

Nondestructive Determination of Solids and Carotenoids in Tomato Products by Near-Infrared Spectroscopy and Multivariate Calibration

André M. K. Pedro^{†,‡} and Márcia M. C. Ferreira^{*,†}

Department of Physical Chemistry, Chemistry Institute, Universidade Estadual de Campinas (UNICAMP), Campinas (SP), CEP 13083-970, Caixa Postal 6154, Brazil, and Latin American Foods Innovation Centre, Unilever Bestfoods Brazil, Valinhos (SP), 13271-450, Brazil

Tomato is an important player in the agricultural market. It is the second most consumed vegetable in the world and is a source of important micronutrients such as lycopene and β -carotene. Recent research has demonstrated that these carotenoids can act as free-radical quenchers in the body and prevent aging, tissue damage, heart disease, and certain cancers. Besides these micro-components, tomato is composed of soluble and insoluble solids. In industry, these solids govern factory yield and play a major role in the tomato trade. Nowadays, standard methods for determining tomato solids and carotenoids are time and labor consuming. In this work, we present the development of a simultaneous and nondestructive method for determining total and soluble solids, as well as lycopene and β -carotene, in tomato products by near-infrared spectroscopy. PLS-1 was the calibration technique chosen. For spectra preprocessing, MSC and second derivative were applied. As variable selection techniques, the correlogram cutoff, the successive projections algorithm, the dimension wise selection, and spectra splitting approach were applied. Best models presented satisfactory prediction abilities evaluated through its RMSEP and r values: total solids 0.4157, 0.9998; soluble solids 0.6333, 0.9996; lycopene 21.5779, 0.9996; β -carotene 0.7296, 0.9981.

Tomato (*Lycopersicon esculentum*) is the second most consumed vegetable in the world, right after potato.¹ The amounts involved in its production and commercialisation are impressive, corroborating its economical role within the agricultural business. In 2003, world production was ~110 million tons, and in Brazil, 3.3 millions tons were commercialised.^{2,3} Most of its volume is used by the food industry in the manufacture of products such as ketchup, tomato concentrates, and tomato sauce, but a significant volume is still purchased *in natura* by consumers.^{1,4}

The fruit is composed, broadly speaking, of tomato solids—soluble and insoluble—and water. Soluble solids are mainly sugars, like sucrose and fructose, and salts, like KCl, and are traditionally determined by refractometry and expressed as degrees Brix ($^{\circ}$ Brix).⁵ Insoluble solids are mainly constituted by fibers, such as cellulose and pectin. Total solids are the summation of soluble and insoluble solids and are determined by oven drying and expressed as percent in mass.⁵ Usually, excluding seeds and skin, tomato presents 4.5–8.5% total solids, depending on variety, soil, and climate conditions.^{1,4}

In industry, tomato solids dictate the factory yield: the highest the tomato solids amount the less tomato has to be used to produce processed tomato products. Besides, in industry, this parameter has to be controlled in many different production steps and thus it has considerable influence in the tomato trade.¹

Tomatoes are also important sources of healthy micronutrients such as carotenoids, mainly lycopene and β -carotene.^{6,7} While β -carotene is a provitamin A, recent research has shown that lycopene may prevent heart diseases, arteriosclerosis, and some types of cancer, including the prostate, lung, and colon.^{8–10} Synergies between carotenoids and other antioxidants, especially vitamins C and E, resulting in various healthy benefits have been reported but are still under investigation.^{11,12} The highly unsaturated structure of these carotenoids may also prevent aging and tissue damage by quenching free radicals formed by oxidative metabolic processes in the body.^{13,14}

There are several methods for determining carotenoids in tomato products reported in the literature, most of them based on solvent extraction followed by HPLC separation.^{15–18} On

- (4) Alencar, R. C. S. *O Tomateiro*; LTC: São Paulo, 1979.
- (5) Nielsen, S. S. *Food Analysis*, 2nd ed.; Gaithersburg Pub.: Aspen, 1998.
- (6) Fennema, O. R. *Food Chemistry*, 3rd ed.; Blakie Academic & Professional: London, 1996.
- (7) Isler, O. *Carotenoids*; Birkhäuser Verlag: Stuttgart, 1971.
- (8) Bramley, P. *Phytochemistry* **2000**, *54*, 233–236.
- (9) Tapiero, H.; Townsend, D. M.; Tew, K. D. *Biomed. Pharmacother.* **2004**, *58*, 100–110.
- (10) Willcox, J.; Catignani, G. L.; Lazarus, S. *Crit. Rev. Food Sci. Nutr.* **2003**, *43*, 1–18.
- (11) Fergusson, L. *Mutat. Res.* **1999**, *42*, 329–338.
- (12) Lee, A.; Thurnhan, D.; Chopra, M. *Free Radical Biol. Med.* **2000**, *29*, 1051–1055.
- (13) Giovannucci, E. *J. Natl. Cancer Inst.* **1999**, *91*, 317–331.
- (14) Jain, C. K.; Argawal, S.; Rao, V. *Nutr. Res.* **1999**, *19*, 1383–1391.
- (15) Sadler, G.; Davis, J.; Dezman, D. *J. Food Sci.* **1990**, *55*, 1460–1461.
- (16) Breithaupt, D. E. *Food Chem.* **2004**, *86*, 3, 449–456.

