Yield modeling of snap bean based on hyperspectral sensing: a greenhouse study

Amirhossein Hassanzadeh,[a,*] Jan van Aardt,[a] Sean Patrick Murphy,[b] and Sarah Jane Pethybridge[b]

[a]Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science, Rochester, New York, United States

[b]Cornell University, Cornell AgriTech at The New York State Agricultural Experiment Station, School of Integrative Plant Sciences, Plant Pathology and Plant-Microbe Biology Section, Geneva, New York, United States

Abstract. Farmers and growers typically use approaches based on the crop environment and local meteorology, many of which are labor-intensive, to predict crop yield. These approaches have found broad acceptance but lack real-time and physiological feedback for near-daily management purposes. This is true for broad-acre crops, such as snap bean, which is valued at hundreds of millions of dollars in the annual agricultural market. We aim to investigate the relationships between snap bean yield and plant spectral and biophysical information, collected using a hyperspectral spectroradiometer (400 to 2500 nm). The experiment focused on 48 single snap bean plants (cv. Huntington) in a controlled greenhouse environment during the growth period (69 days). We used applicable accuracy and precision metrics from partial least squares regression and cross-validation methods to evaluate the predictive ability of two harvest stages, namely an early-harvest and late-harvest stage, against our yield indicator (bean pod weight). Four different spectral data sets were used to investigate whether such oversampled, hyperspectral data sets could accurately and precisely model observed variability in yield, in terms of the coefficient of determination (R^2) and root-mean-square error (RMSE). The objective of our approach hinges on the philosophy that selected spectral bands from this study, i.e., those that best explain yield variability, can be downsampled from a hyperspectral system for use in a more cost-effective, operational multispectral sensor. Our results suggested the optimal period for spectral evaluation of snap bean yield is 20 to 25 or 32 days prior to harvest for the early- and late-harvest stages, respectively, with the best model performing at a low RMSE (3.02 g plant^{-1}) and a high coefficient of determination ($R^2 = 0.72$). An unmanned aerial systems-mounted, affordable, and wavelength-programmable multispectral imager, with bands corresponding to those identified, could provide a near real-time and reliable yield estimate prior to harvest.
 [DOI: 10.1117/1.JRS.14.024519]

Keywords: precision agriculture; yield prediction; snap bean; hyperspectral; partial least squares regression.

Paper 200160 received Feb. 27, 2020; accepted for publication May 25, 2020; published online Jun. 5, 2020.

1 Introduction

It is forecasted that the global population growth will approach ~24% by the year 2050.[1,2] This notion can profoundly alter social structures, from a society's economic growth, to public health and our interaction with the environment; changes that could diminish employment opportunities, cause agricultural decline, and result in a continuous need for more productive and efficient farms.[3,4] As a result, our ability to use the available croplands, while optimizing yield, will become a requirement for sustainable use of scarce agricultural resources. There exists a range of agricultural crops, but our focus is on snap bean, one of the vegetables that hold a large market

*Address all correspondence to Amirhossein Hassanzadeh, E-mail: ah7557@rit.edu

share in being the fifth largest vegetable crop nationally in terms of acreage, with 158,920 acres harvested for processing and 71,170 acres harvested for fresh market across the US in 2014, with a combined value of $416 million.[5] Precision agriculture, or site-specific management, is one strategy to help satisfy this need by helping farmers to improve per-acre yield, reduce uncertainty and risk, and optimize the input–output crop yield continuum.[6,7] It is in the context of precision agriculture that remote sensing, a technology that collects spectral and structural information of an object remotely, has garnered attention over the past two decades.[8] One classification of remote sensing systems revolves around the platform, e.g., ground-based, airborne, or space-borne, with the shared goal of nondestructively obtaining information.[9] This information can be acquired from sensing systems that generate color (RGB) images, three-dimensional (3-D) point clouds from ranging systems [e.g., light detection and ranging (LiDAR)], and spectro-radiometers, which collect hyperspectral data, defined as narrow, contiguous spectral bands or wavelengths.[10]

The core science behind "hyperspectral" remote sensing and analysis is spectroscopy, an approach to evaluate the interaction of light with materials, introduced by Norris (1965), and the corresponding effect on the material's molecular and chemical structure, called chemo-metrics.[11] In short, three scenarios can occur when electromagnetic energy strikes a surface: energy can be absorbed, reflected, or transmitted by the surface.[12] Every material thus has a specific reflectivity (ρ), transmissivity (τ), and absorptivity (σ), for which $\tau + \sigma + \rho = 1$ is valid. While each of these scenarios has been studied intensively, the reflected light per wavelength is the most convenient form to examine remotely, given that property's ease of acquisition, inherent informational properties, and the ubiquitous nature of such reflectance-based systems.[13] Most every material is defined by a specific reflectance spectrum, which theoretically varies as the material is altered. This notion becomes crucially important when it comes to living organisms, e.g., crops, which can exhibit spectral variability due to species, physiological state, disease, water stress, growth stage, etc. For example, absorption in the visible region (400 to 700 nm), red edge peak (680 to 760 nm), and reflectance in the near-infrared (NIR) domain (700 to 1000 nm) are primary indicators of chlorophyll absorption, chlorophyll density, intercellular leaf structure, and even the general health of a crop, and serve as a gauge for how well-developed, vigorous, and even dense the plant material is.[14–16] It is in this context that hyper-spectral imaging systems and point-based spectroradiometers have proven valuable, given their ability to measure high spectral resolution from plant surfaces.

Spectroradiometers, either imaging or nonimaging (i.e., point-based), represent one class of these hyperspectral sensing systems. Hyperspectral imagers typically collect hundreds of spectral bands, with associated spatial context information (pixel coordinates), often by moving a spectral slit across the subject of interest.[17] Spectroradiometers (nonimaging), on the other hand, obtain an average spectrum within the (fore-)optic's field-of-view (FOV). Hyperspectral devices have been intensively used for a variety of crop-related studies, e.g., harvest maturity quantification,[18–27] stress and disease detection,[28–35] genus/species classification,[35–41] and even yield prediction.[42–50] We will focus our discussion on the latter, namely yield forecasting or modeling, given its importance to either enable within-season intervention or harvest planning and logistics.

The utility of remote sensing data (e.g., vegetative indices; Ref. 51) to predict agronomic crop data (e.g., yield) is often quantified using empirical univariate or multivariate modeling. For example, Lai et al.[52] evaluated wheat yield as a linear function of time-integrated normalized difference vegetation index (iNDVI) in 17 farms over 15 years in Australia, using leave-one-out cross-validation. Their model was able to predict yield with a low-average RMSE (0.79 Mg ha^{-1}). Becker-Reshef et al.,[53] in turn, assessed the ability of an empirical model which could estimate yield with a 7% error in winter wheat crop, using MODIS-derived NDVI data from Kansas. Their model was able to forecast yield 6 weeks prior to harvest within a 10% error using MODIS NDVI data in Ukraine. The advantage of empirical modeling is that only a limited number of parameters are required,[54] while pitfalls to this approach include: (i) performance is constrained to the size of the data set;[55] (ii) it is assumed that photosynthetic activity is the primary estimator of the yield, which could result in inaccurate estimations;[53] (iii) these NDVI-based models result in erroneous estimates for crops with high biomass, given that NDVI saturates with high leaf area index;[56] and (iv) this approach is an estimator and arguably not a true forecasting technique.[57]

Statistical approaches, as another technique for yield prediction, are defined here as being more multivariate in nature, and we will focus on partial least squares regression (PLSR)[58] as one of the more proven modeling techniques. There has been extensive research encompassing the use of PLSR for crop yield assessment. For instance, the study by Wenzhi et al.[59] evaluated sunflower yield estimation using both PLSR and neural networks on 16 independent physical variables. The authors utilized the variable-in-projection approach for PLSR, which shows how sensitive each variable is concerning the output variable.[60] It was shown that neural networks can capture more complex relationships and slightly outperform PLSR with a high coefficient of determination ($R^2 = 0.86$, compared to $R^2 = 0.77$). A study of Vásquez et al.[61] used hyperspectral data (handheld spectroradiometer: 350 to 1050 nm) for paddy field yield estimation via a PLSR approach. The study was cross-validated over six different cultivars, for three different growth stages, and replicated three times. Their results showed that the booting stage exhibits a more accurate estimation than other tests ($R^2 = 0.87$). It was shown that PLSR can handle data non-normality, with drawbacks such as neglecting variables that may be considered discriminating, as well as being susceptible to the scale of data. We opted to evaluate a PLSR approach to yield forecasting in snap bean, our proxy crop, and did so in a controlled greenhouse environment, before attempting to scale our results to the field-level using unmanned aerial systems (UAS)-based imaging. We hypothesized that snap bean yield can be accurately predicted early in the season via remote sensing techniques, because crop yield has been shown to be a function of time, spectral, and biophysical attributes, among others.[50,62–65]

The objectives of this study were to: (i) determine the applicability of PLSR for yield assessment of snap bean; (ii) assess the most accurate time to perform snap bean yield estimation, using spectral and plant biophysical data; and (iii) identify spectral regions or features and plant biophysical attributes that result in the most accurate/precise yield predictions for snap bean.

