Determination of quality parameters of tomato paste using guided microwave spectroscopy

Lu Zhang [a], Michelle A. Schultz [b], Rick Cash [b], Diane M. Barrett [a], Michael J. McCarthy [a,c,*]

[a] Department of Food Science and Technology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
[b] Thermo Fisher Scientific, Minneapolis, MN 55433, USA
[c] Department of Biological and Agricultural Engineering, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

ABSTRACT

Guided microwave spectrometry was evaluated for in-line determination of quality attributes, soluble solids content, titratable acidity and Bostwick consistency of tomato pastes in continuous processing. The spectrometer was tested for two processing operations, dilution and evaporation. A total of 34 tomato paste samples were prepared by blending 9 different tomato pastes together in various proportions with final Bostwick consistency ranging from 2 to 11 cm. During the processing operation, spectra of 5 samples were obtained when different soluble solids contents were achieved. At the same time, standard reference methods were used to measure the parameters of interest on the same samples taken from the processing flow loop. The spectrum was recorded from 136 to 2690 MHz in 2 MHz steps. Partial Least Square (PLS) was employed to produce prediction models for each quality attribute from the collected microwave spectra. Cross-validation was performed to validate the calibration models. These models demonstrated good predictability of the soluble solids content, titratable acidity and Bostwick consistency in both dilution and evaporation conditions, as shown by their high correlation coefficient (r) and low root mean square error of cross-validation (RMSECV) values. Interval PLS (iPLS) was applied to optimize the models by selecting the most relevant and non-noisy regions in the spectra. The iPLS models showed improved accuracy for prediction of all three of the parameters, especially for Bostwick consistency ($r > 0.92$ and RMSECV < 0.7 cm). These observations suggest that guided microwave spectroscopy gave accurate estimation of soluble solids content, titratable acidity and Bostwick consistency of tomato paste. Different processing conditions required varying prediction models for the changes in the dielectric properties of tomato pastes.

1. Introduction

The majority of the processing tomatoes grown in California are thermally processed into concentrated tomato paste. The concentrated pastes are usually stored and used as an intermediate product with water and other ingredients to be reconstituted into final products, such as ketchups and sauces. Since tomato paste is the main ingredient in the final products, maintaining the quality of the paste is crucial for the tomato processing industry. Factors like the cultivar of the tomato and the processing conditions introduce great variation in the quality of the paste (Anthon, Diaz, & Barrett, 2008). These variations pose difficulties in achieving a consistent

paste quality during production. As a quality control routine, pastes are sampled from each production batch and a range of quality factors are tested; these include color, soluble solids content, pH, acidity and consistency. Because the classification of tomato products, e.g., paste or puree, is done according to its soluble solids content, it is a key characteristic of tomato products. During evaporation, the real-time value of the soluble solids helps to determine the end point of the processing. In addition, the flavor of tomato paste is closely associated with the soluble solids content and titratable acidity.

Among all factors, consistency is a major quality component of tomato paste, sauce and ketchup, determining the rheological properties of the intermediate and final products. Consistency describes the flow property of non-Newtonian fluids with suspended particles and dissolved long chain molecules (Barrett, Garcia, & Wayne, 1998). The consistency of the paste is typically evaluated using a Bostwick consistometer. This empirical method is widely used in the tomato industry, and USDA has established quality

* Corresponding author. Department of Food Science and Technology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA. Tel.: +1 530 752 8921; fax: +1 530 752 4759.
E-mail address: mjmccarthy@ucdavis.edu (M.J. McCarthy).

standards for tomato products based on the consistency as measured by the Bostwick consistometer (USDA, 1992). The Bostwick consistency of tomato paste decreases exponentially with the increase in concentration, which imposes a limit on the validity of this method on tomato concentrates with more than 15% total solids (Rao, Bourne, & Cooley, 1981; Tanglertpaibul & Rao, 1987). Moreover, the consistency of tomato paste is sensitive to temperature (Tehrani & Ghandi, 2007). As a result, the Bostwick consistency measurement is usually performed on a 12 °Brix dilution of tomato concentrate at 20 °C. The sample preparation and temperature requirement makes it impossible to obtain real-time or even near real-time data using the Bostwick consistency method.

An in-line process measurement of tomato paste properties, especially the consistency, is very valuable for improving the quality of products and minimizing production cost. Infrared spectroscopy has been used as a rapid alternative to assess the quality of tomatoes, such as total solids, soluble solids, titratable acidity, pH, levels of sugars and acids (Ścibisz et al., 2011; Slaughter, Barrett, & Boersig, 1996; Wilkerson et al., 2013). The application has been extended to measurement of properties of tomato concentrate and juice, including total solids, soluble solids, sugar, acid, protein and carotenoids (Goula & Adamopoulos, 2003; Pedro & Ferreira, 2005). In-line consistency prediction of tomato products has been studied using a variety of different approaches, most of which are based on measurement of rheological properties, e.g., differential pressure measurement (Barringer, Azam, Heskitt, & Sastry, 1998), vibrational viscometer (Cullen, Duffy, & O'Donnell, 2001), ultrasonics (Dogan, McCarthy, & Powell, 2002; Dogan, McCarthy, & Powell, 2003) and Flow Magnetic Resonance Imaging (McCarthy & McCarthy, 2009). A correlation between optical density and Bostwick consistency was found in tomato products using in-line absorption photometry (Haley & Smith, 2003). However, no method has been reported to simultaneously measure both the chemical properties and the rheological property, i.e., consistency, of tomato products.

Singh, Bhamidipati, Singh, Smith, and Nelson (1996) evaluated the performance of several in-line sensors for prediction of soluble solids and moisture of fruit juice. Guided Microwave (GM) spectroscopy gave good results for soluble solids and moisture prediction. GM spectrometer applies multiple frequencies in the lower portion of the microwave band to the sample contained in the chamber body and measures the amplitude of microwave signal transmitted across the sample (Liang, Anantheswaran, Bradley, & Long, 2002). The interaction between molecules and microwaves cause molecules to rotate and align with the electromagnetic field. The polarization and depolarization of molecules reduce the wave velocity across the chamber, and the dielectric constant is used to describe this effect. The energy loss due to the friction of orienting molecules causes a reduction in the magnitude of the wave. Dielectric loss factor quantifies the efficiency of the energy loss (Wellock & Walmsley, 2004). The signal amplitude is a function of the change of dielectric constant and dielectric loss factor of the sample (Singh et al., 1996). The ability of microwaves to penetrate thick material allows GM spectroscopy to analyze the bulk properties of a sample instead of a surface measurement.