* To whom correspondence should be sent: (e-mail) marcia@iqm.unicamp.br.

[†] Universidade Estadual de Campinas.

[‡] Unilever Bestfoods Brazil.

- (1) Gould, W. A. *Tomato Production, Processing & Technology*, 3rd ed.; CTI Pub. Inc.: Baltimore, 1992.
- (2) United States Department of Agriculture (USDA) online, <http://www.fas.usda.gov>, 11/06/2004.
- (3) Brazilian Institute of Geography and Statistics (IBGE) online, <http://www.ibge.gov.br>, 11/06/2004.

average, it takes 40–50 min to have the results of one sample if the preparation time and the chromatographic run are considered.^{16–18}

Because of these health claims, tomato producers, research institutions, and the food industry are working together on breeding programs for the development of new “high lycopene” varieties that could be used in the manufacture of functional tomato-based foods as well as a less expensive lycopene source for the pharmaceutical and cosmetics industries. In such programs, success is usually determined by standard, time, and labor consuming analytical methods.

But as in breeding programs the number of screening varieties can be huge,^{1,4} the application of quick, solvent-free, and *on field* analytical techniques would present an enormous advantage over the conventional, wet chemistry-based approaches. NIR spectroscopy, allied to multivariate calibration techniques, fulfills the above requirements^{19–21}

Food chemistry was, indeed, one of the most benefited fields by the application of NIR spectroscopy for the determination of a series of properties in different food matrixes.^{22–27} The very first publication on tomato quality parameters is by Hong and Tsou.²⁸ More recently, Goula and Adamopoulos²⁹ have determined moisture, sugars, total acidity, salt, and protein in tomato products and Jha and Matsuoka³⁰ have calibrated the acid/Brix ratio, an important sensorial parameter of consumer acceptance, of various tomato juices. Nevertheless, carotenoids were never calibrated by near-infrared spectroscopy (NIR) or any other fast technique.

The objective of this work is to set up a method for determining, simultaneously and nondestructively, the contents of total and soluble solids, as well as the amounts of lycopene and β -carotene in tomato concentrate products via NIR and multivariate calibration techniques.

EXPERIMENTAL SECTION

Forty-two samples of tomato concentrate products, with total solids content varying from 6.9 to 35.9% (6.8 and 31.1 °Brix, respectively) were purchased in various markets from Brazil, Argentina, the United States, and Europe (The Netherlands, Italy, and Greece). These samples presented a broad range of variation for calibrating the properties of interest.

As tomato products are quite sensitive to molding, reference analyses and spectra acquisition (see below) were performed right after the packages' opening.

Reference Analysis. Total solids (%) were determined in triplicate by using a Fanem EV8 oven (Fanem, Co., São Paulo, Brazil). Approximately 3 g of sample was weighed, with 0.0001-g accuracy, in an aluminum capsule containing ~0.6 g of diatomaceous earth.³¹ After dispersion of the sample in the diatomaceous earth, aluminum capsules were kept at 70 °C under vacuum (~150 mmHg absolute pressure) using an Edwards E2M8 vacuum pump, until constant weight (~4 h).

Soluble solids were determined in duplicate by using an Abbe benchtop refractometer (American Optical, Inc., Baltimore, MD), with 0.1° Brix accuracy using temperature correction. All the readings were performed at room temperature (20–24 °C) after filtration through hydrophilic cotton.^{1,32}

Lycopene and β -carotene (mg kg⁻¹) were determined using a Shimadzu HPLC (Shimadzu, Co., Kyoto, Japan) equipped with a CTO-10A column oven, a Sil-10A automatic injector, LC-10AD pumps, and a SPD-10AV UV–visible detector at 473 nm. Separation was achieved using a RP18 Zorbax ODS column (5 μ m, 15 \times 0.46 cm). The mobile phase was MetOH/THF/H₂O (67:27:6), isocratic at 1.0 mL/min. Typical retention times were 17.5 min for lycopene and 20.8 min for β -carotene. Sigma-Aldrich standards were used to build analytical curves for lycopene (Catalog No. C 0251) and β -carotene (Catalog No. L 9879).

For the extraction procedure, the method suggested by Sadler et al.¹⁵ was applied. A 5-g sample was weighed, with 0.0001-g accuracy, in a 250-mL Erlenmeyer flask; 100 mL of solvent mixture hexane/acetone/ethanol (50:25:25) was added. The Erlenmeyer flask was kept away from light while the solvent mixture was kept in contact, under agitation by a magnetic stirrer, with the sample. After 10-min agitation, 10 mL of distilled water was added, and the mixture was kept under agitation for 5 min more. Agitation was then turned off, and the solution was allowed to separate into the organic and aqueous phases. An aliquot of the hexane phase was immediately collected to a 2-mL vial with silicon septa and 5 μ L was injected.

