Near-infrared (NIR) Spectrometric Technique for Nondestructive Determination of Soluble Solids Content in Processing Tomatoes

K.H.S. Peiris, G.G. Dull, R.G. Leffler, and S.J. Kays

Department of Horticulture, The University of Georgia, Athens, GA 30602-7273

ADDITIONAL INDEX WORDS. quality evaluation, internal quality, NIR

ABSTRACT. A nondestructive method for measuring the soluble solids content (SSC) of individual processing tomatoes (*Lycopersicon esculentum* Mill.) was developed using NIR spectrometry. A diode array fiber optic spectrometer was used to measure NIR transmittance. Each fruit was scanned at two locations on opposite sides midway along the proximal-distal axis. After scanning, each fruit was processed and pureed, and SSC was determined using a refractometer. Multiple linear regression (MLR), partial least squares (PLS) regression, and neural network (NN) calibration models were developed using the second derivatives of averaged spectra from 780 to 980 nm. The validation results showed that NN calibration was better than MLR or PLS calibrations. The NN calibration could estimate the processed SSC of individual unprocessed tomatoes with a standard error of prediction of 0.52% and could classify >72% of fruit in an independent population within $\pm 0.5\%$ of SSC.

Tomatoes represent the largest segment of the processed vegetable industry in North America, with the United States producing $\approx 48\%$ of the world's supply of processing tomatoes (Sullivan and Ravara, 1990). In 1997, for example, U.S. tomato processors contracted 121,650 ha yielding 9.3 million t of tomatoes (Anon, 1997).

The productivity of processing tomatoes is a function of the harvestable yield and total solids content of the fruit. Total solids are comprised of all fruit components except water and those volatilized during drying. The most common means of estimating solids content is by refractive index where the value obtained is the total soluble solids content (SSC) or natural tomato soluble solids. SSC values are reported as $^{\circ}\text{Brix}$ or percent and are highly correlated with total solids (Stevens, 1986). High SSC is positively correlated with processed product yield and negatively correlated with the energy cost for dehydration. The latter is an important component in the final product cost in that tomato processing is energy intensive (Singh et al., 1980). The economic benefit to be realized from genetically increasing SSC in processing tomatoes is substantial. For example, an increase in SSC of 1% (e.g., from 5% to 6%) represents a 20% increase in yield for products whose processed product yield is directly influenced by SSC (e.g., pastes, purees, sauces, etc.). Financially this translates into a \$70 to \$80 million a year savings in the United States alone (Wood, 1992).

SSC is also of prime concern in fresh-market tomatoes due to the important contribution of sugars and acids to the overall flavor of the fruit. Fresh tomato quality is a complex characteristic, involving not only physical appearance and firmness but also flavor and nutritive value. Simandle et al., (1966) showed that flavor quality ratings of six fresh tomato cultivars were significantly correlated with fruit SSC. Likewise, an increase in SSC would also improve the nutritional value of tomatoes. Jones and Scott (1983) suggested that tomato breeders interested in improved product quality should select for higher SSC and acid levels.

In tomato breeding, fruit solids content has received more attention than any other component (Stevens, 1986), thus elevated

SSC is a major goal in most processing tomato breeding programs (Berry and Uddin, 1991; Stevens, 1986; Tigchelaar, 1980) and is also considered important for fresh-market tomatoes. Therefore, all processing breeding lines and cultivars need to be assessed for SSC, even if the primary trait of interest is insect or disease resistance. One of the drawbacks to measuring SSC is that the product is generally processed before measurement and, with large numbers of lines, this is a time-consuming, labor-intensive process. In addition, the seeds (reproductive propagules for the line) are destroyed in the process. As a consequence, a rapid, nondestructive method for measuring tomato SSC could greatly enhance the evaluation efficiency in breeding programs allowing a far greater number of lines to be screened. A nondestructive method would also preserve the fruit for other analyses or seed extraction.