The objectives of this study were to evaluate guided microwave spectroscopy for in-line measurement of tomato paste properties, including soluble solids content, titratable acidity and Bostwick consistency, and to develop prediction models for tomato paste processing based on the microwave spectra by multivariate calibration. In tomato paste production, tomatoes are thermally processed and concentrated under vacuum by evaporation. Afterwards, the evaporated products are processed into final products by dilution and mixing with additional ingredients. These two steps are key operations in the manufacture of tomato products, and they contribute to the quality of the final products. Therefore, two processing operations, dilution and evaporation, were chosen to assess the performance of the guided microwave spectrometer as an in-line sensor.

2. Material and methods

2.1. Tomato paste material

Nine different tomato paste samples, with soluble solids content ranging from 24 to 30 °Brix and Bostwick consistency ranging from 2 to 11 cm, were obtained from H. J. Heinz Co. (Los Banos, CA). These samples were from commercial batches produced on different days within a three-month period, representing a wide span of variation in the acidity, consistency and other characteristics of the tomato paste. These 9 samples were blended together in various proportions and constituents to create another 25 new pastes. The new samples were formulated so that the consistency values of all 34 samples were distributed in the range between 2 and 11 cm with intervals of 0.2–0.3 cm.

2.2. Guided Microwave (GM) spectrometry

The GM spectroscopy instrument used in this study was an e scan In-line Food Analyzer (Thermo Fisher Scientific, Minneapolis, MN). The spectrometer has a bandwidth of 31.25 MHz to 4 GHz. The sensor body is composed of a transmitter and a receiver separated by two parallel plates, which define a rectangular section with a cross-section area of 45.72 mm (separation of two plates) × 38.86 mm (distance between probes), acts as a waveguide. The volume of the waveguide is 300 mL. The waveguide design functions to suppress microwave power at frequencies whose corresponding wavelength cannot fit into the chamber. A waveguide can only transmit microwaves with a wavelength shorter than twice the distance between the two parallel plates. Cutoff frequency is defined as the frequency, below which the transmission will not occur. Besides the dimensions of the waveguide, the cutoff frequency is affected by the dielectric constant of the measured sample. When a sample flows through the waveguide, microwaves are transmitted through the sample to the receiver, and the response is recorded in the form of a spectrum. Spectra were collected over the frequency range from 136 to 2690 MHz in 2 MHz steps.

2.3. Experimental setup

2.3.1. Dilution
The flow loop system consisted of a feeding tank, positive displacement pump, GM spectrometer, sampling valve, and additional pipes and fittings were used to connect all the pieces together. Samples were pumped from the feeding tank to the GM spectrometer and then returned back to the tank. The dilution study was conducted at room temperature (21 ± 2 °C). Due to the limitation of the pump used in the flow loop, the original tomato paste with soluble solids content over 24 °Brix could not be tested under flow conditions. Therefore, for pastes above 24 °Brix, each tomato concentrate was loaded into the GM spectrometer sample chamber from the top and tested under static conditions. After that, distilled water was added to the paste in the feeding tank and continuous mixing was applied. The paste was diluted at steps of about 5 °Brix until it reached the target level of 5 °Brix. When desired level was achieved, GM spectroscopy measurements were performed on the diluted samples. Six spectra were acquired for each sample during the measurement. At the same time, samples were collected from the sampling valve located at the downstream

end of the GM spectrometer chamber for reference analysis. Five samples were generated from each paste. For example, a 25 °Brix paste was diluted to approximately 20 °Brix, 15 °Brix, 10 °Brix, and 5 °Brix, successively. The original paste and the four dilutions were tested using GM spectrometer. Therefore, a total of five samples were collected from the dilution study of each paste sample.

2.3.2. Evaporation

A semi-batch forced circulation evaporator was used to evaporate the final product from the dilution operation. Sample was fed into the evaporator by a gear pump. The evaporator is mainly composed of a shell and tube heat exchanger, static mixer, back pressure valve, vacuum separator, vacuum pump and a circulation pump. The product from the vacuum separator was circulated back to the heat exchanger unless it was drawn off from the outlet located between the circulation pump and heat exchanger. The GM spectrometer sample chamber was installed upstream of the heat exchanger and downstream of the circulation pump. The evaporator was operated at a reduced pressure of 10 KPa and temperature was maintained at 46 ± 3 °C. About 20 L of juice at around 5 °Brix from the dilution operation was fed to the evaporator and evaporated gradually to 25 °Brix. GM spectroscopy measurements were performed on the diluted juice and on the product of evaporation when approximately 5 °Brix increments in the soluble solids content were achieved. Five samples were generated from the evaporation of each juice. For example, for a 5 °Brix juice, the five samples were at about 5 °Brix, 10 °Brix, 15 °Brix, 20 °Brix and 25 °Brix. Six spectra were acquired for each sample. No product was discharged except for sampling. Samples were collected from the product outlet for reference analysis after the spectra acquisition.

2.4. Chemical and physical analysis

Collected samples were allowed to equilibrate to room temperature before they were tested for their chemical and physical properties, including soluble solids content, titratable acidity and Bostwick consistency.

Soluble solids content (SSC) was determined using a bench-top temperature compensating refractometer (RFM 730, Bellingham and Stanley, UK). The titratable acidity (TA) of each sample was tested according to AOAC method 942.15 (AOAC, 2002). Two grams of sample was diluted in 50 mL distilled water and well mixed. The dilution was titrated with 0.1 M NaOH to an end point of pH 8.2, and the TA value was expressed as the equivalent percentage of citric acid.

The consistency of samples was measured using a Bostwick consistometer (CSC Scientific company, Inc. Fairfax, VA). In industry standard practice, a 12 °Brix dilution of paste is used in evaluation of the Bostwick consistency of the paste. The 5 samples from the dilution series of the same paste would have the same Bostwick consistency. Therefore, the measurement was performed only on the 34 original tomato paste samples in the dilution operation and the 34 final products from the evaporation operation. Each sample was diluted to 12 °Brix by adding distilled water, and equilibrated to 20 ± 1 °C. The distance the 12 °Brix dilution traveled in the Bostwick consistomer after 30 s was read to the nearest 0.1 cm and recorded as the consistency in cm.

All the measurements were performed in duplicate and the average values were reported.

2.5. Multivariate analysis

Spectra pretreatment and multivariate analysis were performed using Matlab 2010a (The Mathworks, Natick, MA) and PLS toolbox 6.5.2 (Eigenvector Research Inc., Wenatchee, WA).

2.5.1. Spectra pretreatment

The heterogeneity of samples, occurrence of air bubbles, and sudden mechanical motion can affect acquired spectra. For this reason, a series of 6 spectra were recorded for each sample and the most similar 4 spectra were averaged. Principle Component Analysis (PCA) is capable of identifying similarities and outliers in multivariate data (Jolliffe, 2002). A PCA approach was employed to detect the abnormal spectra. The 6 spectra with 1278 frequency variables of each sample were treated as 6 stand alone observations and analyzed using PCA. T^2 scores of each spectrum were calculated based on the first two principle components. The hotelling T^2 measures the variation of the variables within the PCA model. The two spectra with the highest hotelling T^2 scores were excluded and the average of the remaining four spectra was used as the spectrum for further analysis.