NIR Spectra Acquisition. The NIR spectra were acquired immediately after opening of the samples. A suitable aliquot of the sample was added to the bottom of a glass Petri dish (Schott 23 755 48 05), and readings were performed in a Büchi NIRLab N-200 spectrometer (Büchi Labortechnik AG, Postfach) equipped with a MSC-100 diffuse reflectance cell with sample rotation system.³³ This rotating system promotes a better sampling surface and also prevents local heating of the sample by the infrared radiation. Spectra were acquired at room temperature (20–24 °C).

Petri dish infrared radiation absorption was evaluated using MgO (Merck 105.866) as reference material and was considered negligible.

Three spectra were collected, each one using 100 scans in the 4000–10 000 cm⁻¹ range, with 4 cm⁻¹ of resolution.

- (17) Fish, W. W.; Perkins-Veazie, P.; Collins, J. K. *J. Food Compos. Anal.* **2002**, *15*, 309–317.
- (18) Oliver, J.; Palou, A. *J. Chromatogr., A* **2000**, *881*, 543–555.
- (19) Martens, H.; Naes, T., *Multivariate Calibration*; John Wiley & Sons: Chichester, 1993.
- (20) Williams, P., Norris, K., *NIR Technology in the Agricultural and Food Industries*, Am. Ass. Cereal Chemists: St. Paul, 1990.
- (21) Pasquini, C. *J. Braz. Chem. Soc.* **2003**, *14*, 198–219.
- (22) Chen, J. Y.; Iyo, C.; Kawano, S. *J. Near Infrared Spectrosc.* **1999**, *7*, 265–273.
- (23) Lammertyn, J.; et al. *Trans. ASAE* **1998**, *41*, 1089–1094.
- (24) Jha, S. N.; Matsuoka, S. *Food Sci. Technol. Res.* **2000**, *6*, 248–251.
- (25) Kawano, S.; Fujiwara, T.; Iwamoto, M. *J. Jpn. Soc. Hortic. Sci.* **1993**, *62*, 465–470.
- (26) Osborne, S. D.; Kunenmyer, R.; Jordan, R. B. *J. Near Infrared Spectrosc.* **1999**, *7*, 9–15.
- (27) Kawano, S.; Watanabe, H.; Iwamoto, M. *J. Jpn. Soc. Hortic. Sci.* **1992**, *61*, 445–451.
- (28) Hong, T. L.; Tsou, S. C. S. *J. Near Infrared Spectrosc.* **1998**, *6*, A321–A324.
- (29) Goula, A. M.; Adamopoulos, K. G. *J. Near Infrared Spectrosc.* **2003**, *11*, 123–136.
- (30) Jha, S. N.; Matsuoka, T. *Int. J. Food Sci. Technol.* **2004**, *39*, 425–430.

- (31) Diatomaceous earth is only used as a dispersive medium. It has no specific features and can even be replaced by acid-treated, washed sand.
- (32) Hydrophilic cotton has no specific specifications. It is applied as a fast filtration medium, and thus, commercial products can be purchased.
- (33) More information about Büchi NIR spectrometer can be found at <http://www.buchi.com>.

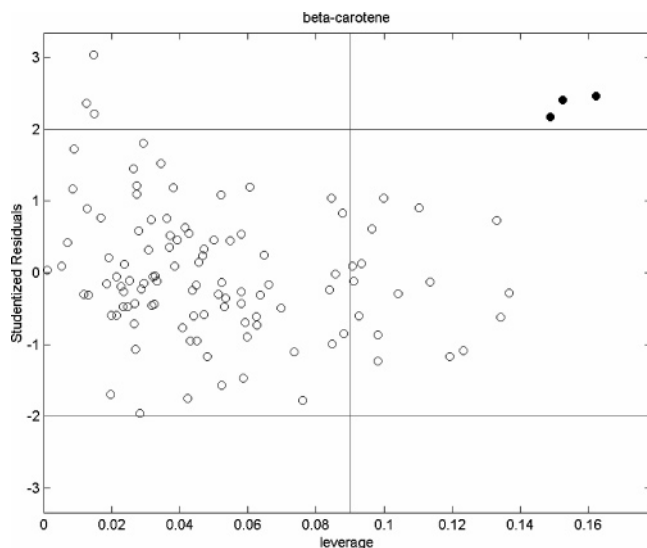


Figure 1. Leverage vs studentized residuals chart. The horizontal lines represent limiting studentized residuals with 95% significance while the vertical line represents a critical leverage value. The three highlighted spectra are from the same sample and were removed as outliers for the construction of the final model.

Multivariate Calibration. Following the traditional convention in linear algebra, in this work vectors are represented by boldface lower case type, matrices by boldface upper case type, scalars by italic lower case letters, and sequences by italic subscripts.