Total dry matter content in existing tomato cultivars is generally $\approx 4\%$ to 7.5% fruit mass but can approach 15% in certain wild species (Hewitt and Garvey, 1987). Soluble solids account for $\approx 75\%$ of the total dry matter and are comprised primarily of the reducing sugars (fructose and glucose), which represent 55% to 65% of the total SSC. Sucrose, in contrast, is present in very small quantities, rarely exceeding 0.1% on a fruit mass basis. A further 10% to 15% of the SSC consists of citric and malic acids, with the remaining amount made up of minor organic acids, amino acids, lipids, and minerals (Gould, 1983). Most of these organic compounds absorb NIR due to the presence of functional groups (e.g., R-OH, R-NH₂, R-CH, R-CH₂, and R-CH₃) on the molecules. It may therefore be possible to develop an analytical technique based on NIR absorption spectrometry to determine SSC of tomatoes nondestructively.

Nondestructive optical methods based on NIR spectrometry have been evaluated for the nondestructive estimation of internal SSC and dry matter content of a number of fruit and vegetable products including citrus (*Citrus* sp.) (Kawano et al., 1992), date (*Phoenix dactylifera* L.) (Schmilovitch et al., 1992), fresh-market tomato (Slaughter et al., 1995), mango (*Mangifera indica* L.) (Guthrie and Walsh, 1997), melons (*Cucumis melo* L. Inodorus and Reticulatus groups) (Dull et al., 1989a, 1992), onion (*Allium cepa* L. Cepa group) (Birth et al., 1985), papaya (*Carica papaya* L.) (Birth et al., 1984), peach [*Prunus persica* (L.) Batsch. var. *persica*] and nectarine [*Prunus persica* (L.) Batsch. var. *nucipersica* (Suckow) C. Schneider] (Peiris et al., 1997; Slaughter, 1995), pineapple [*Ananas comosus* (L.) Merr.] (Guthrie and Walsh,

Received for publication 18 Dec. 1997. Accepted for publication 12 June 1998. We thank Richard Ozminkowski, Jr., and Albert Liptay for providing the breeding lines and their constructive criticism of the manuscript. We also thank Betty Schroeder for assistance in laboratory analysis. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

Fig. 1. Distribution of SSC in tomatoes used in this study.

1997), and potato (*Solanum tuberosum* L.) (Dull et al., 1989b).

Slaughter et al. (1996) used a fiber optic probe interfaced with a commercial NIR spectrophotometer (model 6500; NIR Systems, Silver Spring, Md.) to develop a technique for estimating SSC in fresh-market tomatoes. They used an average of 250 scans, possibly due to the limitations in input energy, to get one spectrum from a single location and used averaged spectra from four equatorial locations to develop the calibration. At the scanning rate of 1.8 scans/s for this particular instrument, it thus takes nearly 10 min to scan

a single fruit. Practical applications, however, require instrumentation capable of analyzing constituents more rapidly.

Based on the successful use of NIR spectrometry for other products, our objective was to develop a rapid, nondestructive NIR absorption spectroscopic technique for determining the SSC of processed tomatoes using unprocessed, fresh individual tomatoes.

Materials and Methods

Seeds of 14 processing lines, provided by Heinz, USA, and Agriculture and Agri-Food Canada, Ontario, were grown at The University of Georgia Horticulture Research Farm, Watkinsville. Mature ripe fruit were hand harvested, immediately brought into the laboratory, washed, dried with a cloth towel, and numerically coded.

Based on preliminary work conducted in 1996 with 326 fruit, a protocol was developed for scanning tomatoes for spectrometric determination of SSC of the whole fruit. The critical information gained concerned the need for two spectra on each fruit, taken at the equator of the fruit and on opposite sides. It was also established that each spectrum should be comprised of 25 consecutive scans.