2.5.2. Partial Least Square Regression

Partial Least Square (PLS) analysis was used to analyze the spectra data of the paste samples. Similar to PCA, PLS handles complex multivariate data like spectra, and it has been applied in the analysis of Infrared, UV, and NMR spectroscopy data as well as developing quantitative predictive models based on references to measure the properties of interest in a sample. In PLS, the regression is based on latent variables, which were constructed to capture maximal covariance between spectral matrix X and the response vector Y (Nicolai et al., 2007). The logarithm to the base 10 of the spectra was used in the analysis. The logarithmic spectrum of each sample was then auto-scaled by subtracting the mean of the spectrum, and the subtracted spectrum was divided by the standard deviation of the spectrum. PLS was applied to establish the relationship between the pretreated spectral data X and the measured quality parameter of interest Y. For SSC and TA calibration models, a 5 fold venetian blinds cross-validation was performed to validate the calibration models. The whole dataset was divided into 5 subsets, and 1 subset was used to validate the calibration model built based on the other 4 subsets. The cross-validation was repeated until all subgroups have been used once as the validation data. The Bostwick consistency models were validated using leave-one-out cross-validation approach. One sample was used as the validation data, and the remaining samples were used to build the calibration model. The validation process was repeated until all samples have been used once as the validation data.

2.5.3. Interval Partial Least Square

Interval Partial Least Square (iPLS) is a variable selection technique searching for important spectral regions and removing interferences from other regions (Norgaard et al., 2000). The iPLS algorithm develops local PLS models based on spectral subintervals of equal width. The prediction performance of the local PLS models and the full-spectrum PLS model is compared in terms of the root mean square error of cross-validation (RMSECV) defined as:

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}$$

where n is the number of the cross-validated samples, y_i is the reference value from lab measurement of the cross-validated sample i, and $\widehat{y}_i$ is the predicted value of the cross-validated sample i. The combination of intervals that gives the lowest RMSECV is selected. Therefore, a simpler model with better performance can be obtained by using only the variables selected by iPLS (Pereira et al., 2008). The selection of interval was performed using the iPLS algorithm in the PLS toolbox 6.5.2 (Eigenvector Research Inc., Wenatchee, WA). The iPLS was run in forward mode on the

logarithm auto-scaled spectra with an interval size of 63 variables. As a result, the whole spectrum with 1278 variables was divided into 20 subintervals of equal width. Cross-validation of the local PLS models was conducted as described in 2.5.2 to validate and evaluate the model performance.

3. Results and discussion

3.1. Spectral characteristics

Fig. 1 shows an example of GM spectra of a paste sample diluted to different soluble solids content. All spectra began with an initial sharp rise in the signal amplitude as a function of frequency. The signal amplitude leveled off around 650 MHz and decayed gradually after that. The region where the initial increase occurred is referred to as the cutoff region in the GM spectra, and the passband region starts from the point where the signal levels off. The shape and the signal amplitude of the spectrum are determined by three dielectric properties of the sample, dielectric constant, electrical conductivity and the molecular relaxation time (Jean, 2006). The cutoff frequency approximates the frequency at which the spectrum started to level (Liang et al., 2002). It is dependent on the dielectric constant of the sample, which is a strong function of the moisture content. The amplitude of the passband region is contingent upon the electrical conductivity of the sample, and can be estimated from the slope of the cutoff. The molecular relaxation time determines the slope of the signal decay after the point signal starts to level off (Singh et al., 1996). The changes in the concentration of the chemical composition result in changes in the dielectric properties of the sample, leading to changes in the spectra.

For example, dilution of tomato paste caused an increase in the moisture content; meanwhile, a decrease occurred in the soluble solids content and the titratable acidity. As illustrated in Fig. 1, the spectrum gradually shifted to the left with decreasing soluble solids content. At the same time, signal amplitude of the passband region decreased. These changes were the joint effects of the changes of the moisture content, soluble solids content and titratable acidity.

Fig. 2 shows the spectra of three tomato pastes with different Bostwick consistency in the dilution study. These three spectra demonstrated differences in the amplitude and the slope of the passband region. Besides the consistency, the samples varied in constituent components, such as soluble solids content and titratable acidity, which may contribute to the divergence in the spectral features. Because multiple properties of the sample may be related to one dielectric property of the sample, consequently leading to
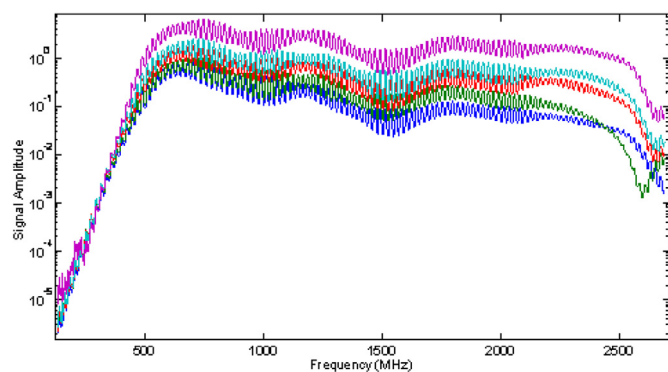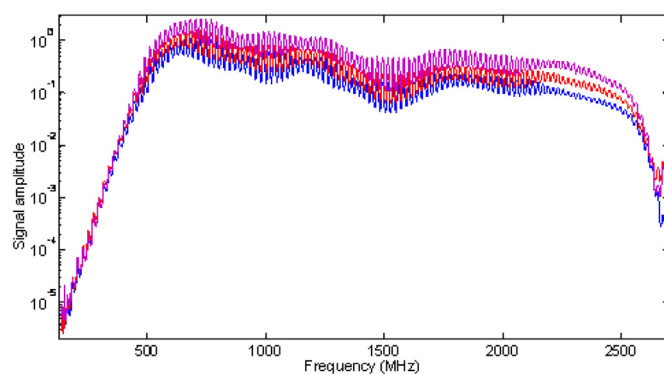

Fig. 2. GM spectra of 3 tomato pastes with different Bostwick consistency. Signal amplitude is the ratio of the input signal to the transmitted signal. From the bottom to the top, Bostwick: 3.3 cm, 7.3 cm, 9.7 cm; soluble solids content: 25.5 °Brix, 30.2 °Brix, 30.2 °Brix; Titratable acidity: 1.30%, 1.65% and 1.34%.

change in the spectra, it is challenging to correlate a specific characteristic or component of the samples with the spectra. Unlike other spectroscopic techniques, no clearly defined peaks are associated with chemical composition of a sample in a GM spectrum. Multivariate calibration is usually needed to correlate the spectra to the chemical composition or properties of a sample (Wellock & Walmsley, 2004).