2 Methods

2.1 *Study Area*

Hyperspectral data were collected in an on-campus greenhouse, located at Rochester Institute of Technology (43.0846° N, 77.6743° W), during the winter and spring of 2018/2019. One hundred forty snap bean seeds (cv. Huntington) were planted on March 6, 2019, in a seed raising mix ("Potting Mix," Miracle-Gro, composed of peat, peat moss, perlite, compost, and 0.48% fertilizer), after which 48 plants were selected and transplanted to 6-in.-diameter pots. Plants were irrigated with 500 ml of water every 2 days thereafter.

2.2 *Plant Growth Characteristics*

The 48 plants were divided into two groups of 24 plants, where each plant was considered a single experimental unit. Harvest dates were determined based on when 50% of plants for early harvest, and 70% of plants for late harvest, showed 50% of their pods being at industry sieve size 4 to 6 (see Table 1). On each harvest date, total pod weight of each plant was measured using a balance ("1500" model, VWR, Radnor, Pennsylvania) with a precision of 0.05 g. Sample yield values ranged 16.65 to 38.35 g and 20.20 to 40.10 g for early-harvest and late-harvest stages, respectively.

2.3 *Data Collection*

The spectral data set was captured using a spectroradiometer ("HR-1024i" model, Spectra Vista Corporation, Poughkeepsie, New York), with a 14-deg FOV fore-optic. The spectroradiometer captures spectral information across the visible, NIR, and shortwave-infrared (SWIR) wavelengths (335 to 2500 nm), for 987 contiguous spectral channels. The fore-optic enabled us to accurately target each plant and resulted in a ~6-cm-diameter sampling size from a distance of 30 cm. This spectroradiometer consists of silicon (350 to 1000 nm), indium gallium arsenide (InGaAs; 1000 to 1890 nm), and extended InGaAs (1890 to 2500 nm) detectors. Select

Table 1 The growth calendar of snap bean trials.

Date	Stage	DAP[a]	DPH[b] (early/late)
March 6th	Planting	0	65/69
March 6th to April 3rd	Vegetative growth	0 to 28	65 to 37/69 to 41
April 3rd to April 15th	Budding	28 to 40	37 to 25/41 to 29
April 15th to April 17th	Flowering	40 to 42	25 to 23/29 to 27
April 17th to end	Pod formation	42 to maturity	23 to 0/27 to 0
May 10th	Early harvest	65	—
May 14th	Late harvest	69	—

[a]DAP, days after planting.
[b]DPH, days prior to harvest.

wavelength ranges were omitted from the sampled spectra prior to analysis: (i) range 335 to 480 nm, due to sensor fall-off noise; (ii) the 850- to 1000-nm range, because of artifacts due to detector overlap between the Si and InGaAs sampled regions; (iii) the 1900- to 2000-nm range, due to a similar overlap artifact between the InGaAs and extended InGaAs detectors; and (iv) the region between 2400 and 2500 nm, again due to low SNR (detector fall-off) in the far SWIR region. This selective data cleaning resulted in 678 remaining spectral bands. The spectroradiometer was mounted on a stationary stand and was covered with black felt to inhibit external light from entering the sampling chamber. Two tungsten–halogen lamps were installed inside the chamber to produce stable and consistent light for measurements, shown in Fig. 1. All samples were calibrated to reflectance using a Spectralon calibration panel as the reference target, shown in Fig. 1; this was done to account for potential illumination differences or sensor drift between samples, either within- or across sample days.[66] Finally, to minimize the influence of potential "mixed pixels," i.e., where sample background (potting mix) could influence the signal collected for each target, a collar-shaped nonreflective black felt was used to cover the potting mix (see Fig. 1); this approach enabled the capture of a pure vegetative spectrum for each sample (one spectral sample per plant) at the canopy-level, 15 to 20 cm away from the canopy top, resulting in a 3- to 4-cm sampling diameter. Averaging of canopy-level reflectance resulted in a closer approximation of what a UAS-based spectrometer would sample, even if across canopy scale.

Spectral and biophysical measurements, such as the number of leaves, plant height, and canopy width, were collected three times a week and then daily for the last 10 days of the growth

Fig. 1 (a) Black felt was used to reduce spectral mixture effects, where potential potting mix background may erroneously be included on the plant spectrum; and (b) Spectralon calibration panel for reference measurements, used for calibrating radiance data to reflectance and thus normalizing for any illumination variability that may have occurred between samples and sample days.

season, throughout the 69-day growing period. The number of leaves per plant was recorded, and plant height and canopy width were measured using a ruler. We normalized the collected spectral and biophysical attributes to the scale of 0 to 1, ensuring comparable feature influence on the final analysis.

2.4 *Data Analysis*

The spectral data were subjected to several preprocessing steps before being used in yield prediction algorithms. First, the data set was tested for multivariate normality using Small's and Chi-squared approaches to evaluate the data distribution.[67] Both approaches utilize the squared Mahalanobis distance between each sample vector and the mean feature vector for the data set $X_{M\times N}$ (M number of samples and N number of features) as

$$D_i^2 = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}), \tag{1}$$

where $i = 1, \ldots, m$, and S^{-1} is the inverse covariance matrix of random variable X, along the feature axis. The difference between these two approaches is that the Chi-squared method compares the squared Mahalanobis distance to percentiles from the Chi-squared distribution, which is limited by a requirement of large sample sizes for a large number of features, while Small's method estimates this by calculating u values, and compares sorted u values with percentiles from a beta distribution, as

$$u_i = \frac{m d_i^2}{(m-1)^2}. \tag{2}$$

The next preprocessing step involved spectral or mathematical enhancement (transformation) of spectral data, including smoothing, first derivative calculations, and absorption feature enhancement. Spectral smoothing was performed using a Savitzky–Golay filter with a window size of 11 bands, fitted with a second-order polynomial.[68] This configuration was chosen based on a preliminary sensitivity analysis using the raw data.[3] First derivatives were calculated, via a Savitzky–Golay filter with the second-order polynomial fit over 11 bands, in order to evaluate not only spectral magnitude but also curve shape, as well as the baseline offset reduction due to scattering.[25] Absorption feature enhancement used a continuum-removal (CR) approach for spectral regions 555 to 755 nm (potential chlorophyll absorption), 1120 to 1265 nm (potential water and cellulose absorption), 1275 to 1675 nm (water, sugar, protein, and nitrogen absorption), 1676 to 1825 nm (lignin, cellulose, and sugar absorption), and 1900 to 2397 nm (structural signatures related to cellulose and lignin).[69] Consequently, these spectral data sets (viz., raw spectra, smoothed spectra, first derivative spectra, and CR-spectra) were appended with concurrent physical data and used as inputs to the yield prediction analysis.

Normalization using scaling ensures that features with different units of measure have the influence on the analysis.[70] We utilized this approach due to the difference in scales between biophysical and spectral data and scaled all features to the 0 to 1 range.

Finally, the PLSR approach was utilized for yield prediction.[58] PLSR projects the data onto an alternative axis/dimension using latent variables (LV), maximizing the covariance between independent variables and the dependent variable. In PLSR, X (the independent variables) and Y (the dependent variable) can be explained as a sum of LVs as

$$X = TP^T + E = \sum t_{lv} p_{lv}^t + E, \tag{3}$$

$$Y = UQ^T + F = \sum u_{lv} q_{lv}^t + F, \tag{4}$$

where P and Q are the loading factor matrices, T and U are called the score matrices, and E and F are the residual matrices. PLSR also outputs a coefficient matrix, C, that explains the weight of each independent variable in the simple linear regression model, as

$$Y = CX + \epsilon. \tag{5}$$

Finally, the band selection stage for this study can be defined as the weight of each band (matrix C), which can be used to identify features that have more discriminating power over other features. The goal is to be able to identify features that contribute most to the modeling effort and best explain yield as a function of several independent feature variables.[32,71]

For each data set (viz., raw, continuum-removed, smoothed, and first derivative), we evaluated the PLSR-based coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), and RMSE of calibration and leave-one-out cross-validated sets,[72] based on the number of components used (i.e., maximum of five components). Consequently, regression results of the number of components corresponding to highest R^2 values for calibration and leave-one-out cross-validation sets were reported.

