Canadian Journal of Plant Science
Revue canadienne de phytotechnie

**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.**

Journal:	<i>Canadian Journal of Plant Science</i>
Manuscript ID	CJPS-2017-0013.R1
Manuscript Type:	Article
Date Submitted by the Author:	24-Mar-2017
Complete List of Authors:	Toivonen, Peter; Summerland Research and Development Centre Batista, Adrian; Summerland Research and Development Centre Lannard, Brenda; Summerland Research and Development Centre
Keywords:	internal fruit quality, 'Lapins' sweet cherry, spectroscopy, model building

SCHOLARONE™
 Manuscripts

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.

Peter M.A. Toivonen¹, Adrian Batista^{1,2} and Brenda Lannard¹

¹Agriculture and Agri-Food Canada, Summerland Research and Development Centre, Summerland, British Columbia V0H 1Z0

² Co-operative Education Student, Chemical Biology Program, Thompson Rivers University, Kamloops, British Columbia

The 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. Absorption 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 to predict dry matter content in these other cultivars, with R^2 values of 0.96, 0.94 and 0.99, and RMSEP 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 cultivars cherries tested in this paper, a model developed using 'Lapins' fruit was reliable to predict dry matter in these other cultivars.

Keywords: internal fruit quality, 'Lapins' sweet cherry, spectroscopy, model building

Prediction and management of post-storage 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 post-storage 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). 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 in order to assess differences between cultivars and orchards and also to verify the association between postharvest quality and measured dry matter values at harvest.

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 use for these non-destructive tools to increase the efficiency and reduce the cost for 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 time, labor 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 in order to allow physical or chemical analysis. This can restrict the number of different assessments that can be done when assessing the new

cultivars phenotype 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 post-storage quality it may also be used as a screening tool for selection.

Compositional analysis of sweet cherries shows that of 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 near infrared spectroscopy can be used to measure sugars and dry matter at harvest with other fruits, dry matter measures appear to provide more consistent prediction of post-storage 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 visible/near infrared spectrometer and test how robust a resultant model was in estimating the dry matter of other cherry cultivars that were not used to develop that 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 located in Summerland, British Columbia. 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 visible/infrared spectrometer after each fruit was acclimatized to one of three temperature controlled rooms at 10, 20 and 30°C. At each temperature they were allowed to equilibrate to that temperature for at least 4 h before spectral measurements were taken.

Absorption spectra were acquired with FELIX F-750 Produce Quality Meter (CID Bio Science, Inc., Camas, WA, USA). It is 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 lab 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 removed from each of the three conditioning temperatures and absorbance spectra acquired on 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 set at one of the other three temperatures. A total of seventy-five 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 (Nicolaï 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, Canada) 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 hours. 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, $(\text{dry weight} / \text{fresh weight}) \times 100$.

Building dry matter models

The “training set” was imported from the F-750 instrument into the F750 Model Building software (F750 Model Builder, version 1.1.0.90, CID Bio Science, INC., Camas, WA, USA). 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 nm to 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 which 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 F-750 instrument. Three sets of sweet cherry fruit from different cultivars were harvested at the Summerland Research and Development Centre 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. Then each fruit was dried and dry matter content 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 (PCA) and use of the nonlinear iterative partial least squares (NIPALS) algorithm (Nicolai et al. 2007). In essence this analysis determines which absorption wavelengths are most influential for the 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; 1) 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), 2) the mean and standard deviation for dry matter values, 3) the coefficient of correlation (R) between reference values and predicted values, and 4) the SDR which is calculated as $SD \div RMSECV$ and this parameter indicates the relative prediction performance of the model (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 software which generated the mean values, standard deviations, 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 (Cary, NC, USA).

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 13.61-24.55 % (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 the Summerland Research and Development Centre since 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 Figure 1, it is clearly evident that the differences in spectra between individual samples in the visible 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 the 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 visible 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, the model built using this range was relatively poor (Table 2). The quality of the model is indicated by three parameters, the root mean square error of cross validation (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 a 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 an 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 970nm (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 visible/near infrared spectroscopic determination of internal quality of fruit have suffered from lack of testing for 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 (Figure 2). More importantly for the analysis of the robustness of the models, the root mean square error of the prediction (RMSEP) was very similar to the RMSECV of the model 2, indicating that the model was quite robust for those two cultivars. Despite the fact 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 measures since 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 (Figure 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 post-storage or post-shipping quality as has been suggested for apples (Palmer et al. 2010, Kumar et al. 2015).

ACKNOWLEDGEMENTS

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

Bentley, O. 1873. A Manual of Botany Including the Structure, Functions, Classification, Properties and Uses of Plants, 3rd Edition. J. & A. Churchill, London, 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.

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. *California Agriculture* **66**, 70-75. doi: 10.3733/ca.v066n02p70

Felix Instruments. 2015. Operation Manual, F-750 Produce Quality Meter, CID Bio Science, Inc., Camas, Washington, USA, 65 pp.

Kappel, F., Fisher-Fleming B., and Hogue E. 1996. Fruit characteristic and sensory attributes of an ideal sweet cherry. *HortScience*. **31**: 443-446.

Kappel, F. and Lane, W.D. 1998. Recent sweet cherry introductions from the breeding program at Summerland, British Columbia. *Acta Hort*. **468**: 105-110.

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 Alimen*. **38**: 415-426.

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.

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.

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.

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.

Peirs, A., Scheerlinck, N., Nicolaï, B.M. 2003. Temperature compensation for near infrared reflectance

measurement of apple fruit soluble solids contents. *Postharvest Biol. Technol.* **30**: 233-248.

Savitzky, A., and Golay, M., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**: 1627-1639.

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**. [online] <http://journal.frontiersin.org/article/10.3389/fpls.2016.00602> [24 Mar. 2017]

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. 2015. Fresh peaches and cherries: World markets and trade. Washington DC: Foreign Agriculture Service/United States Department of Agriculture.

USDA. 2016. Basic Report 09070, Cherries, sweet, raw. In: United States Department of Agriculture National Nutrient Database for Standard Reference, Release 28 slightly revised May 2016. Accessed on August 9, 2016 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 Infra. Spectro.* **12**: 141-148.

Table 1. Dry matter data means, standard deviations, 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. While 75 cherries were used from the SuRDC and Grower 1 sites, only 70 cherries were reported for Grower 2 since a few of the dry matter samples were compromised in the drying oven and those data were excluded from the “training set”¹.

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

¹ 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.

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

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

For Review Only

FIGURE CAPTIONS

Figure 1. Second order derivative transformed and smoothed visible/near infrared 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, 20, and 30 °C) for each cherry, therefore the figure represents 675 spectra overlaid over each other. The spectra were collected in the interactance mode using a Felix F750 spectrometer.

Figure 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 prediction is value of the model is quite high and the root mean square error of prediction (RMSEP) provides indication of 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 3.

Figure 1.

Figure 2.