Video Article

Construction of Models for Nondestructive Prediction of Ingredient Contents in Blueberries by Near-infrared Spectroscopy Based on HPLC Measurements

Wenming Bai[1], Norio Yoshimura[1], Masao Takayanagi[1], Jingai Che[2], Naomi Horiuchi[2], Isao Ogiwara[3]

[1]United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology

[2]Faculty of Agriculture, Tokyo University of Agriculture and Technology

[3]Institute of Agriculture, Tokyo University of Agriculture and Technology

Correspondence to: Masao Takayanagi at masaot@cc.tuat.ac.jp

URL: https://www.jove.com/video/53981
DOI: doi:10.3791/53981

Keywords: Chemistry, Issue 112, Analytical Chemistry, Blueberry, Near-infrared Spectroscopy, Sugar, Organic Acid, Anthocyanin, Chemometrics, Partial Least Squares (PLS) Regression, HPLC

Date Published: 6/28/2016

Citation: Bai, W., Yoshimura, N., Takayanagi, M., Che, J., Horiuchi, N., Ogiwara, I. Construction of Models for Nondestructive Prediction of Ingredient Contents in Blueberries by Near-infrared Spectroscopy Based on HPLC Measurements. *J. Vis. Exp.* (112), e53981, doi:10.3791/53981 (2016).

Abstract

Nondestructive prediction of ingredient contents of farm products is useful to ship and sell the products with guaranteed qualities. Here, near-infrared spectroscopy is used to predict nondestructively total sugar, total organic acid, and total anthocyanin content in each blueberry. The technique is expected to enable the selection of only delicious blueberries from all harvested ones. The near-infrared absorption spectra of blueberries are measured with the diffuse reflectance mode at the positions not on the calyx. The ingredient contents of a blueberry determined by high-performance liquid chromatography are used to construct models to predict the ingredient contents from observed spectra. Partial least squares regression is used for the construction of the models. It is necessary to properly select the pretreatments for the observed spectra and the wavelength regions of the spectra used for analyses. Validations are necessary for the constructed models to confirm that the ingredient contents are predicted with practical accuracies. Here we present a protocol to construct and validate the models for nondestructive prediction of ingredient contents in blueberries by near-infrared spectroscopy.

Video Link

The video component of this article can be found at https://www.jove.com/video/53981/

Introduction

Near-infrared (NIR) spectroscopy is widely applied as a nondestructive technique to analyze contents of fruits and vegetables of various kinds.[1,2] Nondestructive analyses by NIR spectroscopy enable the shipping of only delicious fruits and vegetables with guaranteed qualities. NIR spectroscopy has already been applied to orange, apple, melon, cherry, kiwi fruit, mango, papaya, peach and so on to know their Brix that corresponds to the total sugar content, acidity, TSC (total solids contents), and so on. Recently, we have reported the application of NIR spectroscopy to the quality evaluation of blueberries.[3] We measured not only the total sugar content and the total organic acid content corresponding to acidity, but also the total anthocyanin content. Anthocyanin is a bioactive component which is believed to improve human health. It is convenient for consumers if they can buy delicious blueberries with an assurance of their sugar content, acidity, and anthocyanin content.

In NIR absorption spectra of fruits and vegetables, only broad absorption bands are observed. They are mainly the bands due to fiber and moisture. Although many weak bands due to various ingredients of the non-destructed target are observed simultaneously, the observed bands cannot be assigned to specific vibrational modes of specific components of the target in most cases. Therefore, the traditional technique to determine the content of a specific component using the Lambert-Beer's law is not effective for NIR spectra. Instead, calibration models to predict the contents of the target components from the observed spectra are constructed using chemometrics by examining the correlation between the observed spectra and the ingredient contents corresponding to the spectra.[4,5] Here, a protocol to construct and validate the models for prediction of total sugar content, total organic acid content corresponding to acidity, and total anthocyanin content of blueberries from NIR spectra is presented.

Figure 1 shows the general flow chart to construct reliable and robust calibration models. Samples of sufficient number are collected. Some of them are used for the construction of models while the others are used for the validation of the constructed models. For each of collected samples, an NIR spectrum is measured, and then the target components are analyzed quantitatively with traditional destructive chemical analysis methods. Here, high-performance liquid chromatography (HPLC) is used for the chemical analyses of sugars, organic acids, and anthocyanins. Partial least squares (PLS) regression is used for the construction of calibration models where the correlation between the observed spectra and the ingredient contents determined by chemical analyses is examined. In order to construct robust models with the best prediction ability, the pretreatments of observed spectra and the wavelength regions used for the prediction are also examined. Finally, the constructed models are validated to confirm their sufficient prediction ability. In the validation, the contents predicted from the observed spectrum

by the constructed model (predicted values) are compared with the contents determined by the chemical analyses (observed values). If the sufficient correlation cannot be found between the predicted and observed values, the calibration model should be re-constructed until the sufficient correlation is obtained. Although it is preferable to use different groups of samples for the construction and validation of a model as shown in this figure (external validation), samples in a same group are used both for the construction and the validation (cross validation) when the number of samples is not large enough.