The instrument for scanning, a modified OM-200 onion meter (Agricultural Innovations, Athens, Ga.), had an Ocean Optics SD 1000-TR dual-beam fibre optic spectrometer (Ocean Optics, Dunedin, Fla.) operating in single-channel mode. A 65-W tungsten-halogen lamp with 45 W regulated input power was used as the source of NIR radiation. A pair of cutoff filters limited the wavelength range to 650 to 1000 nm. Radiation was directed toward the fruit using a 20-mm-internal-diameter metal tube. Individual fruit were held in place at the end of the light tube using a spring-loaded plunger. A 400- μ m fiber optic cable fitted with a collecting lens mounted on a moveable spring-loaded arm was positioned at the surface of the fruit at a fixed distance of 30 mm away from the center of the light tube. The spectrometer was operated at 30 kHz with a short integration cycle. About 12 s was required to obtain a spectrum averaging 25 scans from one side of each fruit. Reference spectra were taken using a Teflon block, which served as a neutral density scattering material. Data were

accessed and processed via the SpectraScope Spectral Acquisition software program (Ocean Optics).

After scanning, each fruit was individually processed to obtain puree for SSC determination. Each fruit was weighed, pierced with a fork several times to prevent exploding during cooking, placed in a small glass casserole covered with a lid, cooked in a microwave oven for 2 min, and cooled in an ice-water bath. After the fruit reached room temperature (20 °C), it was reweighed and water lost due to evaporation during cooking and cooling was added to the cooked fruit. The fruit was then pureed using a vegetable juicer and the percent SSC was measured on a sample using a refractometer (Abbey-3L; Bausch and Lomb, Rochester, N.Y.).

The absorbance data from 780 to 980 nm were used in all calibration model developments using averaged spectra (i.e., the average of two spectra recorded per fruit having one SS value). Mathematical transformation of raw optical density spectra to respective second derivative spectra was accomplished via Near Infrared Spectral Analysis Software (NSAS) (NIR Systems) using a segment size of 5 nm and gap size of 0 nm.

In addition to multiple linear regression (MLR) and partial least squares (PLS) regression calibrations, neural network (NN) calibration models were also developed. The second derivative absorbance values at each integer wavelength from 780 to 979 nm (200 data points) were used as input neurons to develop feed forward back propagation NN (Lawrence, 1994) calibration models using the Brainmaker Professional Neural Network Simulation Software program (California Scientific, Nevada City, Calif.). NNs were constructed using all 863 averaged spectra (613 spectra for training, 100 for testing, and 150 for validation data sets). Spectral selection was carried out randomly such that mean, standard deviation, and range of SSC values were nearly equal in all training, testing, and validation spectral data sets. NN models were developed with varying numbers of hidden neurons (Anon, 1993), and the best NN was selected based on its predictive powers using bias, standard error of prediction (SEP), and correlation coefficient (r) between laboratory measured and predicted SSC values. Bias was calculated as the difference between the means of predicted and laboratory determined SSC in the validation data set. SEP was calculated as the standard deviation of residuals (percent SSC predicted – percent SSC laboratory measured), while r (correlation

Fig. 2. Two optical density spectra of tomato having baseline variation with their respective second derivative spectra (upper lines).

coefficient of lab and NIR predicted SSC values) was calculated with a regression analysis of laboratory determined and NIR predicted SSC values. Statistics for NN models were calculated using data regression procedures and the mathematical and statistical functions of Lotus 1-2-3 software program (version 4.1, Lotus Development Corp., Cambridge, Mass.), since the Brainmaker program does not calculate these statistics.

Results and Discussion

The SSC of 863 fruit tested varied from 3.2% to 7.3% with a mean of 5.1% and a standard deviation of 0.7%. The population distribution of SSC is shown in Fig. 1. It shows that a representative population of fruit incorporating the natural variation of SSC found among processing tomato lines was used in the experiment.

A sample of optical density spectra with respective second derivative spectra is shown in Fig. 2. These spectra indicate that the instrument used in this experiment acquired optical density spectra of tomatoes with little spectral noise from 780 to 980 nm.