Python 3.5[73] was used for all preprocessing and regression tasks. Scikit-learn module[74] was used for PLSR, and author's publicly available codes were used for normality assessment and CR (available at https://github.com/yxoos/MultivariateNormality and https://github.com/yxoos/ContinuumRemoval).

3 Results

The weighted average growth rate of snap bean plants ranged from −0.38 to 1.15 cm and −0.40 to 0.65 cm per day of data collection, for early- and late-harvest stages, respectively. Figure 2 shows snap bean biophysical changes including canopy width, plant height, and the number of leaves, over the growth cycle. As can be seen from the two stages [Figs. 2(a) and 2(b)], the variability in canopy width is substantial, while plant height and number of leaves follow a parabolic trend. Specifically, width measurements exhibited more noise than the other two attributes; this might be due to plants' response to the amount of sunlight they receive and available natural light variability throughout the growth period. Maximum plant height was observed between 45 and 55 days after planting (DAP), or between 15 and 25 days prior to harvest (DPH) in the number of leaves curve, for both stages. After that period, the snap bean plants divert more resources to pod formation and maturation, which results in leaf abscission and chlorosis of the foliage [see Figs. 2(a) and 2(b) for number of leaves curve at DAP > 60].

Results from multivariate normality tests, presented in Fig. 3, show p-values that confirm multivariate normality with only minor indications of non-normality.

3.1 *Early-Harvest Stage*

Figure 4 shows the R^2 results as a function of time, while Table 2 tabulates results (R^2, R^2_{adj}, and RMSE) of the best model for each transformation approach. Each subfigure shows the best-fit value attained with five LV and the maximum R^2 for the number of LV smaller than five. We can see that the fit values comparing the trends from fit values with those of cross-validation.

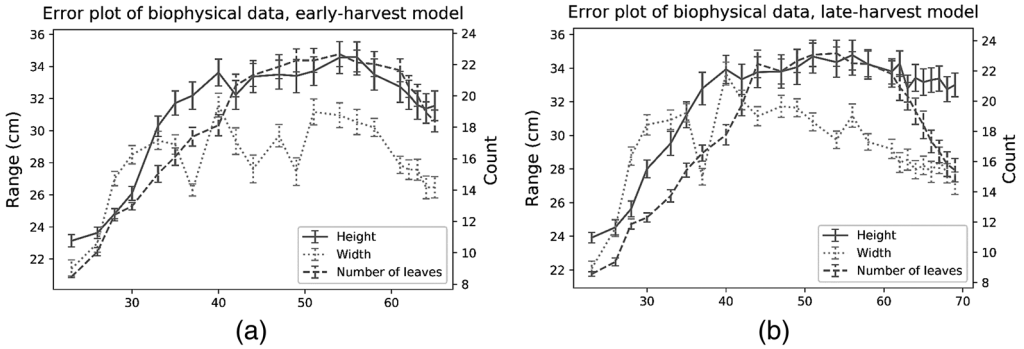

Fig. 2 The average biophysical data with standard error as error bars versus DAP for: (a) the early-harvest stage and (b) the late-harvest stage. Note the maximum number of leaves that occurs 10 to 20 DPH.

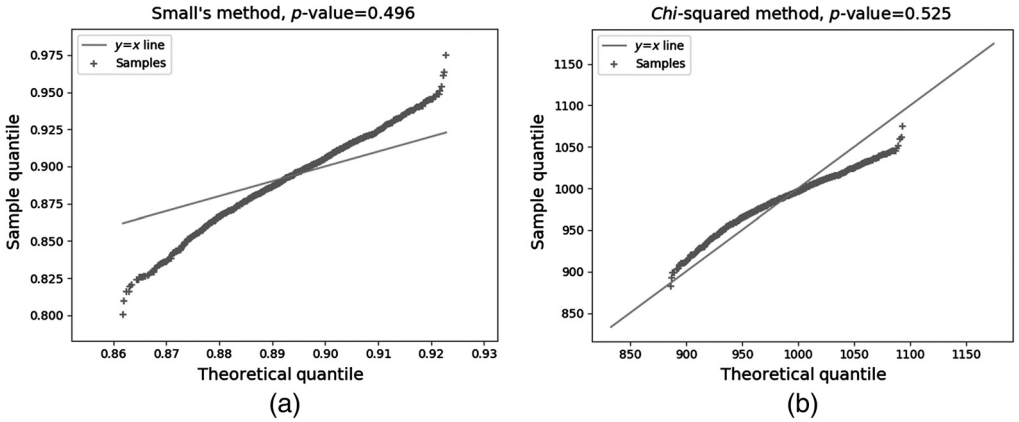

Fig. 3 (a) Small's multivariate normality test. Note that the data follow the non-normal trend to some extent; and (b) the Chi-squared multivariate normality test.

That is why cross-validation plays a crucial role in proposing a generalized model, i.e., it enables a more accurate prediction response when extrapolated to new datasets.[52]

According to Fig. 4, almost all models exhibit local maxima around 33 to 35, 42 to 44, 55 to 57 DAP, and a few around 62 DAP. We can observe a peak around 33 to 35 DAP, which has a lower accuracy when compared to other time frames and only shows superior performance in [Raw-spec. + PHY] and [Smoothed-spec. + PHY], while not being consistent across all transformations (see Table 2). Results from this table also show that raw and smoothed models performed similarly, which are indicative of the relatively low noise level in the acquired data. Moreover, the maximum cross-validation R^2 value at 55 to 57 DAP is considered to be for a period when it is too late to take action, management-wise. The same is valid for 62 DAP, which is at 3 DPH. The maximum cross-validation R^2 value around 42 to 44 DAP (or 23 to 21 DPH) seems to be an excellent temporal assessment period, which arguably is timely in terms of potential management interventions, while also being consistent across transformations in terms of performance. We also observed that the addition of physical attributes increased R^2 (up to 20%) across the growth period, along with increasing R^2 around 62 DAP (or 3 DPH). The model that performed best during that period is [smoothed-spec. + PHY]. However, both CR models exhibited superior performance ($R^2 = 0.40$) and remained as accurate when biophysical predictor features were removed ($R^2 = 0.31$ for [CR-spec.], $R^2 = 0.33$ for [CR-spec. + PHY]). Figure 5 shows the residual and regression plots for the two models that performed best, i.e., [Raw-spec. + PHY] and [CR-spec.], as reported in Table 2. As presented in Table 2, the residual distribution was random for both approaches and did not follow any trend, which confirms that the dependent–independent variable modeling approach properly described the trend or behavior in the variable relationship. Two wavelength regions (1030 to 1075 nm NIR and 1678 to 1725 nm SWIR) were dominant for yield prediction.

3.2 *Late-Harvest Stage*

Results for all four spectral data sets are shown in Fig. 6; this figure shows that several local maxima occur around 37 DAP and 44 to 47 DAP. Correspondingly, the highest accuracy measures are presented in Table 3. It is likely that the 37 DAP exhibits a maximum due to biophysical features, since the performance would deteriorate were it not for these biophysical characteristics. It is also interesting to note that, while significant variability in observed yield was explained at 37 DAP, the R^2 value decreases to 60% of its original value if biophysical data are omitted. However, the [CR-spec.] data do not rely on biophysical data at all and exhibits similar performance to the [Smoothed-spec. + PHY] model, which indicates that this transformation could be more robust.

