

Development of a predictive model for ‘Lapins’ sweet cherry dry matter content using a visible/near-infrared spectrometer and its potential application to other cultivars

P.M.A. Toivonen, A. Batista, and B. Lannard

Abstract: This work was conducted to determine if a useful model could be developed for a portable visible/near-infrared spectrometer for non-destructively predicting dry matter content in ‘Lapins’ and other cultivars of sweet cherries. Absorptions at a range of 948–957 nm (default selected by the instrument model-building program) or 858–1008 nm (user selected) were used to create models associating absorption values with actual measured dry matter contents of cherries. The best model was created using the user selected waveband and this model was determined to be highly predictive of dry matter, with a resolution of 0.5% dry matter content. External validation of this model was carried out using three different sweet cherry cultivars (Staccato™, Sentennial™, and Sovereign™) and the model was found to be robust, i.e., quite accurate in predicting dry matter content in these other cultivars, with R^2 values of 0.96, 0.94, and 0.99, and root mean square error of prediction values of 0.51, 0.74, and 0.56, respectively. The results indicate that dry matter in sweet cherries can be predicted accurately and non-destructively using visible/near-infrared spectroscopy. In the case of the cherry cultivars tested in this paper, a model developed using ‘Lapins’ fruit was reliable in predicting dry matter in these other cultivars.

Key words: internal fruit quality, ‘Lapins’ sweet cherry, spectroscopy, model building.

Résumé : L'étude devait établir s'il est possible d'élaborer un modèle efficace pour un spectromètre portatif dans le spectre visible/proche infrarouge afin de prévoir de manière non destructive la teneur en matière sèche des fruits du cerisier sauvage Lapins et d'autres cultivars. Les auteurs ont retenu les longueurs d'onde de 948–957 nm (valeur par défaut sélectionnée par le programme de modélisation) ou de 858–1008 nm (sélectionnée par l'utilisateur) pour créer les modèles associés à la teneur en matière sèche réelle relevée sur les cerisiers. Le meilleur modèle est celui reposant sur la plage établie par l'utilisateur et ce modèle prédit très bien la teneur en matière sèche des fruits, soit avec une précision de 0,5 %. Le modèle a été validé sur le terrain sur trois cultivars, en l'occurrence Staccato^{MC}, Sentennial^{MC} et Sovereign^{MC}. Le modèle est robuste, c.-à-d. assez précis pour prédire la teneur en matière sèche des fruits de ces cultivars, avec une valeur R^2 respective de 0,96, 0,94 et 0,99, et un écart-type de 0,51, 0,74 et 0,56. Ces résultats indiquent qu'on peut prédire avec précision et de manière non destructive la teneur en matière sèche des cerises par spectroscopie dans le spectre visible/proche infrarouge. Le modèle développé pour la variété Lapins était assez fiable pour prédire la teneur en matière sèche des fruits des autres cultivars testés dans le cadre de cette expérience. [Traduit par la Rédaction]

Mots-clés : qualité interne du fruit, cerisier sauvage Lapins, spectroscopie, modélisation.

Introduction

Prediction and management of poststorage fruit quality has been the focus of much postharvest research for tree fruits. Recently, the concept that tree fruit dry matter is an important at-harvest constituent

that relates to poststorage quality has evolved (Palmer et al. 2010; Crisosto et al. 2012). Direct measures of dry matter are destructive and when applied to large numbers of samples can consume significant numbers of fruit, time, and resources (McGlone and Kawano 1998).

Received 11 January 2017. Accepted 5 April 2017.

P.M.A. Toivonen and B. Lannard. Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada.

A. Batista. Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada; Chemical Biology Program, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada.

Corresponding author: Peter M.A. Toivonen (email: peter.toivonen@agr.gc.ca).

© Her Majesty the Queen in right of Canada 2017. Permission for reuse (free in most cases) can be obtained from [RightsLink](https://www.copyright.com).

Visible/near-infrared (VIS/NIR) spectroscopy has proven to be a reasonable non-destructive approach to measuring dry matter in many fruits (McGlone and Kawano 1998; Walsh et al. 2004). The development of relatively inexpensive, handheld instruments (Nicolai et al. 2007) has provided the opportunity to take routine measures of dry matter to assess differences between cultivars and orchards and also to verify the association between measured dry matter values at harvest and postharvest quality.