Partial least squares (PLS-1) was the regression method elected to build the calibration models. Original spectra were preprocessed by applying a mean smoother with a window width of 15 wavenumbers, followed by multiplicative signal correction (MSC)³⁴ or second derivative according to the algorithm proposed by Savitsky and Golay. Data (spectra and property vector) were mean-centered prior to calibration.^{19,35–38}

The 42 original samples (126 spectra) were split into two groups. Five samples (15 spectra) were separated for external validation and the least 37 samples (111 spectra) were used for calibration. The samples presenting extreme values for each property were included in the calibration set.

Cross-validation following the leave-one-out procedure was performed during the validation step in order to define the optimum number of factors to keep in the model and to detect any outliers. Leverage (eq 1) and studentized residuals (eq 2) were calculated and plotted for each sample (Figure 1). Samples presenting high leverage and studentized residuals that had significant detrimental effects on the model were considered outliers and removed from the model.^{19,37}

$$h_i = \frac{1}{N} + (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (1)$$

where h_i corresponds to the leverage value of the i th sample and \mathbf{X} is the matrix containing the spectra. \mathbf{X} has N lines, corresponding to the number of samples, and K columns, corresponding to

the wavenumbers. \mathbf{x}_i is the spectra for the i th sample, and $\bar{\mathbf{x}}$ is the average spectra.

$$Lresc_i = \sqrt{\frac{(y_i - \hat{y}_i)^2}{(N-1)(1-h_i)}}$$

$$SR_i = \frac{(y_i - \hat{y}_i)}{Lresc_i \sqrt{(1-h_i)}} \quad (2)$$

in which $Lresc_i$ corresponds to the residual for the i th sample, standardized by its leverage value, SR_i stands for the studentized residual, and y_i and \hat{y}_i are, respectively, the measured and the estimate values of the property y for the i th sample.

External validation was performed and the models with best predictive abilities, considering lowest RMSEP (eq 3) and higher validation correlation coefficient (eq 4), were taken.^{19,37–39}

$$RMSEP = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

$$r = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sqrt{s^2(y) \times s^2(\hat{y})}} \quad (4)$$

where RMSEP is the root-mean-squared error of prediction, r is the correlation coefficient between the estimated and the predicted values, and $s^2(y)$ and $s^2(\hat{y})$ are, respectively, the variances for the measured and predicted values for the property y .

Variable Selection Techniques. To find the very best models for predicting the properties of interest, four different variable selection techniques were applied to the preprocessed spectra: correlogram cutoff, the successive projections algorithm (SPA), the dimension wise selection (DWS), and the splitting of the preprocessed spectra.^{19,20,40}

Correlograms are charts that bring the values of the correlation coefficients of each variable with a specific property. Correlograms using the calibration spectra were built for each property, and an absolute value of correlation coefficient was taken as a cutoff parameter (Figure 2a). Only variables with corresponding correlation coefficients larger than this parameter were taken for modeling (Figure 2b). The procedure was repeated for different cutting-off criteria, and the optimum was the one that gave the model its best predictive ability (Figure 2c).

The DWS consists of neglecting small values in the weights vector \mathbf{w} during the factors extraction step in the PLS algorithm. In this case, various cutoff criteria are tested and the one that gives the best RMSEP value is kept. This approach can be geometrically interpreted as a rotation of the regression vector toward the solution with the best fitting.¹⁹

The SPA aims to find variables that are more orthogonal among themselves by finding the ones that give vectors with the

(34) Isakson, T.; Naes, T. *Appl. Spectrosc.* **1988**, *42*, 1273–1286.

(35) Wold, S.; et al. *Chemom. Intell. Lab. Syst.* **2001**, *52*, 131–150.

(36) Thomas, E. V. *Anal. Chem.* **1994**, *66*, 795A–803A.

(37) Ferreira, M. M. C.; et al. *Quim. Nova* **1999**, *22*, 724–731.

(38) Wold, S. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

(39) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; John Wiley & Sons: New York, 1978.

(40) Araújo, M. C. U.; et al. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 65–73.