Multiple wavelength ranges contributed to this analysis in a more general sense, as presented in Table 3. These ranges were 483 to 495 nm, 670 to 680 nm, 748 to 752 nm, 2090 to 2110 nm,

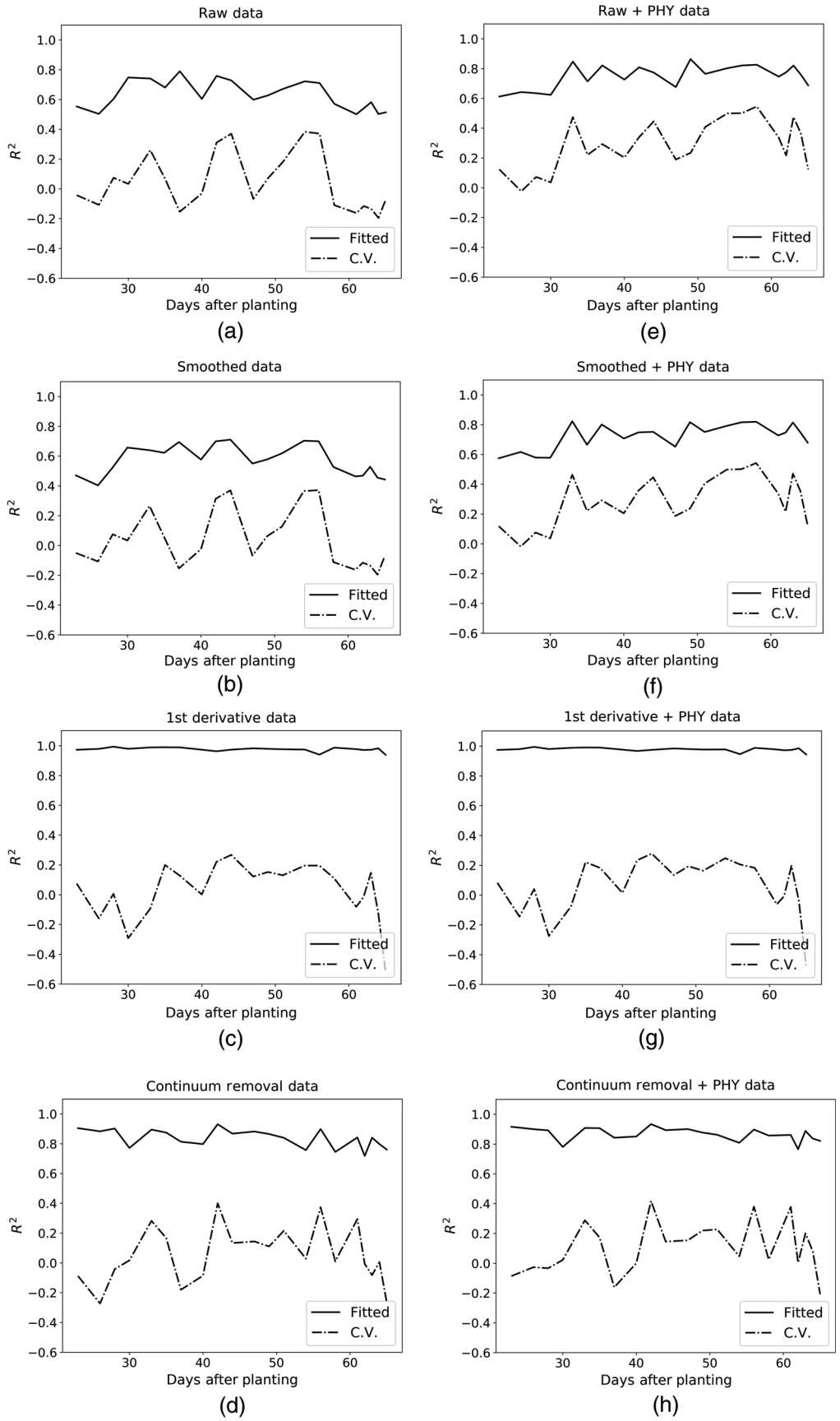

Fig. 4 The early-harvest stage results: (a)–(d) spectral-only results for the raw data, smoothed, first derivative, and CR approaches, (e)–(h) the stage outcomes using for spectral plus biophysical features for the raw data, smoothed, first derivative, and continuum-removed data sets.

Table 2 Regression results for the early-harvest stage, along with spectral and biophysical predictor variables.

Early-harvest stage	R^2	R^2_{adj}	RMSE (g/plant)	LV[a]	DAP[b]/DPH[c]	Top 10 features
Raw – spec.[d]	0.37	0.28	4.62	3	44/21	1066, 1070, 1063, 1036, 1043, 1028, 1074, 1059, 1047, 1032
Raw – spec. + PHY[e]	**0.45**	**0.36**	**4.33**	**3**	**44/21**	**NL[f], H[g], W[h], 1066, 1070, 1063, 1036, 1043, 1028, 1074**
Smoothed – spec.	0.37	0.28	4.62	3	44/21	1066, 1063, 1059, 1070, 1036, 1032, 1040, 1074, 1055, 1043
Smoothed – spec. + PHY	0.45	0.36	4.33	3	44/21	NL, H, W, 1066, 1063, 1059, 1070, 1036, 1032, 1074
1st Deriv.[i] – spec.	0.27	0.23	4.98	1	44/21	1542, 1545, 1590, 1586, 1534, 1549, 1556, 1538, 1582, 1560
1st Deriv. – spec. + PHY	0.28	0.25	4.94	1	44/21	NL, 1542, 1545, 1590, 1586, H, 1534, 1549, 1556, 1538
CR[j] – spec.	**0.40**	**0.31**	**4.51**	**3**	**42/23**	**1707, 1692, 1696, 1718, 1710, 1685, 1677, 1725, 1721, 1699**
CR – spec. + PHY	0.42	0.33	4.44	3	42/23	H, 1707, NL, W, 1692, 1696, 1718, 1710, 1685, 1677

[a]Nu. of LV, number of latent variables.
[b]DAP, days after planting.
[c]DPH, days prior to harvest.
[d]spec., spectral data.
[e]PHY, physical data.
[f]NL, number of leaves.
[g]H, height.
[h]W, weight.
[i]1st Deriv., first derivative of spectral data
[j]CR, continuum-removal
Note: The bold rows indicate the best models in Fig. 5.

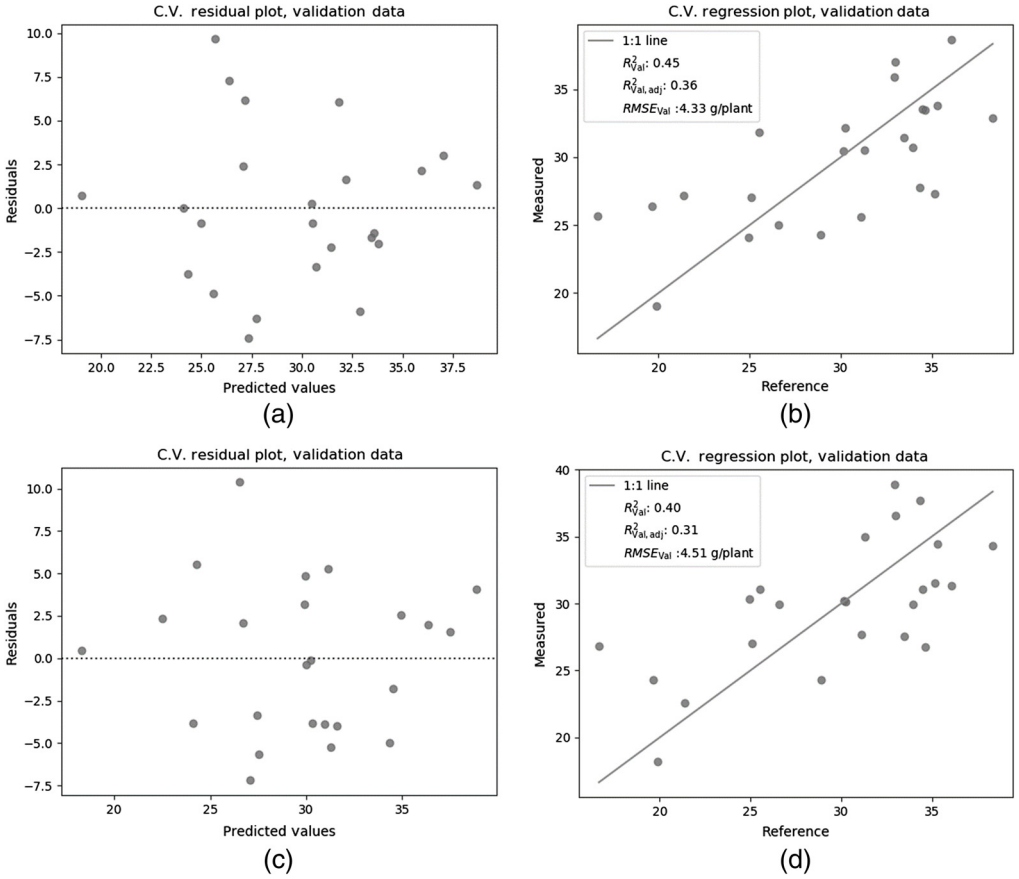

Fig. 5 The early-harvest stage: (a) and (b) residual and regression plots for the cross-validated raw data, with both spectral and biophysical features at 44 DAP/21 DPH. The random nature of the model residuals is evident, which is indicative of a good fit; (c) and (d) the residual and regression plots for the cross-validated continuum-removed data, with spectral-only attributes at 42 DAP/23 DPH. Raw data resulted in the most accurate cross-validated predictions for both spectral and physical features, while the CR exhibited the best results for spectral-only features.

2273 to 2287 nm, and 2380 to 2395 nm. In the next section, the correlations between these spectral features and plant chemical and physiological changes are explained.