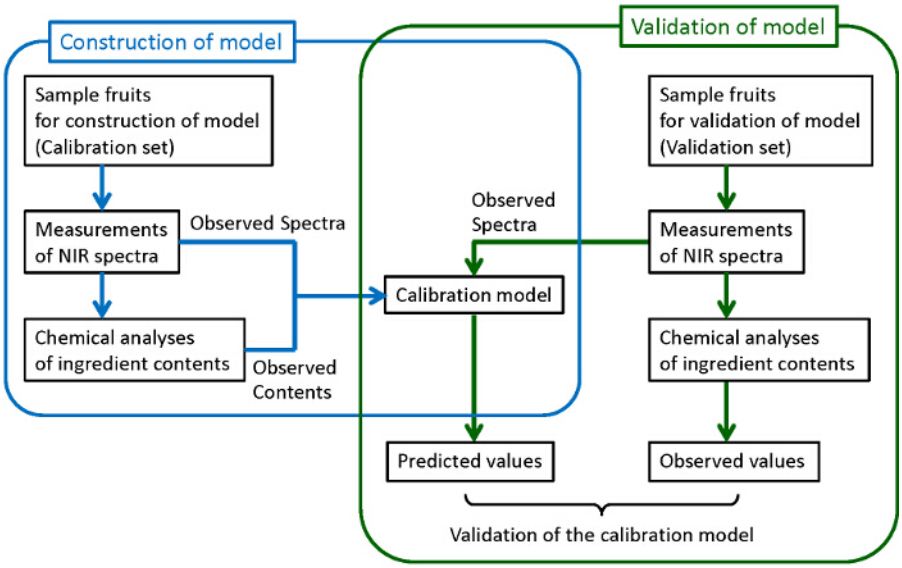

Figure 1. Flow chart for the construction and validation of the calibration model. The procedures surrounded by blue and green lines correspond, respectively, to the construction of a calibration model and its validation. Please click here to view a larger version of this figure.

Protocol

1. Collection of Samples

1. Decide which cultivars will be included in the target of the calibration model.
2. Collect sufficient number and various types of sample blueberries of the target cultivars.
    1. Collect preferably 100 blueberries for the construction of the calibration model, and at least 10 for the validation of the constructed model. In order to construct robust models, collect samples of various types, *i.e.* with various colors, sizes, and at various ripening conditions.
3. Weigh each blueberry. Note: The weights measured are used later for the calculation of content percent of ingredients of each blueberry.

2. Measurements of Spectra

1. Warm-up the spectrophotometer sufficiently (more than 1 hr) before the measurements to get reliable spectra.
2. Set the spectrophotometer. Ensure that the conditions are constant all through the measurements. An example of typical conditions for measurements is given below.
    1. Set the range of measurements to 12,500-3,600 cm^{-1}.
    2. Set the spectral resolution to 16 cm^{-1}.
    3. Set the accumulation to 32 times.
3. Select diffuse reflectance as the mode of measurement.
4. Put the standard reflector on the window of the spectrophotometer for diffuse reflectance measurements. By using the "background single channel" command, measure the background spectrum which is automatically used for the calculation of relative reflectance spectra from the spectra of sample blueberries measured later.
5. Put a blueberry sample in the center of the window of the spectrophotometer for diffuse reflectance measurements. By using the "sample single channel" command, measure spectra of each blueberry preferably at several points of the fruit.
   Note: Kubelka-Munk transformation[6,7] will also be performed automatically for the observed spectra of sample blueberries if the condition of spectral acquisition is set to do so. Kubelka-Munk transformation alters the spectra measured in the diffuse reflectance mode to the spectra equivalent to those measured in the transmission mode and is needed for the analyses of spectra with high accuracy. Spectra in the absorbance scale are used for analyses.
6. Calculate the average spectrum of the spectra of each sample using a data processing program such as MS Excel if the spectra of a blueberry sample are measured at several points. Use the averaged spectrum for analyses.

3. Pretreatment for HPLC Measurements of Sugars and Organic Acids[8]

Note: Extract sugars and organic acids of each blueberry, which are soluble in water, with ultrapure water as follows. The whole of each blueberry is used for analyses.

1. Keep the blueberries in a freezer below -30 °C ready for chemical analyses if they are not analyzed just after the spectral measurements.
2. Cut a blueberry into several pieces so it can be easily homogenized in the following steps. Cut the blueberry without defrosting when it is frozen.
3. Put the pieces into a 50 ml beaker.
4. Add ca. 10 ml of ultrapure water (distilled water whose electrical conductivity is less than 0.1 μS/cm) to the beaker.
5. Heat the cut blueberry in ultrapure water in a microwave oven for 20 sec to deactivate the enzymes that might decompose sugars during the analyses.
6. Add *ca.* 10 ml of ultrapure water to the beaker.
7. Homogenize the mixture for 5 min at 12,000 rpm with a homogenizer equipped with a standard shaft and generator.
8. Centrifuge the homogenized mixture for 10 min at 3,000 rpm (2,000 × g).
9. Collect filtrate by vacuum filtration of the centrifuged sample using a 5B paper filter.
10. Repeat the steps 3.6-3.9 twice on the filtration residue to collect all sugars and organic acids, and combine all filtrates.
11. Measure the pH of the filtrate and adjust it to 7 with dilute hydrochloric acid (0.1 and 0.01 mol L^{-1}) and dilute aqueous solutions of sodium hydroxide (0.1 and 0.01 mol L^{-1}).
12. Dilute the filtrate to 50 ml with ultrapure water.
13. Divide the sample into two; one for the analysis of sugars and the other for the analysis of organic acids.
14. Pass the first sample solution through columns (two C18, CM and QMA) connected in series to exclude pigments, cations, and anions. Throw away the first 1 ml of the sample solution from the columns. Then use the sample solution from the columns for the analysis of sugars by HPLC.
15. Pass the second sample solution through columns (two C18 and CM) connected in series to exclude pigments and cations. Throw away the first 1 ml of the sample solution from the columns. Then use the sample solution from the columns for the analysis of organic acids by HPLC.
16. Centrifuge each solution from steps 3.14 and 3.15 at 6,600 rpm (5,800 × g), for 10 min in a microtube equipped with a 0.45 μm filter with a mini-centrifuge before the analysis by HPLC.