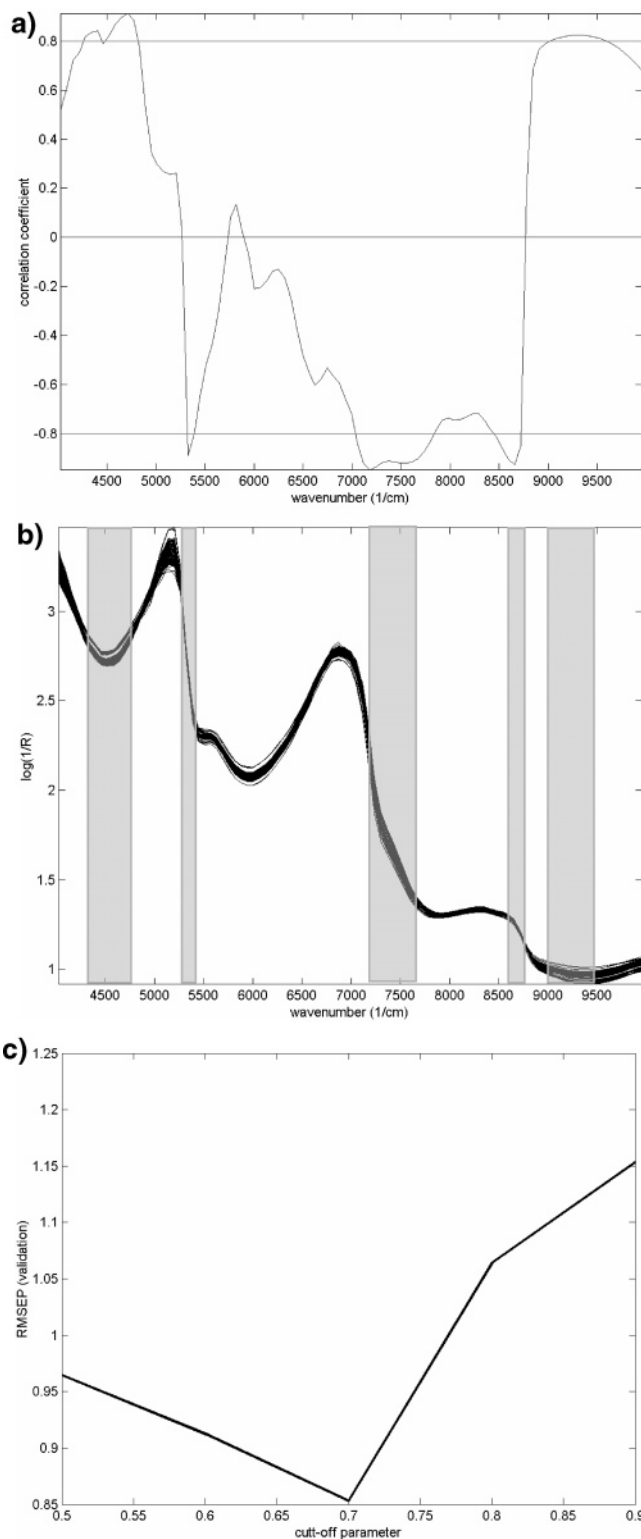


Figure 2. Example of cutting-off procedure for variable selection based on the correlation coefficient. (a) Variables with absolute correlation coefficient smaller than 0.8 (cutoff) are neglected and thus (b) only wavenumbers within the dark regions were used in the calibration model. (c) RMSEP vs cutoff chart. The best cutoff criterion in this example would be 0.7.

highest norms when projected in orthogonal spaces. The variable for starting the projections, as well as the optimum number of variables to extract from the original set of wavenumbers, are systematically determined for each property.⁴⁰

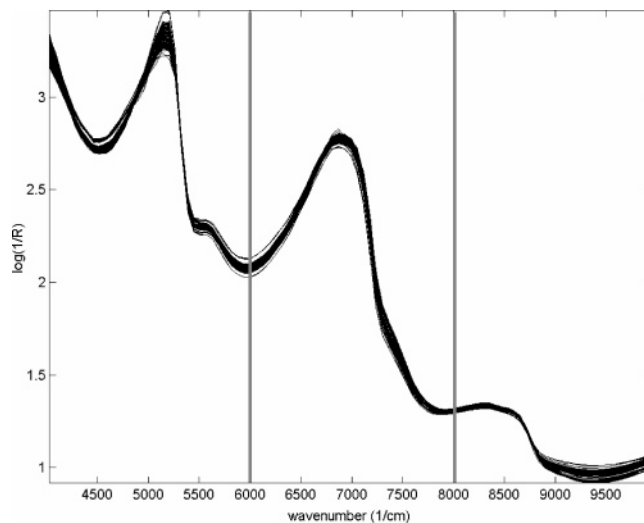


Figure 3. Split spectra as variable selection method for calibration.

In the splitting approach, which might be understood as a particular case of the intermediate PLS algorithm (iPLS),⁴¹ the spectra were divided into three equal regions and calibrations performed in each one as well as in combinations of them. Region one ranged from 4000 to 6000 cm⁻¹, region two from 6000 to 8000 cm⁻¹, and region three from 8000 to 10 000 cm⁻¹. Notice that each region accounts for a feature of the spectra—one peak or one peak and a shoulder (Figure 3).

All the calculations were performed in the Matlab Software v. 6.1 (The MathWorks, Co., Natick, MA) using routines implemented by the authors. The results were compared with those given by the chemometric software NirCal v. 4.21⁴² supplied with the NIR instrument. As similar results were obtained, herein only Matlab results will be presented.

RESULTS AND DISCUSSION

Figure 4a shows the 126 original spectra collected for the 42 samples of tomato products. Considerable noise is present in the regions between 4000 and 5500 and 6300–7100 cm⁻¹. The spectra were quite homogeneous, and no outliers were identified *a priori* by visual inspection.