We depicted the top two models in Fig. 7; we elected to show the results for both the [CR-spec.] and [Raw-spec. + PHY] models, since the former is not as reliant on biophysical features, and the latter outperformed the other models. As can be seen, findings show high coefficients of determination for the raw and CR approach ($R^2 = 0.67$ and $R^2_{adj} = 0.58$ for raw data, $R^2 = 0.58$ and $R^2_{adj} = 0.63$ for continuum-removed data). Also, raw data exhibited a low RMSE (0.265 g plant^{-1}), 25% lower than the best result in the early-harvest stage, shown in Fig. 7(b). Again, the distribution of residuals has a random nature to them [see Figs. 7(a) and 7(c)], which supports the conclusion that the algorithm or model captured the trend in the dependent variable.

4 Discussion

Yield prediction of snap bean via hyperspectral remote sensing techniques in a greenhouse environment was evaluated and proved satisfactory in terms of the explained variability in the dependent (yield) variable. Our findings show that 20 to 25 and 32 DPH represent the optimum time periods for accurate snap bean yield prediction, with high cross-validated coefficients of determination ($R^2_{val} = 0.40$ to 0.72) and low RMSE (3.02 to 4.51 g plant^{-1}). Moreover, our results demonstrate that the snap bean yield prediction can be improved by the addition of biophysical data (viz., canopy width, height, and the number of leaves) to spectral predictor features.

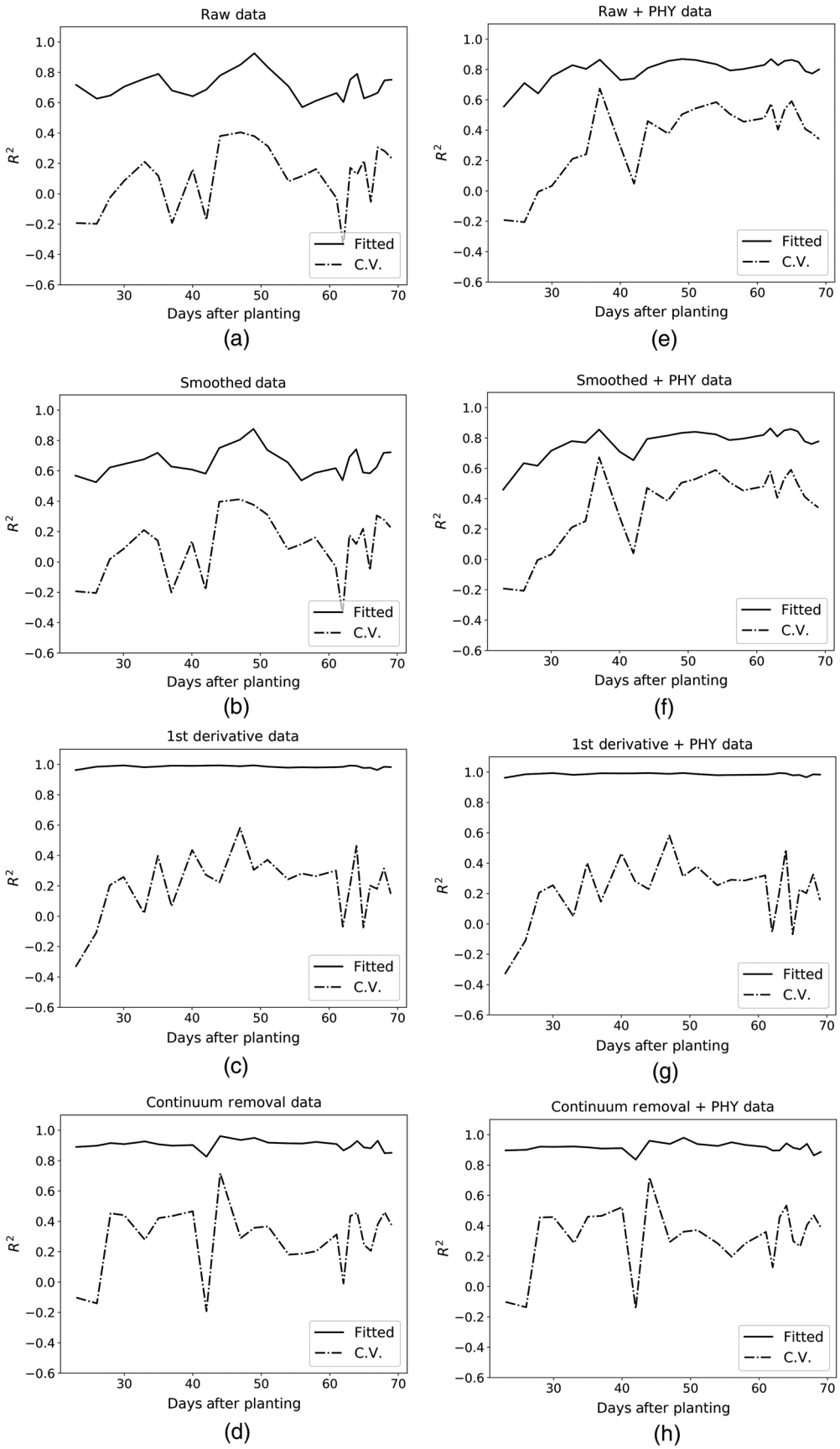

Fig. 6 The late-harvest stage results: (a)–(d) the spectral only results for raw data, smoothed, first derivative, and CR approaches, (e)–(h) spectral plus biophysical features included in the for raw data, smoothed, first derivative, and CR stages.

Table 3 Regression results for the late-harvest stage, along with significant spectral and biophysical predictor variables.

Late-harvest stage	R^2	R^2_{adj}	RMSE (g/plant)	LV[a]	DAP[b]/DPH[c]	Top 10 features
Raw – spec.[d]	0.40	0.27	4.39	4	47/22	483, 485, 677, 675, 672, 680, 490, 669, 494, 676
Raw – spec. + PHY[e]	**0.67**	**0.58**	**3.24**	**5**	**37/32**	**H[f], NL[g], 483, 485, 677, 675, 672, 680, 669, 676**
Smoothed – spec.	0.41	0.28	4.35	4	47/22	483, 484, 677, 675, 676, 679, 673, 485, 672, 680
Smoothed – spec. + PHY	0.67	0.60	3.25	4	37/32	NL, W[h], H, 2282, 2285, 2280, 2278, 2287, 2275, 2273
1st Deriv.[i] – spec.	0.58	0.54	3.68	2	47/22	648, 2089, 638, 637, 2110, 649, 1674, 2092, 2107, 646
1st Deriv. – spec. + PHY	0.58	0.54	3.67	2	47/22	648, 2089, 638, 637, 2110, 649, 1674, 2092, 2107, 646
CR[j] – spec.	**0.72**	**0.63**	**3.02**	**5**	**44/25**	**2394, 752, 2392, 749, 556, 2379, 2381, 2390, 2388, 557**
CR – spec. + PHY	0.72	0.64	3.01	5	44/25	2394, H, 752, 2392, 749, 556, NL, 2379, 2381, 2390

[a]LV, number of latent variables.
[b]DAP, days after planting.
[c]DPH, days prior to harvest.
[d]spec., spectral data.
[e]PHY, physical data.
[f]H, height.
[g]NL, number of leaves.
[h]W, weight.
[i]1st Deriv., First derivative of spectral data.
[j]CR, continuum-removal.
Note: The bold rows indicate the best models in Fig. 7.