4. HPLC Measurements of Sugars

Note: In this study, sum content of sucrose, glucose and fructose of each blueberry is considered as the total sugar content. Therefore, the working curve for each of three sugars is obtained first, and then sum content of the sugars in each blueberry is obtained. The standard contents are reported as 0.3-0.4 wt% (sucrose), 3.8-4.8 wt% (glucose), and 4.2-5.3 wt% (fructose).[9]

1. Measure about 200 mg of sucrose accurately, and dissolve it in 50 ml ultrapure water to prepare a standard solution. Dilute 5 ml of the solution to 50 ml with ultrapure water to prepare the second standard solutions. Prepare similarly the third standard solution from the second standard solution.
2. Prepare the standard solutions of glucose and fructose, similarly.
3. Arrange the HPLC system as follows:
    1. Use a gel permeation column in the column oven at 40 °C.
    2. Use degassed ultrapure water with the flow rate of 0.1 ml/min as the eluate.
    3. Use a refractive index detector.
4. Measure the HPLC spectrograms of standard solutions by injecting a 20 μl aliquot for each measurement. Note: Here, PAC Solution is used as the software for the measurement.
5. Get the area intensity of the band of sugar on the chromatogram of each standard solution by clicking 're-analysis' with the right button of the mouse.
6. Plot the area intensities against the corresponding concentrations to get the working curve for each sugar by the linear regression, where the equation representing the relation between the area intensity and the concentration is obtained for each sugar.
7. Measure the HPLC spectrograms of sample solutions by injecting a 20 μl aliquot for each measurement.
8. Get the area intensities of the bands of sugars on the chromatogram of each sample solution as previously described in step 4.5.
9. Obtain the concentrations of the sugars in the solutions using the equations corresponding to the working curves obtained in step 4.6.
10. Obtain the amount of each sugar in each blueberry from the concentrations of the sample solution obtained in the previous step and the total volume of the sample solution (50 ml, see step 3.12).
11. Obtain the total sugar amounts of each fruit by summing up the contents of three sugars.
12. Obtain the content percent of the total sugar of each blueberry by using the weight measured in step 1.3.

5. HPLC Measurements of Organic Acids

Note: In this study, sum content of citric acid, quinic acid, malic acid, and succinic acid are considered as the total organic acid content. Therefore, working curve for each of four organic acids is obtained first, and then the organic acid content in each blueberry is measured. The standard contents are reported as 0.42-0.62 wt% (citric acid), 0-0.15 wt% (quinic acid), 0.08-0.23 wt% (malic acid), and 0.06-0.25 wt% (succinic acid).[9]

1. Measure about 5 mg of citric acid accurately, and dissolve it into 50 ml ultrapure water to prepare a standard solution. Dilute 5 ml of the solution to 50 ml with ultrapure water to prepare the second standard solutions. Prepare similarly the third standard solution from the second standard solution.

2. Prepare the standard solutions of quinic acid, malic acid, and succinic acid, similarly.
3. Arrange the HPLC system as follows:
    1. Use two anion-exchange columns connected in series in the column oven at 40 °C.
    2. Use degassed 0.1% aqueous solution of phosphoric acid with the flow rate of 0.02 ml/min as the eluate.
    3. Use an ultraviolet-visible detector set at 210 nm.
4. Measure the HPLC spectrograms of standard solutions by injecting a 20 µl aliquot of standard solution for each measurement.
5. Get the area intensity of the band of organic acid on the chromatogram of each standard solution by clicking 're-analysis' with the right button of the mouse.
6. Plot the area intensities against the corresponding concentrations to get the working curve for each organic acid by the linear regression, where the equation representing the relation between the area intensity and the concentration is obtained for each organic acid.
7. Measure the HPLC spectrograms of sample solutions by injecting a 20 µl aliquot of the sample for each measurement.
8. Get the area intensities of the bands of organic acids on the chromatogram of each sample solution as described previously in step 5.5.
9. Obtain the concentrations of the organic acids in the solutions using the equations corresponding to the working curves obtained in step 5.6.
10. Obtain the amount of each organic acid in each blueberry from the concentrations of the sample solution obtained in the previous step and the total volume of the sample solution (50 ml, see step 3.12).
11. Obtain the amount of total organic acid in each blueberry by summing up the contents of the four organic acids.
12. Obtain the content percent of total organic acid of each blueberry by using the weight measured in step 1.3.

6. Pretreatment for HPLC Measurements of Anthocyanins

1. Keep the blueberries in a freezer below -80 °C ready for chemical analyses if they are not analyzed just after the spectral measurements.
2. Dry each frozen fruit with a vacuum lyophilizer for 12 hr.
3. Extract anthocyanin from the dried blueberry in 1% methanol solution of trifluoroacetic acid [weight of blueberry (g)/volume of the solution (ml) = 1/10] by leaving the mixture in a refrigerator at 4 °C for 12 hr.
4. Centrifuge the extract for 15 min in a 2 ml microtube using an ultracentrifuge at -8 °C and 15,000 rpm (21,900 × g).
5. Filter the extract through a 0.45 µm filter to get the sample for HPLC measurements.

7. HPLC Measurements of Anthocyanins

Note: About 13 kind anthocyanins are included in blueberries. Since it is difficult to get working curves for all anthocyanins, a working curve for only cyanidin-3-*O*-glucoside chloride, one of the most popular anthocyanins in blueberries, is obtained. The working curve is applied for approximate quantifications of other anthocyanins.