There is an obvious benefit in using non-destructive tools for quality assessment for commercial fruit harvest, handling, and distribution (McGlone and Kawano 1998); however, there is also a potential for these tools to increase the efficiency and reduce the cost of phenotyping in breeding programs where there are large numbers of progeny generated from selected crosses (Cobb et al. 2013; Kumar et al. 2015). The Summerland Sweet Cherry Germplasm Development Program has been very successful at generating high-quality sweet cherries for commercial adoption in Canada and worldwide (Kappel and Lane 1998). The challenge for this and other germplasm development programs is that, over time, financial and resource allocations have declined and so there need to be shifts in protocols for quality assessment and phenotyping that will reduce the time, labour, and resources required to make assessments. One of the other limitations imposed by existing destructive methods is that each measurement requires that a number of fruit must be destroyed to allow physical or chemical analysis. This can restrict the number of different assessments that can be done when assessing new cultivar phenotypes at the first stage of selection, when only a limited number of fruit are available per seedling. Non-destructive sampling would expand the potential number of analyses performed on fruit in the first selection stage and if the component being measured can be verified to be associated with poststorage quality, it may also be used as a screening tool for selection.

Compositional analysis of sweet cherries shows that total dry matter content and soluble solids make up 80% of the dry matter content of the fruit (USDA 2016). Therefore, dry matter content can be considered to be mostly consisting of sugars in sweet cherry but also includes cell wall residues (dietary fiber), acids, and a very small amount of protein. It has been found that while NIR spectroscopy can be used to measure sugars and dry matter at harvest with other fruits, dry matter measures appear to provide a more consistent prediction of poststorage quality (Palmer et al. 2010; Kumar et al. 2015).

The experiments were conducted to develop a model to estimate dry matter in sweet cherries using a newly available VIS/NIR spectrometer and test how robust the resultant model was in estimating the dry matter of cherry cultivars that were not used to develop the model.

Materials and Methods

Fruit sampling

In 2016, 'Lapins' sweet cherries were harvested from the Summerland Research and Development Center (SuRDC) as well as from two commercial orchards also located in Summerland, BC. At each location, approximately 100 clean fruit were harvested randomly from several different positions of a single tree in early July 2016 to ensure a range in dry matter contents for the harvested cherries. After harvest, the sweet cherries were immediately transported to SuRDC where 75 large fruit of varying maturity stages, as described previously (Toivonen et al. 2004), were selected for analysis. Each sweet cherry was individually identified using a numbered tape label attached to the stem of the fruit. The absorption spectrum for each fruit was non-destructively measured using a VIS/NIR spectrometer after each fruit was acclimatized to one of three temperature-controlled rooms at 10 °C, 20 °C, and 30 °C, respectively. The cherries were allowed to equilibrate to that temperature for at least 4 h before spectral measurements were taken.

Absorption spectra were acquired with a FELIX F-750 produce quality meter (CID Bio Science, Inc., Camas, WA), an interactance optical design spectrometer having a wavelength range of 285–1200 nm (Felix Instruments 2015). The process of developing a predictive model with a spectrometer involves collecting measurements from a set of samples (i.e., "training set") from a population of fruit having a wide range in content of a particular constituent of interest. The spectrometer was placed upright on a laboratory benchtop with the measurement lens facing up. The instrument was positioned in the laboratory such that no direct light from the surrounding environment impinged on the lens area to avoid any stray light interference during measurement (i.e., it was positioned out of direct sunlight). The sweet cherries were removed from the conditioning temperatures and absorbance spectra were acquired from the side of the fruit directly opposite to the ventral suture line (Bentley 1873) at all three temperatures. The cherries were placed directly on the instrument lens. The spectral data from each sampling temperature was inputted into one of three temperature subsets (min, mid, or max) for model building and a sample number recorded that was unique to each fruit measured. After each sweet cherry from one temperature was analyzed, they were then placed into the next controlled-temperature room. At total of 75 fruit from each of three orchards were used in the "training set" and each was measured at each of the three temperatures, resulting in a total of 675 spectra being acquired for model building and internal validation, which are essential processes in building useful models (Nicolai et al. 2007).

Reference dry matter values were determined on the individually numbered fruit for which spectral data had

been acquired. The stem and pit were removed from each sweet cherry and the flesh was cut into quarters and placed into labelled 57 mm diameter aluminum dishes. Fresh weight was determined on a top-loading balance (model LE62025, Sartorius Canada Inc., Mississauga, ON) to an accuracy of two decimal places and the weighed samples were placed into a forced-air drying oven (model OMH100, Thermo Electron LED GmbH, Langenselbold, Germany) set at 65 °C for 80 h. Samples were removed from the oven in small groups and weighed immediately to prevent the absorption of moisture by the sample from ambient air before weight was recorded. The percent dry matter content was then calculated as (dry weight/fresh weight) × 100.