Regression calibration models were developed with MLR and PLS techniques using averaged second derivative spectra. According to preliminary investigations, the average of two individual spectra (i.e., two spectra recorded for the opposite side of each fruit that had the same laboratory measured SSC) gave better calibrations with higher multiple correlation coefficient (R) values and lower standard error of calibration (SEC) values than the models developed using individual spectra. MLR calibration had a R value of 0.80 with a SEC value of 0.67%. This MLR calibration predicted SSC in a validation data set ($n = 330$) with SEP = 0.79%, $r = 0.39$, and bias of 0.22%. Higher multiple correlation coefficients have been previously reported for MLR calibration models for a number of fruit and vegetable products [e.g., $R = 0.98$ for satsuma orange (Kawano et al., 1992), $R = 0.91$ for mango (Guthrie and Walsh,

1997), $R = 0.91$ for honeydew melon (Dull et al., 1992), $R = 0.98$ for onion (Birth et al., 1985), $R = 0.93$ for papaya (Birth et al., 1984), $R = 0.95$ for peach (Peiris et al., 1997), and $R = 0.92$ for potato (Dull et al., 1989b)]. Most of the products mentioned above have internal tissue with more or less uniform texture and, as a consequence, variation in NIR scattering by the tissues in these products may be more uniform than in tomatoes. In tomatoes there is great variation in fruit size, shape, texture, and the physical condition of the internal tissues. Fruit mass in this experiment ranged from 34 to 263 g, while the average fruit mass of individual lines ranged from 48 to 124 g. Fruit shape varied from round to cylindrical to pear shaped. Likewise, the number of locules per fruit ranged from two to seven among lines, while the locular contents varied from 14.6% to 24.4% of the fruit fresh mass. Stevens et al. (1977) reported that locular content of fresh-market

Fig. 3. The relationship between laboratory determined and NIR predicted soluble solids content (SSC) for multiple linear regression (MLR), partial least squares (PLS) regression, and neural network (NN) calibrations (left) and their respective validations (right).

Fig. 4. Cumulative percentage of fruit against absolute residual as classified by multiple linear regression (MLR), partial least squares (PLS) regression, and neural network (NN) calibrations.

tomato lines can range from 14.4% to 35.0%. The pericarp thickness varied from <4 to 9 mm depending on the line. In some lines, air spaces were observed between pericarp and locular tissues. In addition, the texture of pericarp, locular fluid, and seeds vary substantially. Previous studies on the composition (individual sugars, organic acids, and SSC) of pericarp and locular tissues have yielded inconsistent results (Stevens et al., 1977), indicating that the differences in composition, if present, depend on cultivar and environmental conditions. Given the wide variation in chemical and physical characteristics observed within and among tomato lines, it is probable that NIR light-scattering properties vary considerably within and among cultivars, increasing spectral noise over that which might be expected in products with uniform tissue density. Differences in bulk density and texture are two significant sources of errors in NIR reflectance measurements (Williams, 1987). Under such circumstances it is difficult to expect higher correlations in a calibration model with a few selected wavelengths as in the case of MLR models. Therefore, multivariate full-spectrum regression methods need to be tested with such data.

PLS is a multivariate full-spectrum regression technique, which was primarily designed for causal-predictive analyses of complex problems that are rich in data but scarce in theoretical knowledge (Frank et al., 1984). Hence, PLS seems more appropriate than MLR for such data as that obtained with tomatoes. Second derivative optical density data from 780 to 980 nm were used for PLS regressions.

It is important to determine the number of PLS factors that fit the model best. Too few factors result in imprecision since not all information, both needed and available, is used. Conversely, too many factors decrease the prediction precision due to the use of information not relevant to the model under investigation. To prevent overfitting, the PLS algorithm in NSAS calculates the mean squared error of cross-validations (MSECV) for each PLS

factor using one part of the sample in the calibration data set at a time. The number of PLS factors for the model is selected based on this MSECV value. Eight PLS factors were selected for the PLS calibration with a MSECV = 1.57%, $R = 0.92$, and $SEC = 0.46\%$. An independent data set was then used to validate the calibration models. Bias was very low (-0.05%), with $SEP = 0.74\%$ and $r = 0.57$.