3.2. PLS regression

3.2.1. Dilution

PLS was performed on the spectra with 1278 frequency variables to explore the correlation between the properties of interest of the tomato paste samples and their GM spectra. Sample outliers were excluded from the model based on their student t residual and leverage. For the dilution study, tomato paste samples used for PLS regression had SSC values ranging from 5.2 to 30.2 °Brix, TA values ranging from 0.23 to 1.86%, and consistency ranging from 1.8 to 9.9 cm. PLS calibration models were developed using the optimum number of latent variables based on cross-validation. The criterion is to select the minimum number of latent variables that minimize the RMSECV.

A 4-factor PLS model was built for SSC, capturing 95% of the variance in SSC. The PLS model predicted SSC values were plotted against the SSC from lab measurements (Fig. 3a). The model demonstrated better performance in the range of 5–18 °Brix than in higher SSC region. Broad scatter of data points were observed along the diagonal line above 25 °Brix. A PLS model for TA was developed on the first 4 latent variables, accounting for 93% of the variance in TA. Similarly, the TA prediction model had higher accuracy in the low acidity range, and the difference between the predicted and measured TA became evident when the TA was greater than 1% (Fig. 3b). In the dilution study, the tomato paste samples with higher SSC (>25 °Brix) were analyzed under static conditions due to their high viscosity, while measurement of the diluted pastes were conducted in the process of continuous flow and mixing. The agitation and continuous flow facilitated better mixing of the sample, which made the material within the spectrometer chamber more uniform, giving more representative results. For Bostwick consistency, a 5-factor model was chosen from cross-validation. Although 93% of the variance in the consistency can be explained by this model, the accuracy of the consistency model was lower than the SSC and TA models (Fig. 3c).

The performance parameters of the cross-validated prediction models were summarized in Table 1. RMSEC (root mean square

Fig. 1. GM spectra of a paste sample diluted to different soluble solids content. Signal amplitude is the ratio of the input signal to the transmitted signal. From the bottom to the top, soluble solids content: 27.3 °Brix, 23.9 °Brix, 15.5 °Brix, 12.6 °Brix and 6.6 °Brix; Titratable acidity: 1.65%, 1.45%, 0.98%, 0.81% and 0.44%.
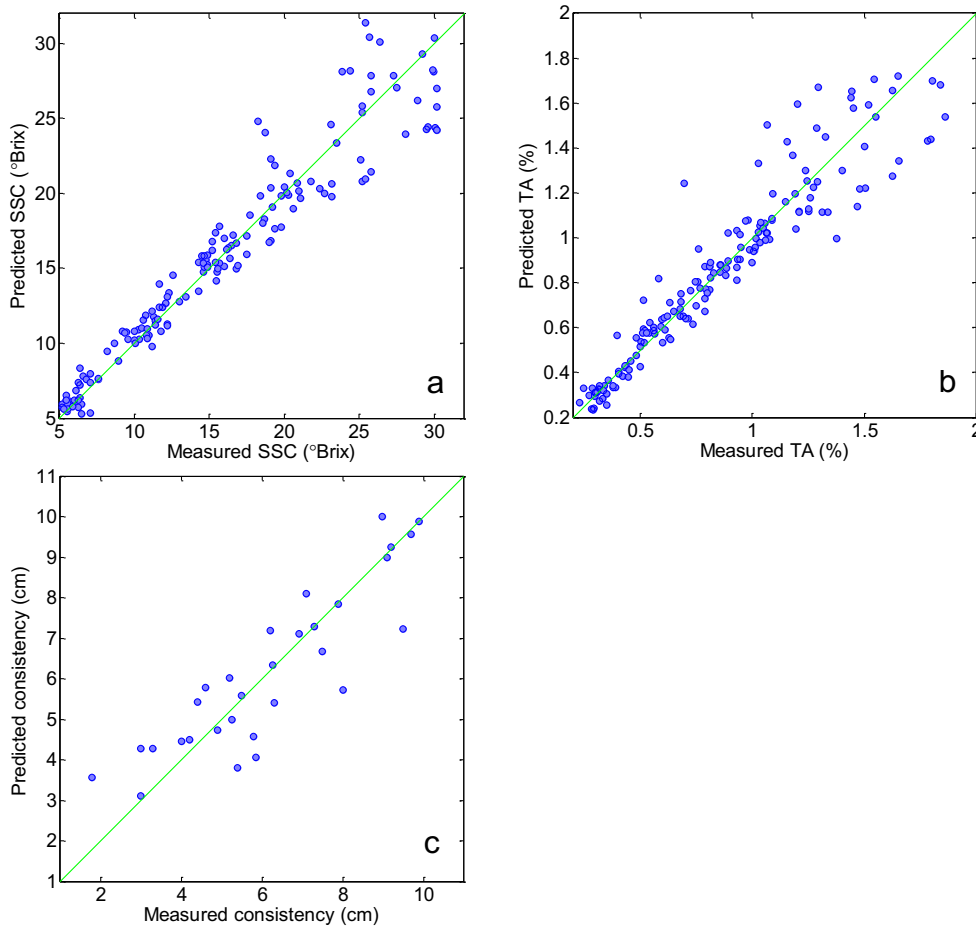
Fig. 3. PLS model predicted value versus reference value of (a) soluble solids content, (b) titratable acidity and (c) Bostwick consistency in dilution study.

error of calibration) measures the difference between the actual measured value and predicted value for the calibration samples and has the same unit as the measured parameter. RMSECV is calculated from the cross-validated samples. Relative RMSECV is calculated as the ratio of RMSECV to the measured mean value, describing the prediction error relative to the measured parameter. Both the SSC and TA models had a correlation coefficient of cross-validation (r_{cv}) higher than 0.94. The RMSECV values of these two models were

Table 1
Performance of PLS models for quality parameters.

Condition	Parameter	N[a]	RMSEC[b]	RMSECV[c]	Relative RMSECV[d]	r_{cv}[e]
Dilution	SSC (°Brix)	167	1.8	2.1	0.135	0.960
	TA (%)	167	0.11	0.13	0.153	0.947
	Bostwick consistency (cm)	30	0.6	1.0	0.174	0.880
Evaporation	SSC (°Brix)	170	1.1	1.3	0.084	0.976
	TA (%)	170	0.08	0.10	0.133	0.958
	Bostwick consistency (cm)	30	0.2	1.1	0.193	0.839

[a] N: number of samples in the model.
[b] RMSEC: root mean square error of calibration.
[c] RMSECV: root mean square error of cross validation.
[d] Relative RMSECV: relative root mean square error of cross validation (RMSECV/mean).
[e] r_{cv}: correlation coefficient of cross validation. RMSEC and RMSECV are in units of the predicted parameters.

similar to their RMSEC, indicating the performance of the models did not decay on the cross-validation test data. The lower r_{cv} and higher relative RMSECV of the consistency model suggested that the performance of the model was inferior to that of SSC and TA.