Building dry matter models

The “training set” was imported from the F750 instrument into the model-building software (F750 Model Builder, version 1.1.0.90, CID Bio Science). Once the “training set” had been imported, the reference dry matter data was appended to the spectral data set so that every spectrum had a corresponding reference dry matter value associated with it. The spectral correlation to the reference dry matter content was modeled using the software-selected default procedure wavelength range for dry matter, which happened to be a spectral range of 948–957 nm. Several other manually selected wavelength ranges were then tested in an iterative process, resulting in a more optimal second model being developed using a wavelength range of 858–1008 nm. This wavelength range coincided with wavelength ranges for dry matter modelling of other fruit (McGlone and Kawano 1998). The model-building software performed a partial least squares analysis to evaluate the optimal number of principle components and it highlighted obvious outlier data points that could be manually removed (one value was removed in this work). Root mean square errors, linearity, and prediction errors were generated within an internal validation procedure by the model-building software. The models were then evaluated and compared to determine if one was superior to the other.

External validation and model robustness testing

The resultant dry matter model for the manually selected wavelength range (858–1008 nm) was saved and uploaded to the F750 instrument. Three sets of sweet cherry fruit from different cultivars were harvested at SuRDC to externally validate the model and test its robustness (Nicolai et al. 2007). Fifteen clean, disorder-free fruit from each of the three different cultivars (Staccato™, Sentennial™, and Sovereign™) were used in this experiment. Each measurement was acquired from the side of the cherry opposite the ventral suture line as was the case in the model-building phase of this work, and then each fruit was dried and its dry matter content was determined as described above.

Statistical analysis

The model-building software for the F750 spectrometer is proprietary in nature but in principle is based on methodology common to spectrometric model development, which includes principle components analysis and use of the nonlinear iterative partial least squares algorithm (Nicolai et al. 2007). In essence, this analysis determines which absorption wavelengths are most influential for consistently predicting the measured variable (dry matter) and provides coefficients for a resultant model that can be inputted into the spectrometer to then subsequently directly predict the variable (i.e., the dry matter in the sample). The model-building software generated values to characterize each model and these values were (i) the root mean square error of cross-validation (RMSECV), which indicates how well a half of a data set fits a calibration model created using the other half of the data (a standard procedure in spectroscopic model building), (ii) the mean and standard deviation for dry matter values, (iii) the coefficient of correlation (*R*) between reference values and predicted values, and (iv) the relative prediction performance (SDR) of the model, which is calculated as standard deviation (SD) ÷ RMSECV (Nicolai et al. 2007). The dry matter content for ‘Lapins’ sweet cherries used for the calibration model were characterized with descriptive statistics using statistical procedures in Microsoft Excel, which generated the mean values, SD, and maximum and minimum values for fruit from each of the sampling sites. External validation the performance of the model for each of the three cultivars of sweet cherries, not used to develop the calibration model, was evaluated on the regression coefficient of measured versus predicted values and the root mean square error of prediction (RMSEP) using the PROC REG procedure of SAS version 9.1 (SAS Institute, Inc., Cary, NC).

Results and Discussion

Measured absorption spectra and assessment of model quality

The original sample size of 225 fruit had to be reduced to 220 due to some samples being contaminated while in the drying oven. This corresponds to 220 spectra per each sweet cherry at each of the three temperatures, resulting in a total of 660 spectra being used in the model. The dry matter content of the fruit as determined by destructive measurements ranged from 13.61% to 24.55% (Table 1). Therefore the range of dry matter values used to develop this model bracket values that would occur in sweet cherry cultivars developed by the breeding program at SuRDC as soluble solids contents and dry matter values in sweet cherries are found to be similar or within 1%–2% of each other (Kappel et al. 2002; Kovács et al. 2009).

The raw spectra for all ‘Lapins’ cherries sampled at the three temperatures, comprising the entire “training set”, are shown in Fig. 1, from which it is clearly evident that

Table 1. Dry matter data means, standard deviations, and maximum and minimum values for ‘Lapins’ sweet cherries collected in 2016 at three sites: the Summerland Research and Development Centre (SuRDC) and two local grower sites in the Summerland area.^a

Sample source	n	Dry matter (%)			
		Mean	Standard deviation	Maximum	Minimum
SuRDC	75	20.87	1.55	24.55	17.72
Grower 1	75	19.73	1.46	22.51	15.76
Grower 2	70	18.58	1.77	22.85	13.61

Note: While 75 cherries were used from SuRDC and the Grower 1 site, only 70 cherries were reported for Grower 2 as a few of the dry matter samples were compromised in the drying oven and those data were excluded from the training set.