1. Measure about 1.5 mg of cyanidin-3-*O*-glucoside chloride accurately, and dissolve it into 10 ml of 1% methanol solution of trifluoroacetic acid to prepare a standard solution. Dilute 5 ml of the solution to 10 ml with 1% methanol solution of trifluoroacetic acid to prepare the second standard solutions. Similarly, prepare the third and the fourth standard solutions sequentially.
2. Arrange the HPLC system as follows:
    1. Use a C18 reverse phase column in a column oven at 40 °C.
    2. Apply the gradient method using eluates of 0.1% aqueous trifluoroacetic acid (elute A) and 0.5% trifluoroacetic acid in acetonitrile (elute B) with the flow rate of 0.1 ml/min, where the ratio of elute B increases from 8% to 15% during 0-50 min after the injection, and from 15% to 75% during 50-60 min after the injection.
    3. Use a photodiode array detector monitoring at 520 nm.
3. Measure the HPLC spectrograms of standard solutions by injecting a 10 µl aliquot for each measurement. "LC Solution" is used as the software for the measurement.
4. Get the area intensity of the band of cyanidin-3-*O*-glucoside chloride on the chromatogram of each standard solution by clicking 're-analysis' with the right button of the mouse.
5. Plot the area intensities against the corresponding concentrations to get the working curve for cyanidin-3-*O*-glucoside chloride by the linear regression, where the equation representing the relation between area intensity and concentration is obtained for cyanidin-3-*O*-glucoside chloride.
6. Measure the HPLC spectrograms of sample solutions by injecting a 10 µl aliquot for each measurement.
7. Get the area intensity of the band of each anthocyanin on the chromatogram of each sample solution as previously described in step 7.4.
8. Obtain the concentrations of the anthocyanins in the solutions using the equation corresponding to the working curve obtained in step 7.5.
9. Obtain the amounts of each anthocyanin in each blueberry from the concentration obtained in the previous step and the total volume of the sample solution used in step 6.3.
10. Obtain the total amount of anthocyanin in each blueberry by summing up the contents of the thirteen anthocyanins.
11. Obtain the content percent of the total anthocyanin of each blueberry by using the weight measured in step 1.3.

8. Construction of Calibration Models for Prediction of Ingredient Contents

Note: PLS regression,[4,5] which is a kind of multiple regression technique using latent variants, is used for the construction of calibration models for each ingredient from the observed spectra and the ingredient contents determined by chemical analyses. PLS regression is performed either with the commercial programs or with the home-made programs. See references[5,10] for the detailed processes of the construction of models.

1. Examine which pretreatments for observed spectra are most effective for accurate and robust prediction.
    1. Construct calibration models by applying one or two of the following pretreatments: MSC (Multiplicative Scatter Correction),[1,2,5] SNV (Standard Normal Variate),[1,2,5] MMN (Min-Max Normalization), COE (Constant Offset Elimination), and the first or the second

derivative calculation by SG (Savitzky-Golay) method.[1,2,5] Predict the ingredient contents of the validation set from their spectra with the constructed models.
Note: In MMN, a spectrum is normalized so that the minimum and maximum values become 0 and 1, respectively. In COE, the ordinate of a spectrum is shifted so that the minimum value becomes zero.

2. Calculate coefficient of determination, R^2, and residual predictive deviation, *RPD,* between the observed and predicted values of the validation set to examine which pretreatments for observed spectra are most effective. Choose the combination of pretreatments giving greater R^2 and *RPD*.

2. Examine which wavenumber regions are effective for the accurate and robust prediction by applying, for example, moving-windows PLS technique[11] to search the effective regions.
Note: The procedure corresponds to removing the wavenumber regions where spectra contain no effective information for predictions or contain information that interferes with predictions.

9. Validation of the Constructed Calibration Models

Note: See references[5,10] for the detailed processes of the validation of constructed models.

1. Predict ingredient contents of the validation set from their spectra with the constructed calibration models with the best combination of pretreatments and for the wavenumber regions effective for the prediction.[5,10]
2. Calculate R^2 and *RPD* between the observed and predicted values of the validation set.[5,10]
3. Examine whether the general criteria for the practical performance of calibration models,[12,13] $R^2 > 0.85$ and *RPD* > 2.5, are satisfied. Consider the reconstruction of model if the criteria are not satisfied.

Representative Results

Figure 2 shows as an example a set of NIR absorption spectra of blueberries where spectra of 70 blueberries are shown simultaneously. Since the bands definitely assignable to sugars, organic acids, or anthocyanins are not observed in the NIR spectra, traditional Lambert-Beer's law is not applicable to quantify the ingredient contents. Therefore, the construction of models for the prediction of ingredient contents is necessary.

Figure 3 shows typical chromatograms for the quantitative analysis of sugars in blueberries. Three panels from the top are, respectively, the chromatograms of standard solutions of sucrose, glucose, and fructose. The bottom panel shows a chromatogram of a sample solution, *i.e.* the extract of a blueberry. The kinds and concentrations of sugars in the sample solution are known from the retention times and the area intensities of the observed peaks. Total sugar content is obtained as the sum of sucrose, glucose, and fructose contents.

Figure 4 shows an example of chromatogram for the analysis of organic acids in a blueberry. By referring to the chromatograms of the standard solution (not shown here), the kinds and concentrations of organic acids in the sample solution are known. For the assignments of the observed peaks shown in the figure legend, two peaks are observed for quinic acid in the chromatograms of the standard and sample solutions. They might be assignable to isomers of quinic acid. Total organic acid content is obtained as sum of citric acid, quinic acid, malic acid, and succinic acid contents.

Figure 5 shows an example of chromatogram for the analysis of anthocyanins in a blueberry. Many peaks corresponding to different kind anthocyanins are observed. Since the order of elusion for these anthocyanins were reported[14,15] as shown in **Table 1**, the observed peaks can be assigned to individual anthocyanins. Total anthocyanin content is obtained as the sum of the contents of 13 kinds of anthocyanins.