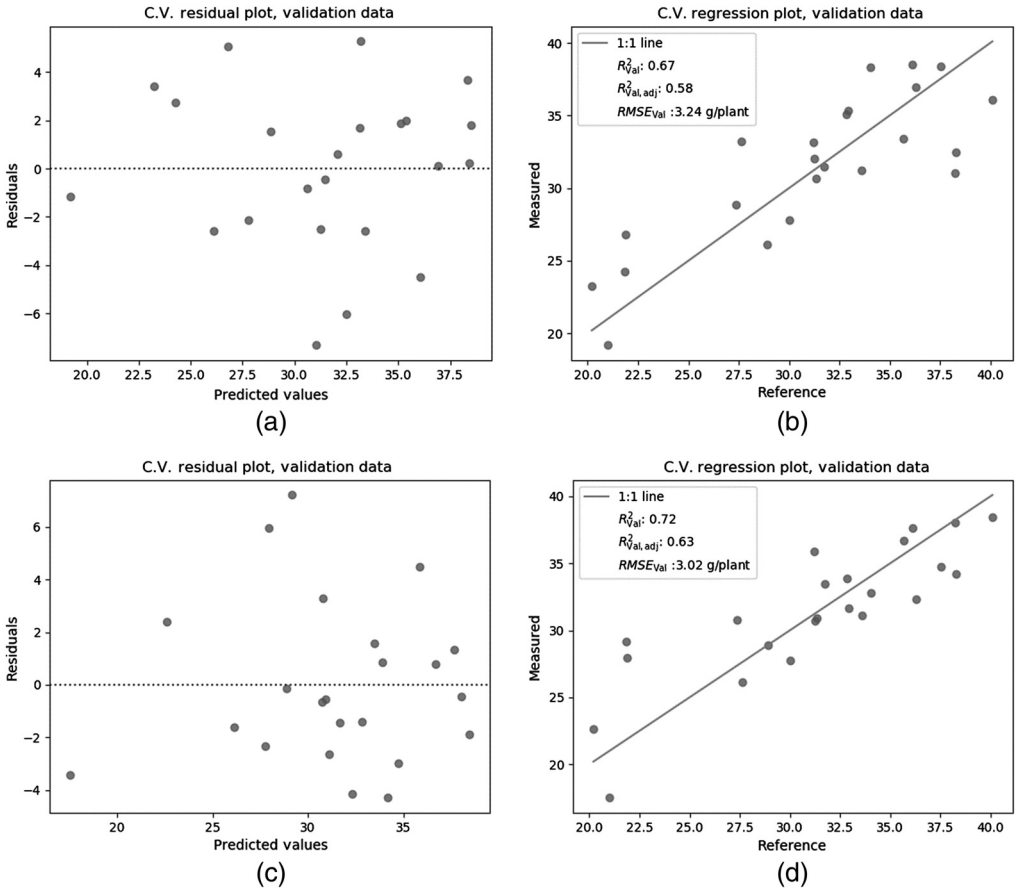

Fig. 7 The late-harvest stage: (a) and (b) the residual and regression plots for the cross-validated raw data, with both spectral and biophysical features at 37 DAP/32 DPH. Note the random nature of the residuals; (c) and (d) the residual and regression plots for the cross-validated continuum-removed data, with spectral-only predictor variables. Raw data exhibited the most accurate cross-validated predictions for both spectral and biophysical features, while the CR yielded the best results for spectral-only features at 44 DAP/25 DPH.

Our yield prediction analysis shows that that continuum-removed data and raw data outperformed other spectral datasets.

The early-harvest stage included spectral features in two wavelength regions, namely the 1030- to 1075-nm NIR and 1678- to 1725-nm SWIR regions. A closer look at these spectral domains and their correlations to plant physiology explains that:

1. The 1030- to 1075-nm region is tied to protein and oil absorption, and their corresponding N─H stretch and C─H stretch/deformation.[75]
2. The 1678- to 1725-nm region, on the other hand, has been shown by several studies to be coupled to lignin and nitrogen absorption.[16,76,77]

Similarly, the late-harvest stage included six wavelength regions of significance to yield prediction, namely 483 to 495 nm (blue), 670 to 680 nm (red), 748 to 752 nm (red edge), and SWIR regions 2090 to 2110 nm, 2273 to 2287 nm, and 2380 to 2395 nm. The corresponding link between the identified spectral domains and plant physiology were scrutinized in the literature:

1. The 483- to 495-nm region suggests strong absorption for chlorophyll-a, and -b.[75] However, this region is also susceptible to noise, due to atmospheric scattering, which results in correspondingly low SNR for many remote sensing sensors, should these results be extended to airborne platforms.[75]

2. The 670- to 680-nm range is responsible for chlorophyll-a absorption and its corresponding electron transition.[75] The mentioned rationale makes sense since, at that stage (early pod formation), the plant accumulates more chlorophyll, as required for photosynthesis and to store energy to turn small pods (i.e., pins) into mature ones. Also, we noticed that as the plant transitions through pod maturity, pod formation is mainly dependent on sunlight. In other words, if enough water and sunlight are available to the plant, pods will remain on the plant, hence higher yield, while pod excision will occur if resources become scarce.[78,79]
3. The 748- to 752-nm region contains the well-known red edge feature,[80] which is tied to the general chlorophyll level in the plant and its associated impact on plant vigor/health.[16]
4. Even though there is a decrease in the SNR ratio in the third detector (extended InGaAs), the range between 2090 and 2110 nm serves as a reliable indicator of starch, cellulose, and nitrogen absorption features, which vibrates O═H, C─O, and C─O─C bands.[75,76]
5. The 2273- to 2287-nm region is mainly governed by the C─H, O─H, and CH_2 stretch for cellulose, sugar, starch, lignin, and nitrogen.[75,77] Note that cellulose and lignin indicate more structural support for pod formation, which mirrors the stage during which the analysis occurred (44 DAP). Sugar and starch absorption are essential factors in pod formation, since almost 72% of snap bean dry weight is composed of these two elements.[81]
6. Finally, limited information exists in the literature regarding the 2380- to 2395-nm region; this lack of information could be due to this region often exhibiting low SNR ratios, due to the SWIR detector fall-off and atmospheric absorption effects.[75] This could suggest an area of further research that focuses on the link between potential structural absorption features in the higher end of SWIR region and plant physiology.

The identified spectral domains and their corresponding link to plant physiology, for both early- and late-harvest stages, are tabulated in Table 4. These identified spectral features/ranges can be used to distill the oversampled hyperspectral system to more operational, cost-effective multispectral sensing solutions. For example, typical UAS-based imaging spectrometers range between $60,000 (NIR) and $170,000 (SWIR), involve complex system calibration steps, require extensive in-flight calibration to reflectance, and have to be flown on expensive airframes, given the payload weight. A multispectral solution, on the other hand, which retains only the above-mentioned spectral features, potentially would be much cheaper, require less preprocessing, and can be housed on more accessible UAS platforms. The mentioned downsampled multispectral solution ideally then would have spectral bands that correspond to the spectral features identified in this study. For example, a prediction aimed at the late-harvest stage would include a drone flight at 25 DPH, with the five narrow-band spectral regions mentioned in Table 4. This effectively could reduce the cost of UAS "yield modeling" platform by an order of

Table 4 Identified spectral domains and their corresponding plant physiological links in the literature.

Harvest stage	Spectral domain (nm)	Plant physiological link	References
Early	1030 to 1075	Protein and oil absorption	75
	1678 to 1725	Lignin and nitrogen absorption	16, 76, and 77
Late	483 to 495	Chlorophyll-a and -b absorption	75
	670 to 680	Chlorophyll-a absorption	75
	748 to 752	Plant general chlorophyll level	16
	2090 to 2110	Starch, cellulose, and nitrogen absorption	75 and 76
	2273 to 2287	Cellulose, sugar, starch, lignin, and nitrogen absorption	75

magnitude, when compared to spectrometers. These results also can be compared to other studies, even though the scales may differ.

For example, crop yield prediction generally falls into two different categories in terms of sensing platform, namely airborne- and satellite-based. Studies focusing on satellite-based crop yield prediction (e.g., Refs. 52, 53, 82 and 83) have proven accurate estimates, but with the drawback of coarse spatial scales, often at no better than the field-level scale. This negates the use of satellite-based assessments for precision agriculture purposes.[84] This is where low-cost, operational airborne crop yield prediction becomes attractive (e.g., Refs. 85–89). However, airborne-level spectral yield assessment has its own drawbacks, namely that these systems (i) are inherently noisy, (ii) can adversely be affected by changing illumination conditions, (iii) are low-cost in general, but may become computationally and financially expensive when collecting high temporal resolution data, and (iv) mostly include preidentified spectral bands not selected for the specific application. Most of these limitations are addressed in our study, i.e., a best-case scenario which identifies spectral bands, downsampled from a hyperspectral system, in low-noise, stable illumination conditions, and at a high temporal resolution, to identify the ideal wavelengths and most accurate timing for yield prediction. Future efforts could focus on translating these concepts to a multispectral setup (see next section).

5 Conclusions

Timely and accurate/precise yield assessment of crops can benefit farmers via optimization of management inputs for maximized profit, but also maximize the actual yield for market consumption, and can contribute to doing so on a sustainable basis. Our study focused on modeling pod weight, as a yield indicator of plants, using hyperspectral and biophysical data. Our approach hinged on PLSR for yield prediction of snap bean (cv. Huntington) at two different harvest times, namely early-and late-harvest, defined as when 50% of plants for early harvest and 70% of plants for late harvest, showed 50% of their pods being at industry sieve size 4 to 6.