NNs (Lawrence, 1994) allow the modeling of nonlinear data and have been shown in certain instances to be superior to MLR, PLS, or principal component regression (PCR) techniques in extracting spectral information (Borggaard, 1994; McClure et al., 1992). NNs were designed with varying numbers of hidden neurons (Anon, 1993) and the best NNs were selected based on their ability to predict SSC in independent samples, based on SEP, bias, and correlation coefficient (r) between actual and predicted values. The optimized NN used one layer of 30 hidden neurons and predicted 150 independent samples with bias = -0.03% , $SEP = 0.52\%$, and $r = 0.69$. The R and SEC values for the NN was 0.80% and 0.42% ($n = 613$); however, these calibration statistics do not translate, since with the progress of training the NN, R gradually increases as SEC decreases while the predictive power of the NN starts to diminish. The optimum number of runs to train the NN was determined using prediction statistics for a test file, periodically saved during the NN training process.

The relationship between the laboratory determined SSC and that predicted by MLR, PLS, or NN calibrations (left) and validations (right) is shown in Fig. 3. The solid line indicates the 1:1 line while the broken line is the regression line for the laboratory- and NIR-determined SSC values. The cumulative percentage of tomatoes in the independent validation sample of fruit plotted against the absolute difference between predicted and actual SSC for the three calibration models is presented in Fig. 4. These results indicate that NNs can predict SSC in whole tomatoes with less error than MLR or PLS calibration models.

These results demonstrate that there is a relationship between SSC of tomatoes and NIR absorbance that can be exploited to estimate SSC of individual intact tomatoes. The NN calibration was better than MLR or PLS models and can be used to determine SSC in >72% of fruit within $\pm 0.5\%$ of SSC.

The utility of this method lies primarily in two applications. Tomato breeders now have a tool for rapid and nondestructive sorting of individual tomatoes for high SSC. A given cultivar will contain tomatoes having a wide range of SSC. The SSC of the cooked, pureed fruit can be predicted using raw, intact fruit. Sampling by groups does not optimize the identification and sorting of high-SSC material. The identification of individual tomatoes by a method that is rapid enough to permit quantity sorting will greatly enhance a breeding program for high SSC. Secondly, for those applications where the identification of an average SSC value for a lot is needed (e.g., loads arriving at a processing plant), the average of the SSC for each of the tomatoes in the representative sample will be acquired more rapidly than in the conventional compositing method and will have the added accuracy, which the mean of individual values will permit.

Literature Cited

- Anon. 1997. Vegetables and specialties; Situation and outlook report. USDA Econ. Res. Serv.
- Berry, S.Z. and M.R. Uddin. 1991. Breeding tomato for quality and processing attributes, p. 197-206. In: G.A. Kaloo (ed.). Genetic improvement of tomato. Springer-Verlag, Berlin.
- Birth, G.S., G.G. Dull, J.B. Magee, H.T. Chan, and C.G. Cavaletto. 1984. An optical method for estimating papaya maturity. J. Amer. Soc. Hort.