Calibration models are constructed from the observed spectra and the chemically determined ingredient contents. **Table 2** shows an example of the examination of pretreatments. Six type pretreatments including "none (without pretreatment)" were examined for the construction of the calibration model of total sugar content using spectra at a fixed wavenumber region of 12,500-3,600 cm^{-1}. Different pretreatments result in different prediction performances. Performances of the models were evaluated with R^2 and *RPD*. The type of pretreatments that gives the best prediction performance is chosen. In **Table 2**, "Second derivative + MSC," which means MSC after the second derivative calculation, gives the best results. Then the wavenumber regions used for the model construction are examined by varying the regions with the fixed pretreatments.

Figure 6 shows as an example a result of cross validation of the calibration model for total sugar content, where the correlation between the values predicted by NIR spectroscopy and those determined by HPLC is shown. The model was constructed with "the second derivative + MSC" as the pretreatments and using the 8,539-7,775 cm^{-1} region of the spectra. The prediction performance of the model is $R^2 = 0.85$ and *RPD* = 2.6, just above the criteria for the practical use. In this example, the number of samples used for the model construction was 30, which is too small to construct high performance models.

Figure 7 shows as an example a result of the cross validation of the calibration model for total organic acid content, where the correlation between the values predicted by NIR spectroscopy and those determined by HPLC is shown. The model was constructed with "the first derivative + MSC" as the pretreatments and using the 7,505-5,446 and 4,605-4,242 cm^{-1} regions of spectra. The prediction performance of the model is $R^2 = 0.92$ and *RPD* = 3.6, which are sufficient for the practical application.

Figure 8 shows as an example a result of the external validation of the calibration model for total anthocyanin content, with the correlation between the values predicted by NIR spectroscopy and those determined by HPLC. The model was constructed with "the first derivative" as the pretreatment and using the 12,489-6,094 and 4,605-4,242 cm^{-1} regions of spectra. The prediction performance of the model is $R^2 = 0.95$ and *RPD* = 4.4, which shows fairly good performance of the constructed model. Since anthocyanin exists mainly in the peel of blueberries, it is easily observed with diffuse reflectance measurements although its content in a blueberry is not high. The good performance shown in **Figure 8** would be caused also by the large number of samples (70) used for model construction.

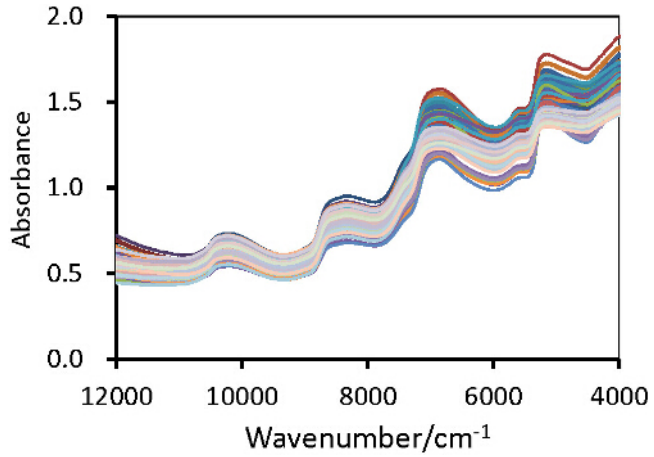

Figure 2. NIR absorption spectra of blueberries. Spectra of 70 blueberries are shown simultaneously. Please click here to view a larger version of this figure.

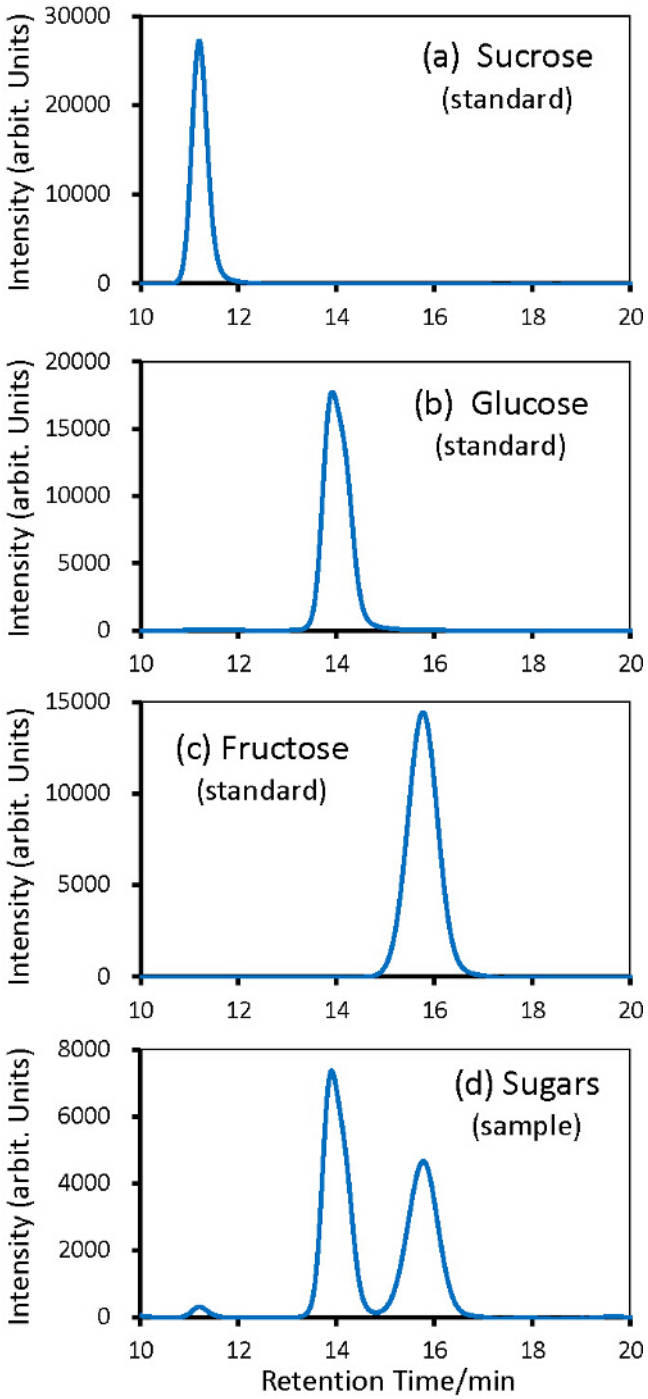

Figure 3. Chromatograms for the quantitative analysis of sugars in blueberries. Chromatograms of standard solutions of (**A**) sucrose, (**B**) glucose, (**C**) fructose, and (**D**) a chromatogram of a sample solution. Please click here to view a larger version of this figure.