Our early-harvest and late-harvest stage approaches confirmed that 20 to 25 DPH is when spectral signatures surface that result in a high accuracy for yield forecasting. Biophysical features proved to be critical in improving the performance of models. Also, the depicted growth trends in biophysical features, and their corresponding maxima around the same period (20 to 25 DPH), corroborated our results based on spectral-only data for the ideal time period to apply yield assessments. We demonstrated that, if using spectral-only data, the CR model, and if incorporating biophysical data with spectral information, the raw or smoothed transformations, should be the preprocessing method of choice. Our spectral model was able to explain ~72% of the variability in observed yield values ($R^2 = 0.72$), which bodes well for extending these results to other, more operational scenarios. In terms of spectral features, the NIR edge peak, chlorophyll absorption features in the visible domain, protein and oil absorption features in the early-SWIR region (~1000 nm), lignin and nitrogen absorption peaks in the ~1700 nm range, and sugar and structural absorption features (e.g., starch, lignin, cellulose, protein, nitrogen, etc.) in the ~2100 and ~2280 nm regions proved to be especially critical to robust modeling efforts. Our study represents a best-case scenario, from a spectral perspective, where the illumination conditions were optimal and controlled during data collection. As such the SNR remained relatively high even at the tail-ends of the detectors' spectral response curves, but would deteriorate in more typical, outdoor conditions, especially as more atmosphere (altitude) is introduced. That led to one of our conclusions, i.e., that while these models showed promising results and could be extensible to more operational platforms, the next phases of the research should focus exactly on that, i.e., scaling of results from a greenhouse scenario to UAS or other airborne modalities. Also, results from this study imply that for future work on the use of structural sensing systems, such as LiDAR and 3-D structure-from-motion approaches, could augment spectral-based models by including estimates of plant biophysical attributes. Finally, implementation of crop growth models, or models that aim to enhance statistical model accuracy by adding more auxiliary environmental variables, such as soil characteristics, geographical data, and other crop attributes, could also be an area of interest for future work.

Acknowledgments

This research was supported by the National Science Foundation PFI Award No. 1827551 and the introduced concepts and ideas are of author(s) and do not reflect the views of the National Science Foundation.

References

1. A. T. Carswell and W. A. Brown, Ed., "United States Census Bureau," The Encyclopedia of Housing, SAGE Publications, Inc., Thousand Oaks (2012).
2. United Nations, "World Population Prospects-Population Division-United Nations," World Population Prospects Revision, https://population.un.org/wpp/Publications/Files/WPP2015_DataBooklet.pdf (2015).
3. D. A. Ahlburg, A. C. Kelley, and K. O. Mason, *The Impact of Population Growth on Well-Being in Developing Countries*, Springer Science & Business Media, Berlin/Heidelberg, Germany (1996).
4. M. Cropper and C. Griffiths, "The interaction of population growth and environmental quality," *Am. Econ. Rev.* **84**(2), 250–254 (1994).
5. USDA, "USDA National Agricultural Statistics Service," vol. 51, USDA Statistical Services (2014).
6. B. M. Whelan and A. B. McBratney, "The 'null hypothesis' of precision agriculture management," *Precis. Agric.* **2**(3), 265–279 (2000).
7. D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps," *Biosyst. Eng.* **114**(4), 358–371 (2013).
8. C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: a review," *Precis. Agric.* **13**(6), 693–712 (2012).
9. F. F. Sabins, *Remote sensing: Principles and Applications*, Waveland Press, Long Grove, Illinois (2007).
10. I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: a review," *ISPRS J. Photogramm. Remote Sens.* **92**, 79–97 (2014).
11. S. Wold, "Chemometrics; what do we mean with it, and what do we want from it?" *Chemom. Intell. Lab. Syst.* **30**(1), 109–115 (1995).
12. B. M. Nicolaï et al., "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review," *Postharvest Biol. Technol.* **46**(2), 99–118 (2007).
13. B. Hapke, "Reflectance methods and applications," in *Encyclopedia of Spectroscopy and Spectrometry*, 3rd ed., J. C. Lindon, G. E. Tranter, and D. W. Koppenaal, Eds., pp. 931–935, Academic Press, Cambridge, Massachusetts (2017).
14. N. C. Coops et al., "Chlorophyll content in eucalypt vegetation at the leaf and canopy scales as derived from high resolution spectral data," *Tree Physiol.* **23**(1), 23–31 (2003).
15. L. Li et al., "Evaluating chlorophyll density in winter oilseed rape (*Brassica napus L.*) using canopy hyperspectral red-edge parameters," *Comput. Electron. Agric.* **126**, 21–31 (2016).
16. A. Pinar and P. J. Curran, "Technical note grass chlorophyll and the reflectance red edge," *Int. J. Remote Sens.* **17**(2), 351–357 (1996).
17. J. A. Hackwell et al., "LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing," *Proc. SPIE* **2819**, 102–107 (1996).
18. L. Lleó et al., "Multispectral images of peach related to firmness and maturity at harvest," *J. Food Eng.* **93**(2), 229–235 (2009).
19. X. Yu, H. Lu, and Q. Liu, "Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (*Brassica napus L.*) leaf," *Chemom. Intell. Lab. Syst.* **172**, 188–193 (2018).
20. Y. Huang, R. Lu, and K. Chen, "Prediction of firmness parameters of tomatoes by portable visible and near-infrared spectroscopy," *J. Food Eng.* **222**, 185–198 (2018).
21. X. Li et al., "SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology," *Postharvest Biol. Technol.* **143**, 112–118 (2018).

22. V. M. Gomes et al., "Comparison of different approaches for the prediction of sugar content in new vintages of whole Port wine grape berries using hyperspectral imaging," *Comput. Electron. Agric.* **140**, 244–254 (2017).
23. A. Rahman et al., "Nondestructive estimation of moisture content, pH and soluble solid contents in intact tomatoes using hyperspectral imaging," *Appl. Sci.* **7**(1), 109 (2017).
24. S. Van Dyk et al., "Determining the harvest maturity of vanilla beans," *Sci. Hortic.* **168**, 249–257 (2014).
25. F. Mendoza, R. Lu, and H. Cen, "Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content," *Postharvest Biol. Technol.* **73**, 89–98 (2012).
26. H. Yang, B. Kuang, and A. M. Mouazen, "In situ determination of growing stages and harvest time of tomato (*Lycopersicon esculentum*) fruits using fiber-optic visible-near-infrared (Vis-NIR) spectroscopy," *Appl. Spectrosc.* **65**(8), 931–938 (2011).
27. R. Guidetti et al., "Prediction of blueberry (*Vaccinium corymbosum*) ripeness by a portable vis-NIR device," *Acta Hortic.* **810**, 877–886 (2009).
28. R. E. Plant et al., "Relationships between remotely sensed reflectance data and cotton growth and yield," *Trans. Am. Soc. Agric. Eng.* **43**(3), 535–546 (2000).
29. P. J. Zarco-Tejada, A. Berjon, and J. R. Miller, "Stress detection in crops with hyperspectral remote sensing and physical simulation models," in *Airborne Imaging Spectrosc. Workshop*, pp. 1–5 (2004).
30. P. J. Zarco-Tejada, V. González-Dugo, and J. A. J. Berni, "Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera," *Remote Sens. Environ.* **117**, 322–337 (2012).
31. D. Zhao et al., "Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum," *Eur. J. Agron.* **22**(4), 391–403 (2005).
32. S. Delalieux et al., "Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications," *Eur. J. Agron.* **27**(1), 130–143 (2007).
33. J. Behmann et al., "Specim IQ: evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection," *Sensors* **18**(2), 441 (2018).
34. M. S. M. Asaari et al., "Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform," *ISPRS J. Photogramm. Remote Sens.* **138**, 121–138 (2018).
35. K. Nagasubramanian et al., "Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps," pp. 3–10, https://arxiv.org/abs/1804.08831 (2018).
36. Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*. **47**(8), 2973–2987 (2009).
37. W. Zhao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*. **54**(8), 4544–4554 (2016).
38. T. Amoriello et al., "Classification and prediction of early-to-late ripening apricot quality using spectroscopic techniques combined with chemometric tools," *Sci. Hortic.* **240**, 310–317 (2018).
39. D. K. Pathak and S. K. Kalita, "Spectral spatial feature based classification of hyperspectral image using support vector machine," in *6th Int. Conf. Signal Process. and Integrated Networks*, pp. 430–435 (2019).
40. X. Li, L. Zhang, and J. You, "Locally weighted discriminant analysis for hyperspectral image classification," *Remote Sens.* **11**(2), 109 (2019).
41. L. Zhi et al., "A dense convolutional neural network for hyperspectral image classification," *Remote Sens. Lett.* **10**(1), 59–66 (2019).
42. X. Ye et al., "Estimation of citrus yield from airborne hyperspectral images using a neural network model," *Ecol. Modell.* **198**(3–4), 426–432 (2006).
43. M. S. Mkhabela et al., "Crop yield forecasting on the Canadian Prairies using MODIS NDVI data," *Agric. For. Meteorol.* **151**(3), 385–393 (2011).