- Sci. 109:62–66.
- Birth, G.S., G.G. Dull, W.T. Renfro, and S.J. Kays. 1985. Nondestructive spectrometric determination of dry matter in onions. *J. Amer. Soc. Hort. Sci.* 110:297–303.
- Borggaard, C. 1994. Modeling non-linear data using neural networks regression in connection with PLS or PCA. *Spectroscopy Europe* 6:21–27.
- California Scientific Software. 1993. Brainmaker Professional Neural Network Simulation Software: Users guide and reference manual. Calif. Sci. Software, Nevada City, Calif.
- Dull, G.G., G.S. Birth, D.E. Smittle, and R.G. Leffler. 1989a. Near infrared analysis of soluble solids in intact cantaloupe. *J. Food. Sci.* 54:393–395.
- Dull, G.G., G.S. Birth, and R.G. Leffler. 1989b. Use of near infrared analysis for the nondestructive measurement of dry matter in potatoes. *Amer. Potato J.* 66:215–225.
- Dull, G.G., R.G. Leffler, G.S. Birth, and D.E. Smittle. 1992. Instrument for nondestructive measurement of soluble solids in honeydew melons. *Trans. Amer. Soc. Agr. Eng.* 35:735–737.
- Frank, E.I., J. Feikema, N. Constantine, and B.R. Kowalski. 1984. Prediction of product quality from spectral data using the partial least-squares method. *J. Chem. Info. Computer Sci.* 24:20–24.
- Gould, W.A. 1983. Tomato production, processing and quality evaluation. AVI, West Port, Conn.
- Guthrie, J. and K. Walsh. 1997. Non-invasive assessment of pineapple and mango fruit quality using near infrared spectroscopy. *Austral. J. Expt. Agr.* 37:253–263.
- Hewitt, J.D. and T.C. Garvey. 1987. Wild sources of high soluble solids in tomato, p. 45–54. In: D.J. Nevins and R.A. Jones (eds.). *Plant biology*, vol. 4. Tomato biotechnology. A.R. Liss, New York.
- Jones, R.A. and S.J. Scott. 1983. Improvement of tomato flavor by genetically increasing sugar and acid contents. *Euphytica* 32:845–855.
- Lawrence, J. 1994. Introduction to neural networks: Design, theory and applications. Calif. Sci. Software Press, Nevada City, Calif.
- Kawano, S., T. Sato, and M. Iwamoto. 1992. Determination of sugars in satsuma orange using NIR transmittance, p. 387–393. In: I. Murray and I.A. Cowe (eds.). *Advances in near-infrared spectroscopy*. VCH, Weinheim, N.Y.
- McClure, W.F., M. Hana, and J. Sugiyama. 1992. Neural networks in NIR spectroscopy, p. 200–209. In: I. Murray and I.A. Cowe, (eds.). *Advances in Near-infrared spectroscopy*. VCH, Weinheim, N.Y.
- Peiris, K.H.S., G.G. Dull, R.G. Leffler, and S.J. Kays. 1997. Nondestructive determination of soluble solids content of peach by near infrared spectroscopy, p. 77–87. In: *Proc. Intl. Conf. Sensors for Nondestructive Testing; Measuring the Quality of Fresh Fruits and Vegetables*. NRAES, Ithaca, N.Y.
- Schmilovitch, Z., Z. Bernstein, M. Austerweil, A. Zaltzman, and G.G. Dull. 1992. Fresh date sorting by NIRS, p. 425–429. In: I. Murray and I.A. Cowe (eds.). *Advances in near-infrared spectroscopy*. VCH, Weinheim, N.Y.
- Simandle, P.A., J.L. Brogdon, J.P. Sweeney, E.O. Mobley, and D.W. Davis. 1966. Quality of six tomato varieties as affected by some compositional factors. *Proc. Amer. Soc. Hort. Sci.* 89:532–538.
- Singh, R.P., P.A. Carroad, M.S. Chhinnan, W.W. Rose, and N.L. Jacob. 1980. Energy accounting in canned tomato products. *J. Food Sci.* 45:735–739.
- Slaughter, D.C. 1995. Nondestructive determination of internal quality in peaches and nectarines. *Trans. Amer. Soc. Agr. Eng.* 38:617–623.
- Slaughter, D.C., D. Barrett, and M. Boersig. 1996. Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *J. Food Sci.* 61:695–697.
- Stevens, M.A., A.A. Kader, and M.A. Holton. 1977. Intercultivar variation in composition of locular and pericarp portions of fresh market tomatoes. *J. Amer. Soc. Hort. Sci.* 102:689–692.
- Stevens, M.A. 1986. Inheritance of tomato fruit quality components. *Plant Breeding Rev.* 4:273–311.
- Sullivan, G.H. and F.A. Ravara. 1990. Organization, structure and trade in the North American tomato processing industry. *Calif. Tomato Grower* 33(4):4–28.
- Tigchelaar, E.C. 1980. Breeding tomatoes for puree and paste. *Acta Hort.* 100:355–361.
- Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis, p. 143–167. In: P. Williams and K. Norris (eds.). *Near-infrared technology in the agricultural and food Industries*. Amer. Assn. Cereal Chem., St. Paul, Minn.
- Wood, M. 1992. Solid future for tomatoes. *Agr. Res.* 40(3):4–5.