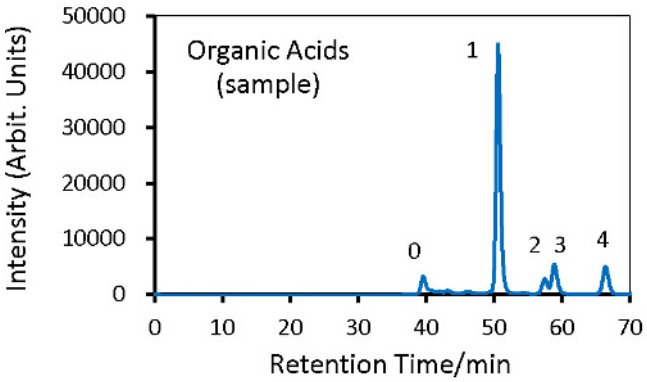

Figure 4. Chromatogram for the quantitative analysis of organic acids in a blueberry. Observed peaks correspond to citric acid (1), malic acid (2), quinic acid (0 and 3), and succinic acid (4). Please click here to view a larger version of this figure.

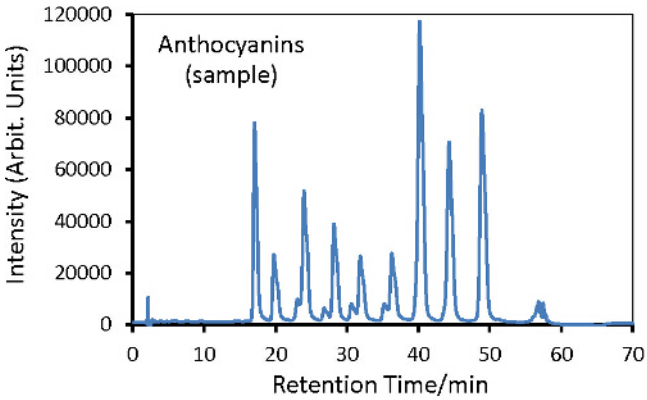

Figure 5. Chromatogram for the quantitative analysis of anthocyanins in a blueberry. Observed peaks are assigned to individual anthocyanins listed in **Table 1** where the standard retention time for each anthocyanin is shown. Please click here to view a larger version of this figure.

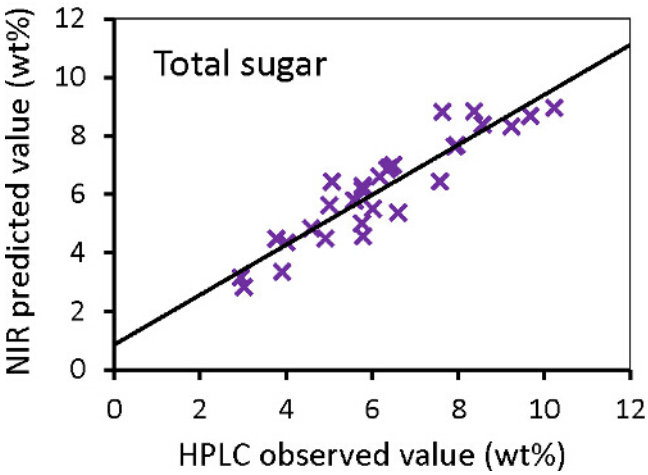

Figure 6. A result of cross validation of the model for total sugar content. The values predicted by NIR spectroscopy are plotted against those determined by HPLC. R^2 = 0.85 and *RPD* = 2.6 are obtained. Please click here to view a larger version of this figure.

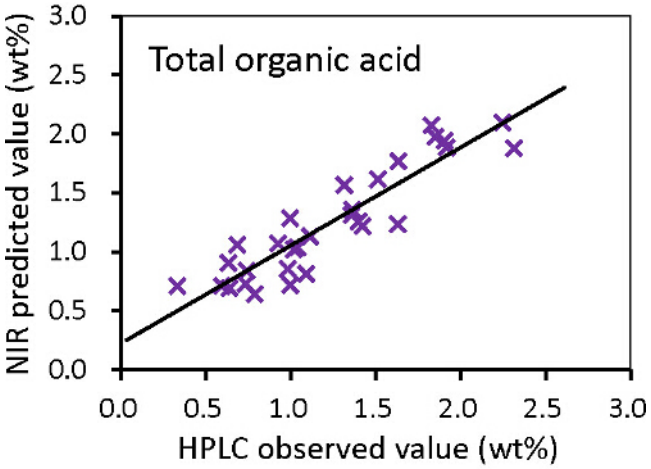

Figure 7. A result of cross validation of the model for total organic acid content. The values predicted by NIR spectroscopy are plotted against those determined by HPLC. R^2 = 0.92 and *RPD* = 3.6 are obtained. Please click here to view a larger version of this figure.