Consistent baseline offsets and bias were present. These are quite common features in NIR spectra acquired by diffuse reflectance techniques.^{20,43} Nevertheless, it was observed that the offset showed some correlation with the amount of water present in the samples: the higher the sample's moisture, the higher its spectral offset. This might be due to the summation of (a) differences in the path length of the infrared radiation because the reduced presence of particulate material allowed the light to penetrate deeper into the bulk sample and (b) due to differences in the scatter profile of the samples and the reference during the reflectance measurements.

Noise and systematic behavior are undesirable features in the spectra. To solve this, the original spectra were preprocessed by a mean smoother (noise and variable number reduction, Figure

(41) Lindgren, F.; et al. *J. Chemom.* **1994**, *8*, 349–363.

(42) Büchi Labortechnik. *Büchi Nirlab Chemometrics Software Manual*; supplied together with NIRLab N-200 and NirCal Software.

(43) Olinger, J. M.; Griffiths, P. R. *Appl. Spectrosc.* **1993**, *47*, 695–701.

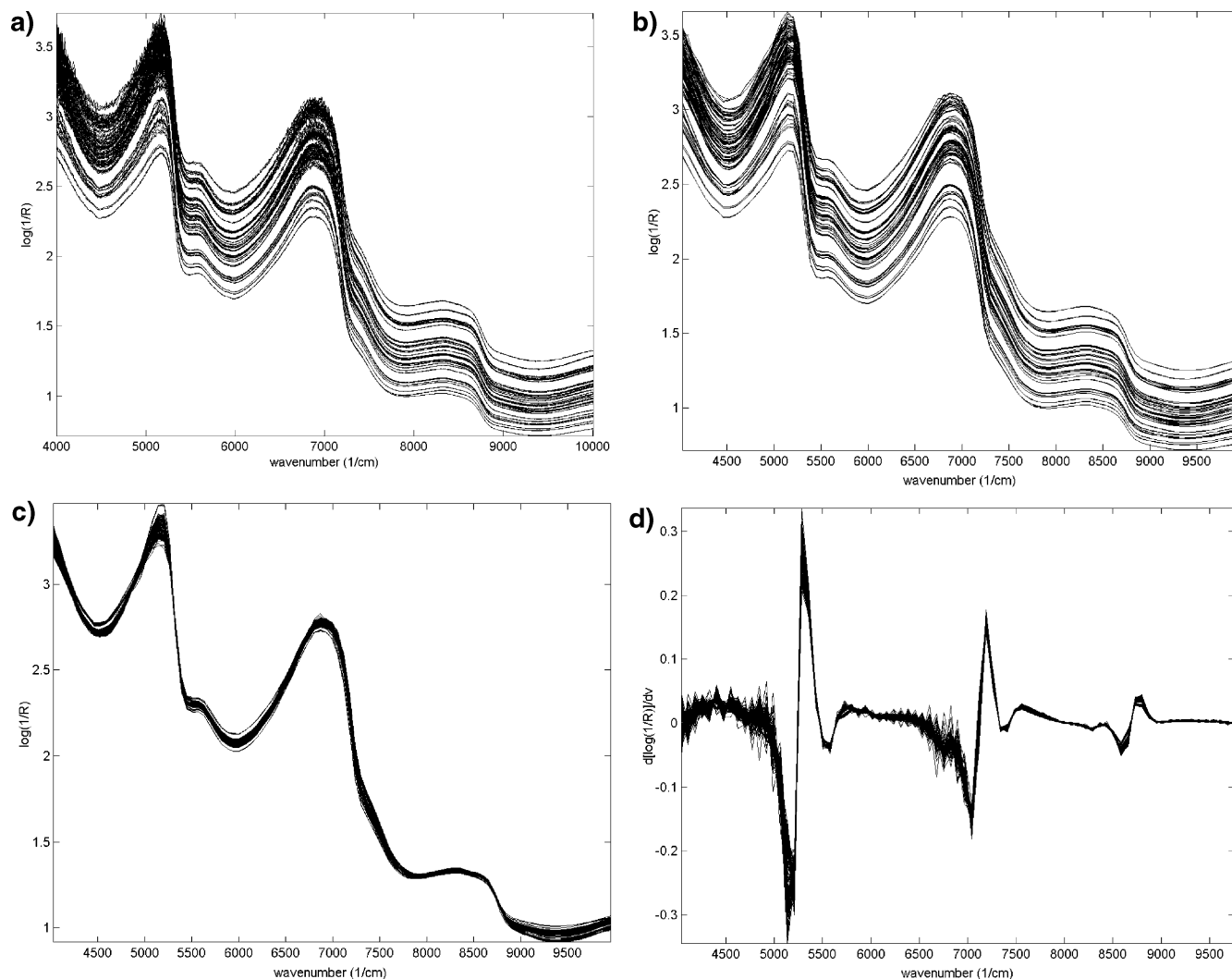


Figure 4. (a) Original spectra of the tomato products; (b) smoothed spectra by average, using a 15-points window. Preprocessed spectra by (c) MSC and (d) second derivative after smoothing.