^aThe process of developing a predictive model with a spectrometer involves collecting measurements from a set of samples (i.e., “training set”) from a population of fruit having a wide range in content of a particular constituent of interest.

Fig. 1. Second-order derivative transformed and smoothed VIS/NIR absorbance spectra of 225 ‘Lapins’ sweet cherries processed with the Savitzky–Golay procedure (Savitzky and Golay 1964). Spectra were collected at each of three temperatures (10 °C, 20 °C, and 30 °C) for each cherry, therefore the figure represents 675 spectra overlaid over each other. The spectra were collected in the intertance mode using a Felix F750 spectrometer.

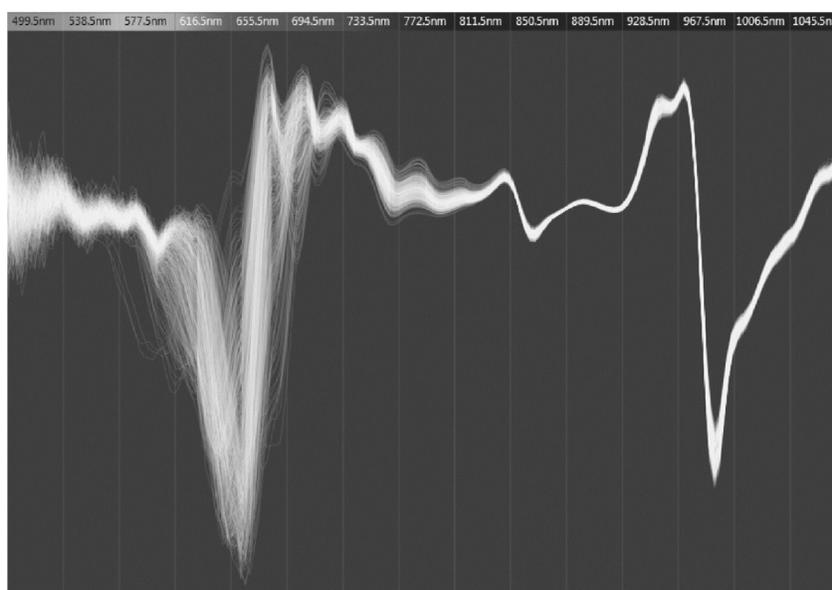


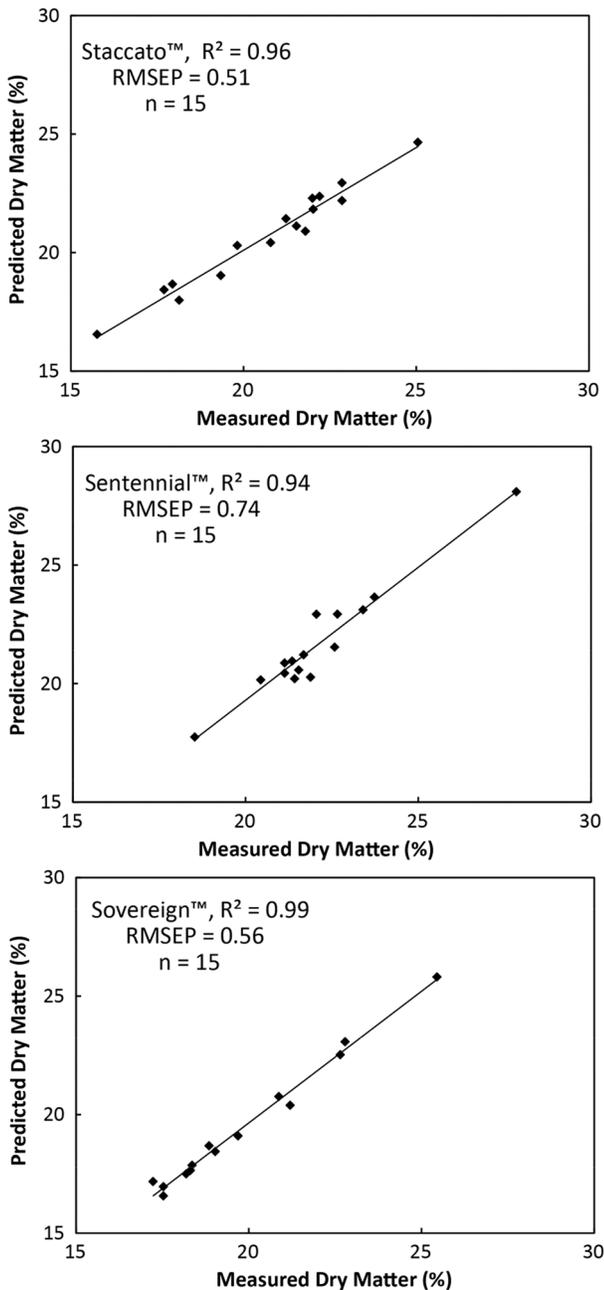
Table 2. Calibration and validation results for two models of dry matter of ‘Lapins’ sweet cherries based on the Savitzky–Golay second-order derivative of absorbance for two different spectral ranges, 948–957 and 858–1008 nm.