3.2.2. Evaporation

In the evaporation study, collected samples were in the range of 5.5–29.7 °Brix for SSC, 0.24–1.628% for TA and 2.25–9.4 cm for Bostwick consistency. A 4-factor PLS model captured 97% of the variance in SSC. For TA, the prediction model base on 4 latent variables accounted for 94% of the variance. As shown in Fig. 4, these models outperformed their counterparts in the dilution study. The predicted values were better correlated to the measured reference values, especially in the high SSC and TA range, than in the dilution study models. The higher accuracy of the SSC and TA model in the evaporation study was confirmed by their lower RMSECV and relative RMSECV (Table 1). The circulation pump and static mixer in the evaporator allowed more uniform flow and better mixing of tomato concentrate than the stirrer in the feeding tank used in the dilution study, and subsequently introduced less variation in the sample within the GM spectrometer chamber during spectrum acquisition and sampling for lab measurement.

The Bostwick consistency model was built on 5 latent variables, which explained 99% of the variance in the consistency. The performance of the model is illustrated in Fig. 4c, the data points scattered loosely along the diagonal line, where the predicted value equals to the measured value, but the linear relationship between
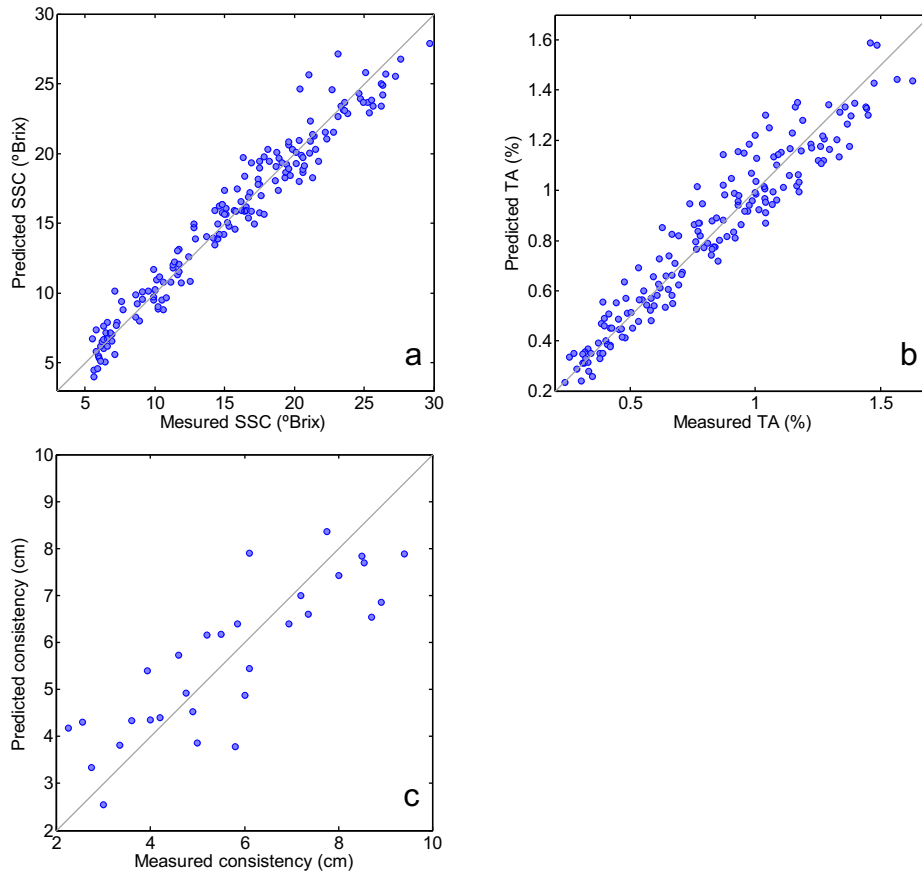
Fig. 4. PLS model predicted value versus reference value of (a) soluble solids content, (b) titratable acidity and (c) Bostwick consistency in evaporation study.

the predicted consistency and lab measured consistency was apparent. The RMSECV of 1.1 cm in the consistency model is relatively high considering the range of consistency in the model is around 7 cm. In addition, the relative RMSECV of the consistency model is 0.19, indicating the prediction error (RMSECV) is almost 20% of the mean consistency. However, the r_{cv} (0.839) indicated that the GM spectrum and the Bostwick consistency were highly

correlated. The improvement in the mixing in the evaporator did not introduce any enhancement in the performance of the consistency model.

Consistency, as an important rheological property of tomato product, is usually correlated with results from a rheology based method, such as viscosity. However, the consistency of tomato paste is affected by its composition. The water insoluble solids are
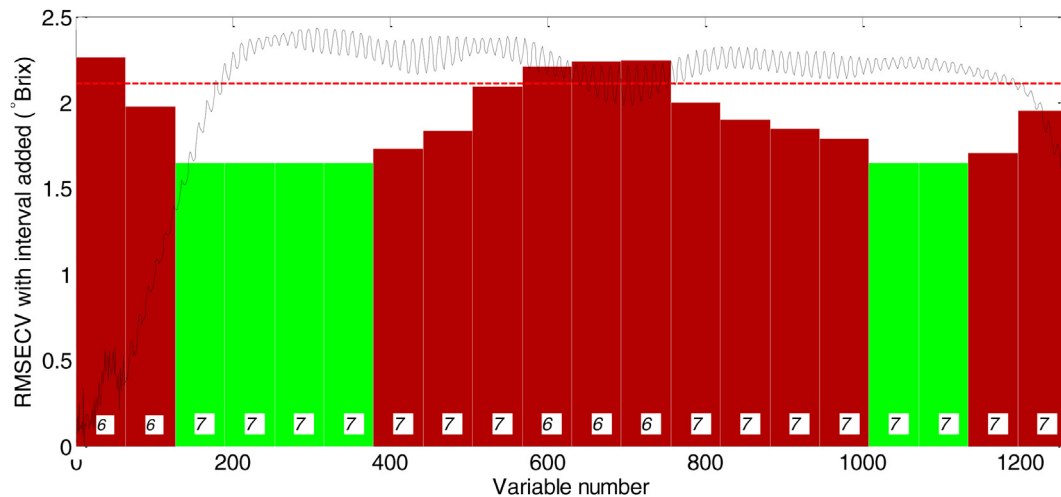

Fig. 5. iPLS result for soluble solids content in dilution study. The columns indicate the RMSECV for the model with the intervals added, which is compared with the RMSECV of the full-spectrum model (dashed line). The number at the bottom of each column denotes the optimal number of latent variables in the model base the interval. A logarithm scaled mean spectrum is overlaid on the plot.