Table 1. Calibration Models for Tomato Properties Using the Preprocessed, Full Spectra

property	model no.	pretreatment	factors	RMSEP	r_{val}	outliers
total solids (%)	1	MSC	10	1.0652	0.9989	1
	2	2nd derivative	13	0.9552	0.9991	0
soluble solids (° Brix)	3	MSC	10	0.5827	0.9996	1
	4	2nd derivative	13	0.9530	0.9989	0
lycopene (mg kg ⁻¹)	5	MSC	8	36.9868	0.9991	1
	6	2nd derivative	14	44.9693	0.9986	1
β -carotene (mg kg ⁻¹)	7	MSC	5	0.7296	0.9981	3
	8	2nd derivative	12	1.9113	0.9867	0

4b), followed by MSC (offset reduction, Figure 4c), or second derivative (offset and bias removal, Figure 4d).

It can be seen that the smoothing window chosen did not eliminate any important feature of the spectra, and thus, all the relevant chemical information was retained for modeling. The number of variables was also reduced from 1557 to 97 wavenumbers. It is also noticeable that the second-derivative preprocessing, despite removing bias in the baseline, has inserted considerable noise to the spectra, a feature that, as will be seen, resulted in poor calibration models.

The preprocessed full spectra (no variable selection) were submitted to the PLS-1 calibration for total solids, soluble solids,

lycopene, and β -carotene. Table 1 shows the results.

Reasonable models were obtained by using the full spectra. However, the best models were mostly given by the MSC-processed spectra (odd models), the only exception being model 2 which is a more complex model. This effect is probably due to the differences in the signal-to-noise ratio on both preprocessing procedures.

PLS calibrations were also performed after the application of the variable selection approaches stated above. Table 2 lists the results for the splitting procedure.

Results using fragments of spectra presented better predictive abilities than their counterparts with full spectra, the

Table 2. Calibration Models for Tomato Properties Using the Spectra Splitting approach

property	mode no.	pretreatment	regions	factors	RMSEP	r_{val}
total solids (%)	9	MSC	1 and 3	10	0.4157	0.9998
	10	2nd derivative	2	13	0.8829	0.9992
soluble solids (° Brix)	11	MSC	2 and 3	11	0.6333	0.9996
	12	2nd derivative	2 and 3	13	0.8634	0.9991
lycopene (mg kg ⁻¹)	13	MSC	1 and 2	5	21.5779	0.9996
β -carotene (mg kg ⁻¹)	14	2nd derivative	2	5	38.0153	0.9987
	15	MSC	1 and 2	5	0.7455	0.9981
	16	2nd derivative	1	6	1.5878	0.9907