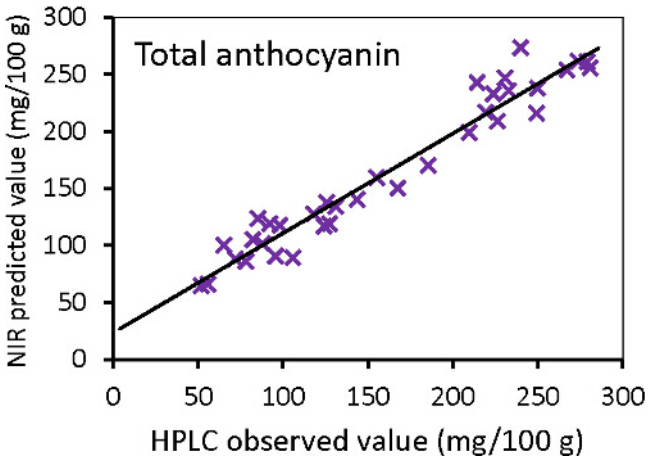

Figure 8. A result of external validation of the model for total anthocyanin content. The values predicted by NIR spectroscopy are plotted against those determined by HPLC. R^2 = 0.95 and *RPD* = 4.4 are obtained. Please click here to view a larger version of this figure.

Formula	Anthocyanin	Representative Retention time (min)
$C_{21}H_{21}O_{12}$	Delphinidin-3-*O*-galactoside	17.3
$C_{21}H_{21}O_{12}$	Delphinidin-3-*O*-glucoside	19.7
$C_{21}H_{21}O_{11}$	Cyanidin-3-*O*-galactoside	22.8
$C_{20}H_{19}O_{11}$	Delphinidin-3-*O*-arabinoside	23.6
$C_{21}H_{21}O_{11}$	Cyanidin-3-*O*-glucoside	24.5
$C_{22}H_{23}O_{12}$	Petunidin-3-*O*-galactoside	28.7
$C_{20}H_{19}O_{10}$	Cyanidin-3-*O*-arabinoside	31.3
$C_{22}H_{23}O_{12}$	Petunidin-3-*O*-glucoside	36.0
$C_{22}H_{23}O_{11}$	Peonidin-3-*O*-galactoside	37.0
$C_{21}H_{21}O_{11}$	Petunidin-3-*O*-arabinoside	40.8
$C_{22}H_{23}O_{11}$	Peonidin-3-*O*-glucoside	43.7
$C_{23}H_{25}O_{12}$	Malvidin-3-*O*-galactoside	45.0
$C_{23}H_{25}O_{12}$	Malvidin-3-*O*-glucoside	49.6

Table 1. Major anthocyanins contained in blueberries. The representative retention times in the HPLC analysis under the present experimental conditions are also listed.

Preprocessing	Wavenumber region used for analysis (cm^{-1})	RPD	R^2
None	12,500-3,600	1.7	0.69
Second derivative	12,500-3,600	2.6	0.85
First derivative	12,500-3,600	2.5	0.84
MSC	12,500-3,600	2.3	0.81
Second derivative + MSC	12,500-3,600	2.8	0.88
First derivative + MSC	12,500-3,600	2.7	0.87

Table 2. An examination of the dependence of prediction performance on the pretreatments of the observed spectra. R^2 and *RPD* for the prediction of total sugar content are listed.

Discussion

Some additional comments on the protocol are described here. Firstly, in step 1.1, it is mentioned to decide the cultivars included in the target. Although it is possible to construct models covering blueberries from many cultivars or without specifying cultivars, the prediction accuracies with the models are sometimes much lower than those with the models for a single cultivar and for limited cultivars. It should also be noted that the calibration models should be constructed for blueberries from each production site to get high prediction performance because blueberries harvested at different production sites have different characteristics which affect prediction performance.[1]

Secondly, in step 2.3, it is mentioned to select the diffuse reflectance mode for the measurements of spectra. The transmission mode is also prepared for measurements on the spectrophotometer. Although spectra measured in the transmission mode are also available for the construction of calibration models, more accurate and more robust models can be constructed with the spectra measured in the diffuse reflectance mode in most cases. The total organic acid contents cannot be predicted with the spectra measured in the transmittance mode.[3]

Thirdly, for the measurements of spectra of blueberries, it is not recommended to measure spectra at the calyx since the surface condition and the contents around the calyx are different from those at other positions. Nevertheless, it is possible to construct calibration models using an ample number of spectra measured at both the calyx and other positions. However, the accuracies of the models are in most cases lower than those of the models constructed with the spectra measured only at positions other than the calyx.

Fourthly, a NIR spectrum of blueberry depends on the temperature. Therefore, for precise prediction either it is important to always measure spectra at the same ambient temperature or to construct calibration models with compensation for temperature variation.[1]

Fifthly, although only R^2 and *RPD* are used for choosing the pretreatments and assessing the performance of constructed models here, some other values such as *SEC* (Standard Error of Calibration), *SEP* (Standard Error of Prediction), *SECV* (Standard Error of Cross-Validation), *RMSEP* (Root Mean Square Error of Prediction), and *RMSECV* (Root Mean Square error of Cross-Validation) are usually used for more detailed examination. In our previous paper,[3] for example, *RMSEP* and *RMSECV* were used for choosing pretreatments and assessing the performance of constructed models.

Nondestructive prediction of total sugar, total organic acid, and total anthocyanin contents in a blueberry was found to be possible if the models for predictions are constructed properly. This technique is applicable for the selection of only delicious blueberries from all harvested ones, which cannot be achieved with other traditional analytical techniques.[8,9] Although the procedures of the chemical analyses may seem complicated, they are included in popular analytical techniques and are not accompanied by great difficulties. It is important to get accurate results for the chemical analyses because the results are the basis of the constructed model. In this study, RSD (relative standard deviation) of the HPLC measurements was around 1%. It is also necessary to follow the basic procedure, *e.g.* as shown in **Figure 1**, for the construction of practically applicable models.