Table 2
iPLS results for quality parameters.

Condition	Parameter	Interval selected (variable number)	RMSECV[a]	r_{cv}[b]
Dilution	SSC (°Brix)	127–378, 1009–1134	1.6	0.976
	TA (%)	127–378, 883–1071	0.11	0.969
	Bostwick consistency (cm)	253–378, 505–567	0.7	0.945
Evaporation	SSC (°Brix)	253–504, 694–756, 1072–1197	1.2	0.980
	TA (%)	505–630	0.09	0.965
	Bostwick consistency (cm)	127–189, 505–630, 1135–1260	0.8	0.929

[a] RMSECV: root mean square error of cross validation.
[b] r_{cv}: correlation coefficient of cross validation. RMSECV values are in units of the predicted parameters.

considered to be the greatest contributor to the consistency of tomato products. The insoluble solids in tomato products are composed of intact and broken cells, cell fragments, and long chain polymers of lignin, cellulose, hemicellulose, and water-insoluble pectic materials (Barrett et al., 1998). Both the insoluble solids content and the size distribution and shape of the insoluble solids have effect on the consistency of tomato paste (Hayes, Smith, & Morris, 1998). These insoluble solids are suspended and highly

hydrated (Barrett et al., 1998), thus interacting with water molecules and influencing the mobility of water. As a result, the dielectric behavior of the water varies as a function of insoluble solids and the amplitude and shape of the GM spectrum change accordingly. On the other hand, relaxation time is related to the viscosity of a liquid and the relationship between them can be described as a linear function for idealized sphere molecules (Tang, 2005). The relaxation time of the sample affects the amplitude and slope of the passband region of the GM spectrum. Therefore, GM spectrum contains information for measurement of the rheological properties of tomato pastes.

3.3. iPLS variable selection

Although good prediction models for SSC and TA were obtained from full-spectrum PLS regression, the performance of the Bostwick consistency model was not satisfactory. iPLS was used to improve the model performance by removing noisy and irrelevant intervals in the spectrum. Moreover, iPLS selects the most interesting and relevant part of the spectrum, helping in spectroscopic interpretation (Pereira et al., 2008). It is advised to experiment using a different number of intervals when doing iPLS. If the number of intervals is too low, the interval may be too broad so that small signature features of the spectrum cannot be seen by the
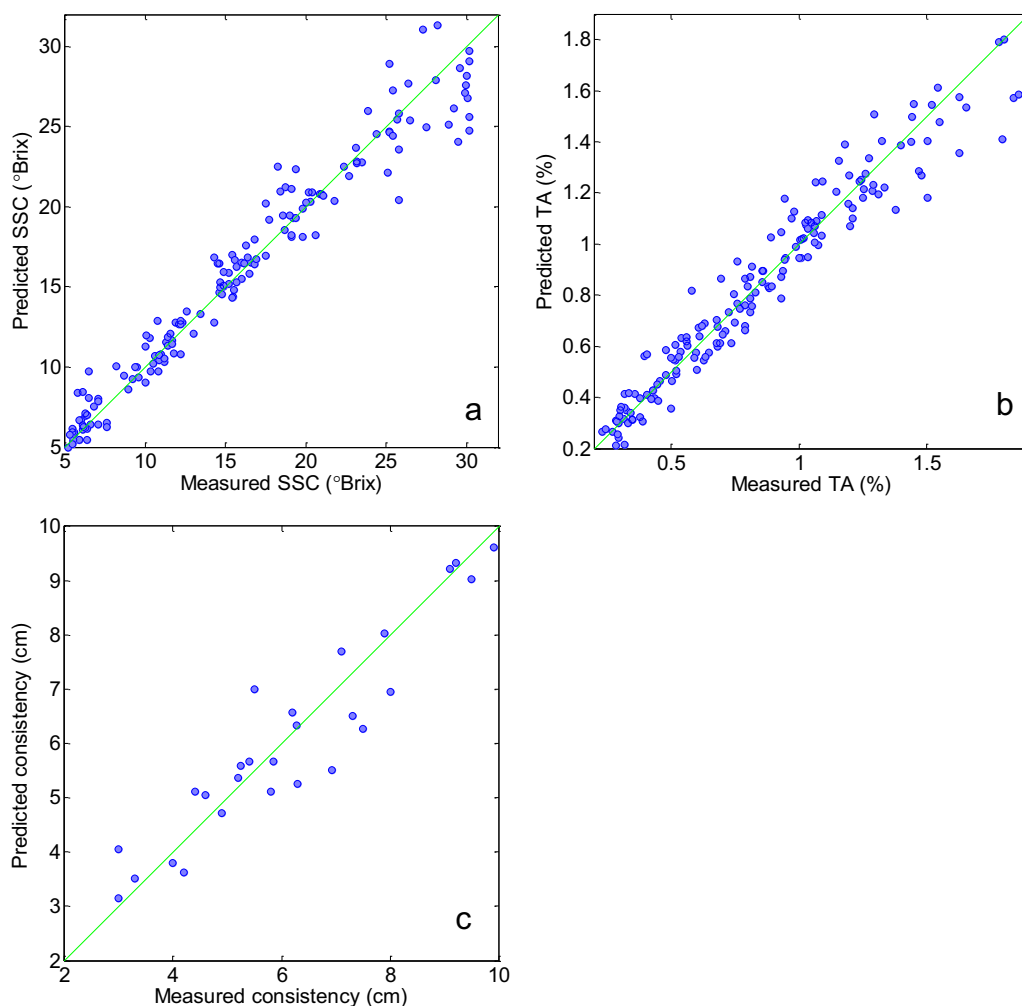
Fig. 6. iPLS optimized model predicted value versus reference value of (a) soluble solids content, (b) titratable acidity and (c) Bostwick consistency in dilution study.

variable selection algorithm. If the number of the intervals is too large, iPLS may not be able to catch all the correlations between the neighboring variables in the spectrum (Andersen & Bro, 2010). An interval size of 63 variables was chosen based on trials with different interval sizes on all data acquired.