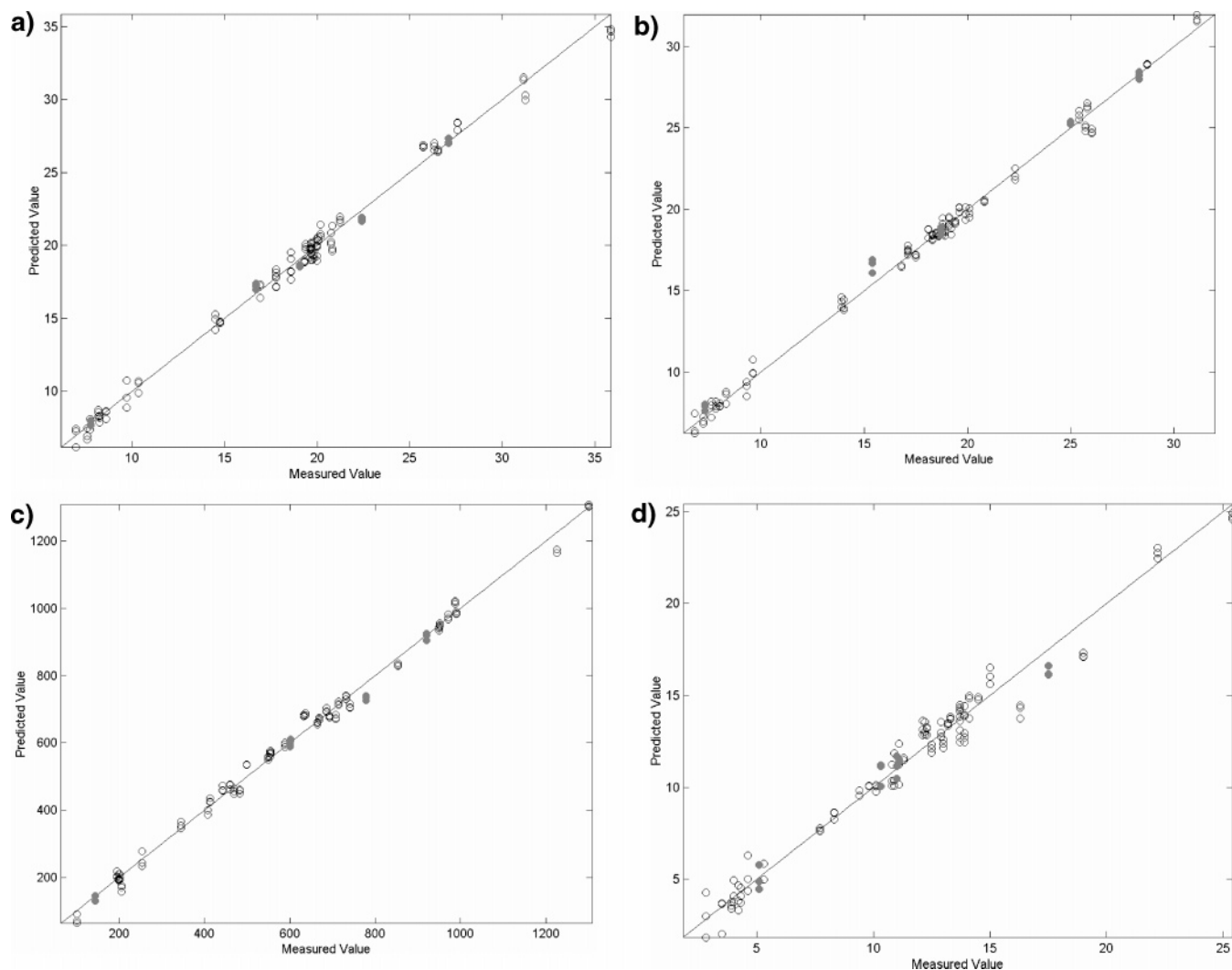


Figure 5. True vs predicted values for (a) total solids (model 9), (b) soluble solids (model 11), (c) lycopene (model 13), and (d) β -carotene (model 7) obtained by the best calibration models. Black open dots represent calibration spectra and solid dots represent the external validation spectra.

only exception being β -carotene. For this property, model 7 has shown better prediction for the external samples than model 15. Besides, second-derivative preprocessing stills presents no good models.

The other selection procedures outlined above presented models with no better predictive abilities than those given by the splitting approach. The DWS procedure presented models with exactly the same features as the ones obtained without its use. Indeed, the regression vectors presented quite subtle changes. As this procedure can be seen as a rotation of the regression vector toward the direction of best fitting, it can be concluded

that the solutions of Table 1 were already oriented or quite close to this direction.

With the SPA, better models were not found probably because this method is based on the search of variables that are more orthogonal among themselves as possible, but they do not necessarily have the best correlation with the property of interest.

Regarding the correlograms, good models, but not better than the ones given by the splitting approach, were obtained only with cutoff parameters such as 0.2 or 0.4. As the cutoff parameter gets smaller, the models tend to the ones obtained with the full spectra.

Table 3. Best Models Absolute and Relative Errors for the Validation Samples for Each Property

property	model no.	sample no.	true value	pred value	ab error	% error
total solids (%)	9	1	7.74	7.78	-0.04	0.57
		2	16.80	17.18	-0.50	3.00
		3	22.40	21.81	0.58	2.63
		4	19.06	18.63	0.42	2.24
		5	27.10	27.13	-0.03	0.11
soluble solids (° Brix)	11	1	7.3	7.9	-0.6	7.76
		2	15.4	16.6	-1.2	7.61
		3	28.3	28.2	0.1	0.27
		4	18.7	18.7	0.0	0.17
		5	25.0	25.3	-0.3	1.11
lycopene (mg kg ⁻¹)	13	1	144	135	9	6.45
		2	602	608	-6	1.03
		3	601	596	5	0.75
		4	779	734	44	5.69
		5	921	916	5	0.49
β -carotene (mg kg ⁻¹)	7	1	5.1	5.0	0.1	1.50
		2	11.1	11.4	-0.3	3.11
		3	11.0	11.1	-0.1	0.93
		4	10.3	10.8	-0.5	4.79
		5	17.5	16.3	1.19	6.83

Once the best models were found, the absolute and relative errors for the five prediction samples were calculated (Table 3). The average values of prediction for each of the three individual spectra were taken per sample.

Absolute and relative errors are within the acceptable experimental ranges for each property; they are also comparable to those presented by the reference methods.

Figure 5 brings the measured versus predicted charts for each property. The diagonal line represents ideal results (predicted = measured value) and so the closer the points are to this, the best is the model. Black open circles represent calibration spectra and solid circles represent validation spectra.

CONCLUSIONS

Calibration models relating spectral characteristics of samples from tomato products with chemical composition, regarding solids and carotenoids, were successfully built using relatively simple models and variable selection techniques.

Multiplicative signal correction was the most effective preprocessing technique in the sense that it produced all the best models for calibrating tomato properties. Splitting the spectra into three distinct regions, a quite simple approach when compared to the other techniques exploited, presented optimum calibration models for most of the properties, the only exception being β -carotene.

The models here presented are already in use at Unilever Brazil and performing quite well. As a good laboratory practice, regular evaluations of its performance are being conducted. No adjustments were necessary.

In the future, these models can be expanded to include raw tomatoes, enabling them to be applied to breeding programs for speeding up the development of new high-solids or high-carotenoids varieties.

Received for review September 12, 2004. Accepted January 26, 2005.

AC048651R