44. J. D. R. Soares et al., "Comparison of techniques used in the prediction of yield in banana plants," *Sci. Hortic.* **167**, 84–90 (2014).
45. S. Marino and A. Alvino, "Hyperspectral vegetation indices for predicting onion (*Allium cepa* L.) yield spatial variability," *Comput. Electron. Agric.* **116**, 109–117 (2015).
46. M. D. Johnson et al., "Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods," *Agric. For. Meteorol.* **218–219**, 74–84 (2016).
47. F. M. Aguate et al., "Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield," *Crop Sci.* **57**(5), 2517–2524 (2017).
48. A. J. Foster, V. G. Kakani, and J. Mosali, "Estimation of bioenergy crop yield and N status by hyperspectral canopy reflectance and partial least square regression," *Precis. Agric.* **18**(2), 192–209 (2017).
49. A. X. Wang et al., "Deep transfer learning for crop yield prediction with remote sensing data," in *Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc.*, pp. 1–5 (2018).
50. K. Kawamura et al., "Canopy hyperspectral sensing of paddy fields at the booting stage and PLS regression can assess grain yield," *Remote Sens.* **10**(8), 1249 (2018).
51. F. L. M. Padilla et al., "Monitoring regional wheat yield in Southern Spain using the GRAMI model and satellite imagery," *F. Crop. Res.* **130**, 145–154 (2012).
52. Y. R. Lai et al., "An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI," *Int. J. Appl. Earth Obs. Geoinf.* **72**, 99–108 (2018).
53. I. Becker-Reshef et al., "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data," *Remote Sens. Environ.* **114**(6), 1312–1323 (2010).
54. F. Kogan et al., "Winter wheat yield forecasting in Ukraine based on Earth observation, meteorologicaldata and biophysical models," *Int. J. Appl. Earth Obs. Geoinf.* **23**(1), 192–203 (2013).
55. M. Moriondo, F. Maselli, and M. Bindi, "A simple model of regional wheat yield based on NDVI data," *Eur. J. Agron.* **26**(3), 266–274 (2007).
56. Q. Wang et al., "On the relationship of NDVI with leaf area index in a deciduous forest site," *Remote Sens. Environ.* **94**(2), 244–255 (2005).
57. M. C. Reeves, M. Zhao, and S. W. Running, "Usefulness and limits on MODIS GPP for estimating wheat yield," *Int. J. Remote Sens.* **26**(7), 1403–1421 (2005).
58. H. Wold, "Systems analysis by partial least squares," in *Measuring the Unmeasurable*, P. Nijkamp, H. Leitner, and N. Wrigley, Eds., pp. 221–251, Marinus Nijhoff, Dordrecht (1985).
59. Z. Wenzhi et al., "Estimation of sunflower seed yield using partial least squares regression and artificial neural network models," *Pedosphere* **28**(5), 764–774 (2018).
60. A. Oussama et al., "Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores," *J. Am. Oil Chem. Soc.* **89**(10), 1807–1812 (2012).
61. N. Vásquez et al., "Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of Swiss-type cheese using spectral profiles," *J. Food Eng.* **219**, 8–15 (2018).
62. X. Zhang and Y. He, "Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves," *Ind. Crops Prod.* **42**(1), 416–420 (2013).
63. V. S. Weber et al., "Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes," *F. Crop. Res.* **128**, 82–90 (2012).
64. N. A. Noureldin et al., "Rice yield forecasting models using satellite imagery in Egypt," *Egypt. J. Remote Sens. Sp. Sci.* **16**(1), 125–131 (2013).
65. L. Wang et al., "Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images," *Fields Crops Res.* **164**(1), 178–188 (2014).
66. A. López-Maestresalas et al., "Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging," *Food Control* **70**, 229–241 (2016).
67. C. C. D. Lelong et al., "Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots," *Sensors* **8**(5), 3557–3585 (2008).

68. A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.* **36**(8), 1627–1639 (1964).
69. R. F. Kokaly and R. N. Clark, "Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression," *Remote Sens. Environ.* **67**(3), 267–287 (1999).
70. P. Bajorski, *Statistics for Imaging, Optics, and Photonics*, Vol. **808**, John Wiley & Sons, Hoboken, New Jersey (2011).
71. S. Munera et al., "Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging," *J. Food Eng.* **214**, 29–39 (2017).
72. M. W. Browne, "Cross-validation methods," *J. Math. Psychol.* **44**(1), 108–132 (2000).
73. G. Van Rossum and F. L. Drake, *Python Tutorial*, Vol. **42**, No. 4, Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands (2010).
74. F. Pedregosa et al., "Scikit-learn: machine learning in Python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
75. P. J. Curran, "Remote sensing of foliar chemistry," *Remote Sens. Environ.* **30**(3), 271–278 (1989).
76. D. H. Card et al., "Prediction of leaf chemistry by the use of visible and near infrared reflectance spectroscopy," *Remote Sens. Environ.* **26**(2), 123–147 (1988).
77. M. E. Martin and J. D. Aber, "High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes," *Ecol. Appl.* **7**(2), 431–443 (1997).
78. S. Saleh et al., "Effect of irrigation on growth, yield, and chemical composition of two green bean cultivars," *Horticulturae* **4**(1), 3 (2018).
79. S. Nasrullahzadeh et al., "Effects of shade stress on ground cover and grain yield of faba bean (*Vicia faba* L.)," *J. Food Agric. Environ.* **5**(1), 337–340 (2007).
80. S. Seager et al., "Vegetation's red edge: a possible spectroscopic biosignature of extraterrestrial plants," *Astrobiology* **5**(3), 372–390 (2005).
81. U.S. Department of Agriculture, Agricultural Research Service, "USDA National Nutrient Database for Standard Reference, Release 18 USDA National Nutrient Database for Standard Reference, Release 18," http://www.ars.usda.gov/nea/bhnrc/mafcl (2011).
82. J. Chang et al., "Corn (*Zea mays* L.) yield prediction using multispectral and multidate reflectance," *Agron. J.* **95**, 1447–1453 (2003).
83. A. Dobermann and J. L. Ping, "Geostatistical integration of yield monitor data and remote sensing improves yield maps," *Agron. J.* **96**, 285–297 (2004).
84. W. S. Lee et al., "Sensing technologies for precision specialty crop production," *Comput. Electron. Agric.* **74**(1), 2–33 (2010).
85. J. F. Shanahan et al., "Use of remote-sensing imagery to estimate corn grain yield," *Agron. J.* **93**, 583–589 (2001).
86. C. T. Leon et al., "Utility of remote sensing in predicting crop and soil characteristics," *Precis. Agric.* **4**, 359–384 (2003).
87. D. Inman et al., "Normalized difference vegetation index and soil color-based management zones in irrigated maize," *Agron. J.* **100**, 60–66 (2008).
88. P. K. Goel et al., "Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing," *Trans. Am. Soc. Agric. Eng.* **46**, 1235–1246 (2003).
89. C. Yang, J. H. Everitt, and J. M. Bradford, "Airborne hyperspectral imagery and linear spectral unmixing for mapping variation in crop yield," *Precis. Agric.* **8**, 279–296 (2007).

Amirhossein Hassnazadeh is a PhD student at the Chester F. Carlson Center for Imaging Science at Rochester Institute of Technology. He received his BSc degree in chemical engineering from the University of Guilan, Iran. His interests are remote sensing, computer vision, and machine learning.

Jan van Aardt is a professor in the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology. He received his BSc and Hons. forestry degrees from the University of Stellenbosch, South Africa. These were followed by MS and PhD forestry degrees at Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA. His research interests include imaging spectroscopy (hyperspectral) and LiDAR applications in forestry and precision agriculture.

Sean Patrick Murphy is a technician in plant pathology at Cornell AgriTech, Cornell University, Geneva, New York, USA, and focuses on diseases of broad-acre vegetable crops, such as table beets, snap beans, dry beans, and allium crops. He joined Cornell University in 2011 in the Horticulture Department working on peas, beans, and sweet corn before moving to plant pathology in 2017. He received his bachelor of technology degree in horticulture business management from SUNY Morrisville in 2016.

Sarah Jane Pethybridge is an associate professor of plant pathology at Cornell AgriTech, Cornell University, Geneva, New York, USA, and focuses on diseases of broad-acre vegetable crops. She specializes in vegetable disease epidemiology and management. She joined Cornell University in 2014 after roles in academia, industry, and government in Australia and New Zealand. She received her bachelor of agricultural science with first class honors and PhD in plant pathology in 1995 and 2000, respectively.