3.3.1. Dilution

Fig. 5 illustrates the iPLS variable selection result for SSC in the dilution study. All columns had the same window size of 63 variables, and the height of each column represented the RMSECV value obtained from a local model using the corresponding interval. The combination of intervals that gave the lowest RMSECV was selected to develop a model with reduced number of variables. For SSC, variables 127–378 and 1009–1134 were selected, shown as green columns (in the web version) in Fig. 5. The selected variables corresponded to the frequency range of 388–890 MHz and 2152–2402 MHz. Removing noisy or irrelevant intervals in the spectrum helped to improve the performance of the SSC prediction model as proven by the new model's lower RMSECV and higher r values (Table 2), compared to those of the full PLS model (Table 1). The optimized model (Fig. 6a) alleviated the problem of the relatively high error of prediction in the high SSC range seen in the full model (Fig. 3a). For TA, a slightly different set of intervals were selected, frequency in the range of 388–890 MHz and 1900–2276 MHz. Similar to the new model of SSC, the TA model with optimized number of intervals showed better correlation between the lab measured TA and predicted TA along with a lower error, especially in the high TA range (Fig. 6b). The most notable improvement in the

performance was observed in the optimized model of Bostwick consistency, which had considerably lower RMSECV and higher r (Table 2) than the full model (Table 1). The optimized model for consistency was based on the frequency range, 640–890 MHz and 1144–1268 MHz.

The selected intervals for SSC, TA and Bostwick consistency shared some common regions (Table 2). The variables 127–378, corresponding to frequency 388–890 MHz, were both included in the optimized model of SSC and TA. The frequency range included part of the cutoff slope and the beginning of the passband region (Fig. 5). Since this region is usually used to calculate the electrical conductivity and contains the information on dielectric constant, the changes in dielectric constant and conductivity of diluted tomato paste may associate with variation in SSC and TA. The other selected interval for SSC (2152–2402 MHz) and TA (1900–2276 MHz) are located towards the end of the passband region. The intervals selected for Bostwick consistency prediction contain a region close to the initial amplitude peak (640–890 MHz) and another frequency range in the far passband region (1144–1268 MHz). As mentioned in 3.2.2, relaxation time is a function of viscosity, which is a correlate of the Bostwick consistency of tomato paste. The selected regions for consistency can be used to calculate the slope of the signal rolloff after the initial amplitude peak in the passband region, which is determined by the relaxation time. The iPLS result for consistency demonstrated the theoretical relationship between relaxation time and viscosity. Furthermore, close to the peak region 640–890 MHz may contain information related both to the dielectric constant (coupled with the cutoff frequency)
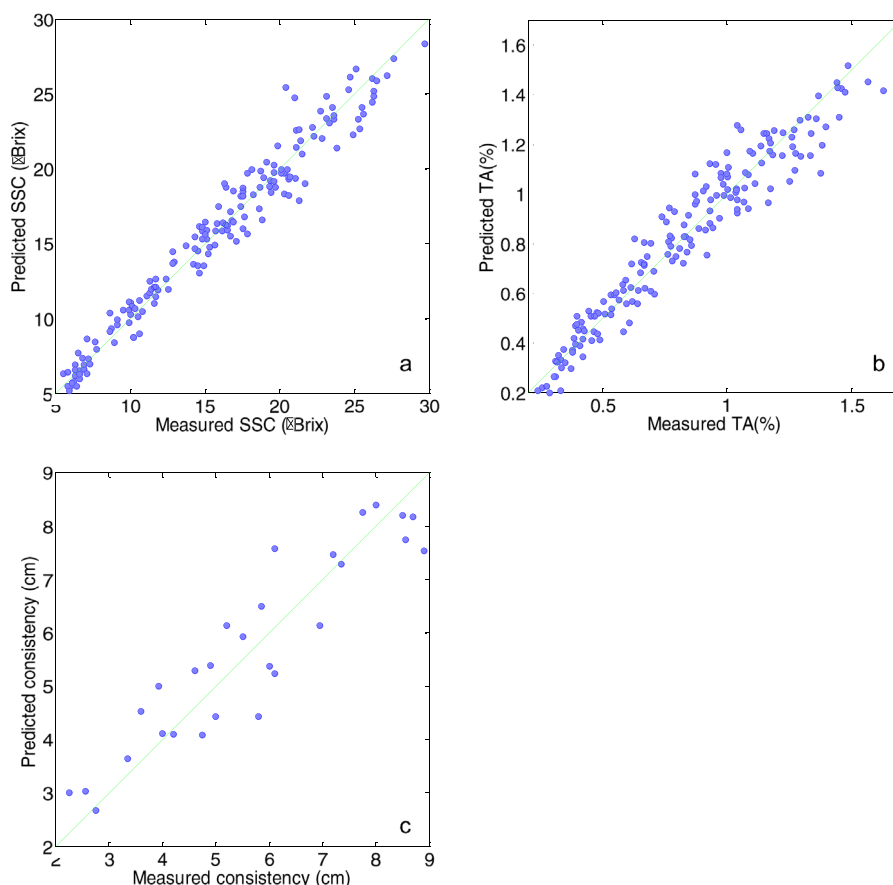

Fig. 7. iPLS optimized model predicted value versus reference value of (a) soluble solids content, (b) titratable acidity and (c) Bostwick consistency in evaporation study.

and the conductivity (correlated with the slope of cutoff region). Thus, Bostwick consistency is affected by other factors, including water content and chemical constituents, which may change the conductivity of tomato paste.

3.3.2. Evaporation

For all three parameters, models based on selected intervals presented better prediction ability than their full models, especially the consistency model (Fig. 7). Different sets of intervals were selected by iPLS for each parameter in the evaporation study, comparing with the results from the dilution study (Table 2). Dielectric properties are temperature and density dependent (Ryynänen, 1995). The evaporation study was conducted at an elevated temperature under reduced pressure, both of which altered the dielectric properties of the tomato paste sample. The amplitude and shape of the GM spectra changed as a result, leading to shifts in the frequency ranges related to the parameters of interest.

In both the evaporation study and dilution study, different regions of the spectrum were selected by iPLS for optimization of models for the three parameters. Correlations may exist between SSC, TA and consistency of the tomato paste samples. It is known that there is a linear relationship between the soluble solids content of tomato paste and the logarithm of Bostwick consistency (Marsh, Buhlert, & Leonard, 1980). The SSC of these concentrated tomato pastes used in the dilution study varied within the range of 20–30 °Brix. The iPLS variable selection results indicated that the ability of GM spectroscopy to quantify the Bostwick consistency is not based on the correlation between the spectra and SSC even though considerable variation in the SSC existed in the samples.

