Van Sistine DR. Map database for surficial materials in the conterminous United States. *US Geol Surv Data Ser*. 2009;425.

28. Marek RS, Richardson JB. Investigating surficial geologic controls on soil properties, inorganic nutrient uptake, and northern hardwood growth in Western Massachusetts, USA. *J Soil Sci Plant Nutr*. 2020;20:901–911.

29. Rechcigl J, Payne G. Comparison of a microwave digestion system to other digestion methods for plant tissue analysis. *Commun Soil Sci Plant Anal*. 1990;21(19–20):2209–2218.

30. Malvern Panalytical Ltd. FieldSpec 4 User Guide. 7 Jun 2016. https://www.malvernpanalytical.com/en/learn/knowledge-center/user-manuals/fieldspec-4-user-guide

31. Yang J, Du L, Gong W, Shi S, Sun J, Chen B. Analyzing the performance of the first-derivative fluorescence spectrum for estimating leaf nitrogen concentration. *Opt Express*. 2019;27(4):3978–3990.

32. Yang J, Cheng Y, du L, Gong W, Shi S, Sun J, Chen B. Selection of the optimal bands of first-derivative fluorescence characteristics for leaf nitrogen concentration estimation. *Appl Opt*. 2019;58(21):5720–5727.

33. Zhou W, Zhang J, Zou M, Liu X, du X, Wang Q, Liu Y, Liu Y, Li J. Prediction of cadmium concentration in brown rice before harvest by hyperspectral remote sensing. *Environ Sci Pollut Res Int*. 2019;26(2):1848–1856.

34. Shi T, Chen Y, Liu Y, Wu G. Visible and near-infrared reflectance spectroscopy-an alternative for monitoring soil contamination by heavy metals. *J Hazard Mater*. 2014;265:166–176.

35. Pedregosa F, Varaquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: Machine learning in python. *J Mach Learn Res*. 2011;12(85):2825–2830.

36. Mehmood T, Liland KH, Snipen L, Sæbø S. A review of variable selection methods in partial least squares regression. *Chemom Intell Lab Syst*. 2012;118:62–69.

37. Saeys W, Mouazen AM, Ramon H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosyst Eng*. 2005;91(4):393–402.

38. Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger J-M, McBratney A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal Chem*. 2010;29(9):1073–1081.

39. Hallett RA, Hornbeck JW. Foliar and soil nutrient relationships in red oak and white pine forests. *Can J For Res*. 1997;27(8):1233–1244.

40. Rice A, Richardson J. Weathering rates of rocks and soils: Determining base cation release to northern hardwood forests across lithologic and pedogenic gradients. Paper presented at: ASA, CSSA, SSSA International Annual Meeting; 2022; St Louis, MO.

41. Dmuchowski W, Gozdowski D, Brągoszewska P, Baczewska AH, Suwara I. Phytoremediation of zinc contaminated soils using silver birch (*Betula pendula* Roth). *Ecol Eng*. 2014;71:32–35.

42. Brand JJ, Becker DW. Evidence for direct roles of calcium in photosynthesis. *J Bioenerg Biomembr*. 1984;16(4):239–249.

43. Najafpour M, Moghaddam AN, Dau H, Zaharieva I. Fragments of layered manganese oxide are the real water oxidation catalyst after transformation of molecular precursor on clay. *J Am Chem Soc*. 2014;136(20):7245–7248.

44. Hochmal AK, Schulze S, Trompelt K, Hippler M. Calcium-dependent regulation of photosynthesis. *Biochim Biophys Acta*. 2015;1847(9):993–1003.

45. Zahedzadeh F, Kakavand F, Mahna N. Effects of carbohydrate, light, nitrogen and magnesium on in vitro production of anthocyanin in apple. *Int J Biosci*. 2015;6(5):250–260.

46. Xu S, Zhao Y, Wang M, Shi X. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma*. 2018;310:29–43.

47. Oliveira M, Cruz-Tirado J, Roque J, Teófilo R, Barbin D. Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder. *J Food Compos Anal*. 2020;87:Article 103403.

48. Hively WD, Lamb BT, Daughtry CST, Serbin G, Dennison P, Kokaly RF, Wu Z, Masek JG. Evaluation of SWIR crop residue bands for the Landsat next mission. *Remote Sens*. 2021;13(18):18.

49. Jacquemoud S, Baret F. Prospect—A model of leaf optical-properties spectra. *Remote Sens Environ*. 1990;34(2):75–91.

50. Feret J-B, François C, Asner GP, Gitelson AA, Martin RE, Bidel LPR, Ustin SL, le Maire G, Jacquemoud S. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens Environ*. 2008;112(6):3030–3043.

51. Kirkby EA, Pilbeam DJ. Calcium as a plant nutrient. *Plant Cell Environ*. 1984;7(6):397–405.

52. Sivasankaramoorthy S. Effect of supplementary calcium enhances plant growth, photosynthetic pigments and uptake of nutrient in *Oryza sativa* L. under NaCl stress. *Int J Chem Life Sci*. 2013;2(7):1189–1192.

53. Millaleo R, Reyes-Diaz M, Ivanov AG, Mora ML, Alberdi M. Manganese as essential and toxic element for plants: Transport. *J Soil Sci Plant Nutr*. 2010;10(4):470–481.

54. Kaur H, Garg N. Zinc toxicity in plants: A review. *Planta*. 2021;253(6):129.

55. Teck RM, Hilt DE, Hilt DE. *Individual tree-diameter growth model for the Northeastern United States* USDA; 1991.

56. Park BB, Yanai RD. Nutrient concentrations in roots, leaves and wood of seedling and mature sugar maple and American beech at two contrasting sites. *For Ecol Manag*. 2009;258(7):1153–1160.

57. Duchesne L, Ouimet R. Present-day expansion of American beech in northeastern hardwood forests: Does soil base status matter? *Can J For Res*. 2009;39(12):2273–2282.

58. Peters RD, Noble SD. Characterization of leaf surface phenotypes based on light interaction. *Plant Methods*. 2023;19(1):26.

59. Teng W, Yu Q, Mischenko I, Rice A, Richardson J. New England foliar spectral reflectance and nutrient concentrations data. *Data and Datasets*. 2022;152.