Model range (nm)	n	Outliers	Mean	SD	RMSECV	R	Terms	SDR
948–957	110	1	19.7	1.59	1.12	0.78	3	1.4
858–1008	110	1	19.7	1.59	0.53	0.95	4	3.0

the differences in spectra between individual samples in the VIS wavelength range (499–733 nm) were quite random and variable, likely because that band range contains absorption bands for chlorophylls and anthocyanins, both of which change significantly in concentrations as sweet cherry fruit ripens (Tijero et al. 2016). The

absorption bands in the NIR range were more regular and much less variable. This observation demonstrates why the VIS wavelength range of the absorption spectrum would not be useful for model development in sweet cherries. Despite the reduction in randomness and variability in the 948–957 nm absorption band range,

Fig. 2. Measured versus predicted values of the dry matter content (%) in Staccato™, Sentennial™, and Sovereign™ sweet cherries using a model developed for ‘Lapins’. The high R^2 values for all three cultivars indicates that the prediction value of the model is quite high, and the root mean square error of prediction (RMSEP) provides an indication of the robustness of the model for these cultivars when compared against the root mean square error of cross validation (RMSECV) reported for the model creation in Table 2.



the model built using this range was relatively poor (Table 2). The quality of the model is indicated by three parameters: RMSECV, the correlation coefficient (R), and the ratio of the standard deviation of the data set and

the RMSECV, which is termed the SDR. Most significantly, the R value was 0.78, which shows a moderate but not high correlation between measured and predicted dry matter content. More importantly, this model has an SDR of 1.4, which is considered to indicate that the model will not clearly discriminate between low and high values of dry matter (Nicolai et al. 2007). In contrast, the second model, based on the absorption spectral range of 858–1008 nm, resulted in a better quality model, having a much lower RMSECV value, a high R value of 0.95, and also a high SDR value, which is indicative of excellent prediction accuracy (Nicolai et al. 2007). The second model, made using a manually selected spectral range, was much stronger and of such quality that it is considered to be highly predictive of sweet cherry dry matter content. The likely reason for this second model being much better is that it is known that important carbohydrate absorption bands exist at 880, 900–930, and 970 nm (McGlone and Kawano 1998); therefore, these wavelengths should be included in a dry matter model.

External validation of model 2

It has been stated that the development of models to allow VIS/NIR spectroscopic determination of internal quality of fruit have suffered from lack of testing of the robustness of the model (Nicolai et al. 2007). One area where robustness is important when using this non-destructive tool in sweet cherry breeding is the applicability of the model to other cultivars of this fruit. Therefore model 2 was tested for its robustness in accurately predicting the dry matter of cultivars of sweet cherries other than ‘Lapins’. Staccato™, Sentennial™, and Sovereign™ sweet cherries showed coefficients of determination (R^2) all exceeding 0.90 (Fig. 2). More importantly for the analysis of the robustness of the models, RMSEP was very similar to the RMSECV of model 2, indicating that the model was quite robust for those two cultivars. Despite the fact that the RMSEP for Sentennial™ was higher than the RMSECV of the model for ‘Lapins’, the robustness for this cultivar may not be poorer than for the two other cultivars measured as it must be noted that the external validation sample population was very clustered and had one dry matter content value that exceeded the highest values used to construct the calibration model, whereas the sample populations for the other two cultivars were very well distributed within the range of dry matter values tested (Fig. 2). In our experience on using VIS/NIR spectroscopy to measure constituents in apples, when clustered sample sets are used, the RMSECV for the model or RMSEP for the external validation will increase.