Simple and quick methods instead of HPLC can be applied for the chemical analyses. Total sugar content and acidity can be measured, respectively, with a refractometer (Brix meter) and a pH meter. The pH differential method[16,17] is applicable for the measurement of total anthocyanin content. Application of the simple methods make the construction of models much easier although the accuracy of values predicted by the models might be lower than those predicted by the models constructed on the basis of the HPLC measurements shown here. Nevertheless, the accuracy of the models constructed on the basis of simple chemical analyses may be practically applicable at production sites and circulation processes because high accuracies are not always needed there. The methods for the chemical analyses, therefore, should be selected according to the accuracies needed for the models to be constructed.

Although some fruits such as apple and orange are sold generally with guaranteed sugar contents and acidities, blueberries have not been sold with guaranteed qualities. As a result, the quality of the commercial blueberries does not seem stable at least in Japan; sometimes low quality blueberries are sold in markets. The nondestructive analytical methods by NIR spectroscopy shown here is expected to enable, in principle, the shipment and sale of blueberries with guaranteed qualities.

Finally, there are limitations of this method. Firstly, as mentioned above, the construction of prediction models is rather troublesome. Moreover, the prediction model should be constructed for each site and each year of cultivation because the difference in the amounts of coexisting components (which depend on the site and the year of cultivation) affects the precision of the prediction. Therefore, some effort is needed for the maintenance of prediction models. Secondly, although we have shown that near-infrared spectroscopy is, in principle, applicable for the quality check of blueberries, the equipment and techniques shown here are only used in the laboratory and not applicable at production sites because

a quick check of the large amounts of berries at production sites is impossible. Practical development of suitable equipment and development of robust calibration models suitable for use in production sites and circulation processes are future directions.

Disclosures

We have nothing to disclose.

Acknowledgements

This work was partially supported by the project "A Scheme to Revitalize Agriculture and Fisheries in Disaster Area through Deploying Highly Advanced Technology" of Ministry of Agriculture, Forestry and Fisheries, Japan.

References

1. Ozaki, Y., McClure, W.F., & Christy, A. A. *Near-infrared Spectroscopy in Food Science and Technology.* John Wiley & Sons, Inc., Hoboken (2007).
2. Sun, D.-W. *Infrared Spectroscopy for Food Quality Analysis and Control.* Academic Press, New York (2009).
3. Bai, W., Yoshimura, N., & Takayanagi, M. Quantitative analysis of ingredients of blueberry fruits by near infrared spectroscopy. *J. Near Infrared Spectrosc.* **22**, 357-365 (2014).
4. Hasegawa, T. Chemometrics in infrared spectroscopic analysis. In: *Introduction to Experimental Infrared Spectroscopy.* Tasumi, M., ed., John Wiley & Sons, Chichester, 97-113, (2015).
5. Varmuza, K., & Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics.* CRC Press, Boca Raton (2009).
6. Kubelka, P. New contributions to the optics of intensely light-scattering materials. Part I. *J. Opt. Soc. Am.* **38**, 448-457 (1948).
7. Juang, R.-H., & Storey, D.E. Quantitative determination of the extent of neutralization of carboxylic acid functionality in carbopol 974P NF by diffuse reflectance fourier transform infrared spectrometry using Kubelka-Munk function. *Pharm Res.*, **15**, 1714-1720 (1998).
8. Ogiwara, I., Ohtsuka, Y., Yoneda, Y., Sakurai, K., Hakoda, N., & Shimura, I. Extraction method by water followed by microwave heating for analyzing sugars in strawberry fruits. *J. Jpn. Soc. Hort. Sci.* **68**, 949-953 (1999).
9. Che, J., Suzuki, S., Ishikawa, S., Koike, H., & Ogiwara, I. Fruit ripening and quality profile of 64 cultivars in three species of blueberries grown in Tokyo. *Hort. Res.* (Japan) **8**, 257-265 (2009).
10. Pomerantsev, A.L. *Chemometrics in Excel.* John Wiley & Sons, Inc., Hoboken (2014).
11. Jiang, H.J., Berry, R.J., Siesler, H.W., & Ozaki, Y. Wavelength Interval Selection in Multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Anal. Chem.*, 74, 3555-3565 (2002).
12. Edney, M.J., Morgan, J.E., Williams, P.C., & Campbell, L.D. Analysis of feed barley by near infrared reflectance spectroscopy, *J. Near-Infrared Spectrosc.* **2**, 33-41 (1994).
13. Mathison, G.W., *et al.* Prediction of composition and ruminal degradability characteristics of barley straw by near infrared reflectance spectroscopy, *Can. J. Anim. Sci.* **79**, 519-523 (1999).
14. Chiara, F., *et al.* Analysis of anthocyanins in commercial fruit juices by using nano-liquid chromatography electrospray-mass spectrometry and high performance liquid chromatography with UV-vis detector. *J. Separation Sci.*, **34**, 150-159 (2011).
15. Li, Q., *et al.* Antioxidant anthocyanins screening through spectrum-effect relationships and DPPH-HPLC-DAD analysis on nine cultivars of introduced rabbiteye blueberry in China. *Food Chemistry*, **132**, 759-765 (2013).
16. Sinelli, N., *et. al.* Evaluation of quality and nutraceutical content of blueberries (Vaccinium corymbosum L.) by near and mid-infrared spectroscopy. *Postharvest Biol. Technol.* **50**, 31-36 (2008).
17. Giusti, M.M., & Wrolsted, R.E. Anthocyanins: characterization and measurement with UV-visible spectroscopy. In *Current Protocols in Food Analytical Chemistry.* Wrolstad, R.E., & Schwartz, S.J. eds., John Wiley & Sons, New York, 1-13 (2001).