4. Conclusion

The chemical constituents of food products affect their dielectric properties, which enables GM spectroscopy technology to measure the properties of tomato pastes. This study confirmed the feasibility of the GM spectroscopy for determination of SSC, TA and Bostwick consistency of tomato pastes. Experiments were conducted under two processing operations, dilution and evaporation, with different temperature and pressure. A multivariate calibration method, PLS, was used to correlate the GM spectra with the properties of interest and develop prediction models for the tomato paste quality parameters. The PLS model for quantification of SSC and TA provided fairly accurate prediction and gave r of at least 0.95 between the GM spectroscopy predicted and the lab measured values in both operation conditions. By comparison, the Bostwick consistency model had lower prediction accuracy. However, an r value of above 0.8 validated the correlation between the predicted consistency based on GM spectra and the measured value. The performance of the prediction models for all three parameters was improved by adopting a local regression modeling approach, iPLS. The iPLS method selected the most relevant and least noisy intervals in the spectrum for an optimized regression model. The Bostwick consistency models benefited the most from the variable selection as shown by a significant increase in the r to above 0.92 and decrease in the prediction error. Different spectral intervals were selected for the three parameters, implying the chemical information related to each of the properties is embedded in different region of the spectra of tomato paste. For all quality parameters, the variable selection results were distinct between the dilution study and the evaporation study. The dielectric properties of tomato pastes changed with temperature and pressure, leading to changes in the GM spectra. These changes made it impractical to use the same set of GM spectra base prediction models for processing conditions with different temperature and pressure. In summary, the GM spectroscopy technique proved useful for in-line measurement of tomato paste properties, SSC, TA and Bostwick consistency under different operation conditions. It provided a rapid and effective alternative to the standard methods used in tomato processing industry, especially for consistency determination.

References

Andersen, C. M., & Bro, R. (2010). Variable selection in regression—a tutorial. *Journal of Chemometrics, 24*(11–12), 728–737.

Anthon, G. E., Diaz, J. V., & Barrett, D. M. (2008). Changes in pectins and product consistency during the concentration of tomato juice to paste. *Journal of Agricultural and Food Chemistry, 56*(16), 7100–7105.

AOAC. (2002). *Official methods of analysis of AOAC international* (17th ed. ed). Gaithersburg, MD: AOAC International.

Barrett, D. M., Garcia, E., & Wayne, J. E. (1998). Textural modification of processing tomatoes. *Critical Reviews in Food Science and Nutrition, 38*(3), 173–258.

Barringer, S. A., Azam, A. T. M. S., Heskitt, B., & Sastry, S. (1998). On-line prediction of Bostwick consistency from pressure differential in pipe flow for ketchup and related tomato products. *Journal of Food Processing and Preservation, 22*(3), 211–220.

Cullen, P. J., Duffy, A. P., & O'Donnell, C. P. (2001). In-line consistency monitoring of tomato based products using vibrational process viscometry. *Journal of Food Processing and Preservation, 25*(5), 337–351.

Dogan, N., McCarthy, M. J., & Powell, R. L. (2002). In-line measurement of rheological parameters and modeling of apparent wall slip in diced tomato suspensions using ultrasonics. *Journal of Food Science, 67*(6), 2235–2240.

Dogan, N., McCarthy, M. J., & Powell, R. L. (2003). Comparison of in-line consistency measurement of tomato concentrates using ultrasonics and capillary methods. *Journal of Food Process Engineering, 25*(6), 571–587.

Goula, A., & Adamopoulos, K. G. (2003). Estimating the composition of tomato juice products by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 11*(2), 123–136.

Haley, T. A., & Smith, R. S. (2003). Evaluation of in-line absorption photometry to predict consistency of concentrated tomato products. *LWT - Food Science and Technology, 36*(2), 159–164.

Hayes, W. A., Smith, P. G., & Morris, A. E. J. (1998). The production and quality of tomato concentrates. *Critical Reviews in Food Science and Nutrition, 38*(7), 537–564.

Jean, B. R. (2006). Process composition monitoring at microwave frequencies: a waveguide cutoff method and calibration procedure. *Instrumentation and Measurement, IEEE Transactions on, 55*(1), 180–186.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer.

Liang, L., Anantheswaran, R., Bradley, M., & Long, B. (2002). Characterization of guided microwave spectrometry using water-ethanol mixtures. *The Journal of Microwave Power and Electromagnetic Energy: A Publication of the International Microwave Power Institute, 37*(1), 3.

Marsh, G. L., Buhlert, J. E., & Leonard, S. J. (1980). Effect of composition upon Bostwick consistency of tomato concentrate. *Journal of Food Science, 45*(3), 703–706.

McCarthy, K. L., & McCarthy, M. J. (2009). Relationship between in-line viscosity and bostwick measurement during ketchup production. *Journal of Food Science, 74*(6), E291–E297.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology, 46*(2), 99–118.

Norgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. (2000). Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy, 54*(3), 413–419.

Pedro, A. M. K., & Ferreira, M. M. C. (2005). Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry, 77*(8), 2505–2511.

Pereira, A. F. C., Pontes, M. J. C., Neto, F. F. G., Santos, S. R. B., Galvão, R. K. H., & Araújo, M. C. U. (2008). NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection. *Food Research International, 41*(4), 341–348.

Rao, M. A., Bourne, M. C., & Cooley, H. J. (1981). Flow properties of tomato concentrates. *Journal of Texture Studies, 12*(4), 521–538.

Ryynänen, S. (1995). The electromagnetic properties of food materials: a review of the basic principles. *Journal of Food Engineering, 26*(4), 409–429.

Ścibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, M., Bertrand, D., et al. (2011). Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. *Food Chemistry, 125*(4), 1390–1397.

Singh, P. C., Bhamidipati, S., Singh, R. K., Smith, R. S., & Nelson, P. E. (1996). Evaluation of in-line sensors for prediction of soluble and total solids/moisture in continuous processing of fruit juices. *Food Control, 7*(3), 141–148.

Slaughter, D. C., Barrett, D., & Boersig, M. (1996). Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Journal of Food Science, 61*(4), 695–697.

Tang, J. (2005). Dielectric properties of foods. In H. Schubert, & M. Regier (Eds.), *The microwave processing of foods* (p. 29). Woodhead Pub Limited.

Tanglertpaibul, T., & Rao, M. A. (1987). Flow properties of tomato concentrates: effect of serum viscosity and pulp content. *Journal of Food Science, 52*(2), 318–321.

Tehrani, M. M., & Ghandi, A. (2007). Modification of Bostwick method to determine tomato concentrate consistency. *Journal of Food Engineering, 79*(4), 1483–1486.

United States Department of Agriculture (USDA), Agricultural Marketing Service. (1992). *United States standards for grades of tomato catsup*, 57 FR 2980.

Wellock, R., & Walmsley, A. D. (2004). Applications of microwave spectroscopy in process analysis. *Spectroscopy Europe, 16*(4), 23–26.

Wilkerson, E. D., Anthon, G. E., Barrett, D. M., Sayajon, G. F. G., Santos, A. M., & Rodriguez-Saona, L. E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *Journal of Agricultural and Food Chemistry, 61*(9), 2088–2095.