Conclusions

Dry matter content prediction of sweet cherries by VIS/NIR spectroscopy is very accurate to discern

differences in dry matter content of approximately 0.5%, provided that an appropriate spectral window is chosen and a sample population for calibration is chosen such that it has a relatively uniform distribution of dry matter contents over a large range. Using a default range generated by a commercial instrument was not the most appropriate to build such a model, which indicates that a certain level of knowledge in the use of VIS/NIR spectroscopy is required to make the best decisions in developing the best calibration model to predict internal quality in fruit such as cherries. The data presented in this paper indicates that VIS/NIR spectroscopy has good potential to be used to determine dry matter contents in different cultivars of cherries and so may be useful for screening cherry selections for dry matter content in a breeding program. The next issue to be evaluated is verification as to whether dry matter content in sweet cherries equates to better poststorage or postshipping quality, as has been suggested for apples (Palmer et al. 2010; Kumar et al. 2015).

Acknowledgements

We thank the Thompson Rivers Co-operative Education Program for making A. Batista's internship possible and Agriculture and Agri-Food Canada's Agri-Innovations Program and the BC Fruit Growers' Association for funding the research.

References

- Bentley, O. 1873. A manual of botany including the structure, functions, classification, properties and uses of plants. 3rd ed. J. & A. Churchill, London, UK. 815 pp.
- Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., and McCouch, S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* **126**: 867–887. doi:10.1007/s00122-013-2066-0. PMID:23471459.
- Crisosto, G.M., Hasey, J., Zegbe, J.A., and Crisosto, C.H. 2012. New quality index based on dry matter and acidity proposed for Hayward kiwifruit. *Calif. Agric.* **66**: 70–75. doi:10.3733/ca.v066n02p70.
- Felix Instruments. 2015. Operation manual, F-750 produce quality meter. CID Bio Science, Inc., Camas, WA. 65 pp.
- Kappel, F., and Lane, W.D. 1998. Recent sweet cherry introductions from the breeding program at Summerland, British Columbia. *Acta Hort.* **468**: 105–110. doi:10.17660/ActaHortic.1998.468.10.
- Kappel, F., Toivonen, P., McKenzie, D.-L., and Stan, S. 2002. Storage characteristics of new sweet cherry cultivars. *HortScience*, **37**: 139–143.
- Kovács, E., Muskovics, G., and Perlak, R. 2009. Relationship of colour and other quality parameters of sweet cherry during development and ripening. *Acta Alimentaria*. **38**: 415–426. doi:10.1556/AAlim.38.2009.4.2.
- Kumar, S., McGlone, A., Whitworth, C., and Volz, R. 2015. Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. *Postharvest Biol. Technol.* **100**: 16–22. doi:10.1016/j.postharvbio.2014.09.021.
- McGlone, A.V., and Kawano, S. 1998. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol. Technol.* **13**: 131–141. doi:10.1016/S0925-5214(98)00007-6.
- Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., and Lammertyn, J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol. Technol.* **46**: 99–118. doi:10.1016/j.postharvbio.2007.06.024.
- Palmer, J.W., Harker, R.F., Tustin, S.D., and Johnston, J. 2010. Fruit dry matter concentration: a new quality metric for apples. *J. Sci. Food Agric.* **90**: 2586–2594. doi:10.1002/jsfa.4125. PMID:20718022.
- Savitzky, A., and Golay, M. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**: 1627–1639. doi:10.1021/ac60214a047.
- Tijero, V., Natalia, T., Muñoz, P., and Munné-Bosch, S. 2016. Implication of abscisic acid on ripening and quality in sweet cherries: differential effects during pre- and post-harvest. *Front. Plant Sci.* **7**: 602. doi:10.3389/fpls.2016.00602.
- Toivonen, P.M.A., Kappel, F., Stan, S., McKenzie, D.-L., and Hocking, R. 2004. Firmness, respiration and weight loss of 'Bing', 'Lapins', and 'Sweetheart' cherries in relation to fruit maturity and susceptibility to surface pitting. *HortScience*, **39**: 1066–1069.
- USDA. 2016. Basic report 09070, cherries, sweet, raw. United States Department of Agriculture National Nutrient Database for Standard Reference (Release 28, slightly revised May 2016). [Online]. Available at: <https://ndb.nal.usda.gov/>.
- Walsh, K.B., Golic, M., and Greensill, C.V. 2004. Sorting of fruit using near infrared spectroscopy: application to a range of fruit and vegetables for soluble solids and dry matter content. *J. Near Infrared Spectrosc.* **12**: 141–148. doi:10.1255/jnirs.419.