

Article



Nondestructive Estimation of Moisture Content, pH and Soluble Solid Contents in Intact Tomatoes Using Hyperspectral Imaging

Anisur Rahman^{1,2}, Lalit Mohan Kandpal¹, Santosh Lohumi¹, Moon S. Kim³, Hoonsoo Lee³, Changyeun Mo⁴ and Byoung-Kwan Cho^{1,*}

- ¹ Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea; anis_fpm@bau.edu.bd (A.R.); lalitm85@gmail.com (L.M.K.); santosh.sanny123@gmail.com (S.L.)
- ² Department of Farm Power and Machinery, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh
- ³ Environmental Microbiology and Food Safety Laboratory, BARC-East, Agricultural Research Service, US Department of Agriculture, 10300 Baltimore Avenue, Beltsville, MD 20705, USA; moon.kim@ars.usda.gov (M.S.K.); hoonsoo.lee@ars.usda.gov (H.L.)
- ⁴ National Academy of Agricultural Science, Rural Development Administration, 310 Nongsaengmyeong-ro, Wansan-gu, Jeonju, Jeollabuk-do 54875, Korea; cymoh100@korea.kr
- * Correspondence: chobk@cnu.ac.kr; Tel.: +82-42-821-6715; Fax: +82-42-823-6246

Academic Editor: Hung Nguyen

Received: 3 December 2016; Accepted: 17 January 2017; Published: 21 January 2017

Abstract: The objective of this study was to develop a nondestructive method to evaluate chemical components such as moisture content (MC), pH, and soluble solid content (SSC) in intact tomatoes by using hyperspectral imaging in the range of 1000–1550 nm. The mean spectra of the 95 matured tomato samples were extracted from the hyperspectral images, and multivariate calibration models were built by using partial least squares (PLS) regression with different preprocessing spectra. The results showed that the regression model developed by PLS regression based on Savitzky–Golay (S–G) first-derivative preprocessed spectra resulted in better performance for MC, pH, and the smoothing preprocessed spectra-based model resulted in better performance for SSC in intact tomatoes compared to models developed by other preprocessing methods, with correlation coefficients (r_{pred}) of 0.81, 0.69, and 0.74 with root mean square error of prediction (RMSEP) of 0.63%, 0.06, and 0.33% Brix respectively. The full wavelengths were used to create chemical images by applying regression coefficients resulting from the best PLS regression model. These results obtained from this study clearly revealed that hyperspectral imaging, together with suitable analysis model, is a promising technology for the nondestructive prediction of chemical components in intact tomatoes.

Keywords: hyperspectral imaging; intact tomato; diffuser lighting; moisture content; pH; soluble solid content; partial least squares regression

1. Introduction

Tomatoes are the most important fruit grown worldwide with approximately 170.8 million tons produced in 2014, which are widely consumed in either fresh or processed form [1]. They are considered healthy because they contain high levels of lycopene, which is a natural antioxidant, as well as β -carotene, vitamin C, and vitamin E [2]. Selecting quality tomatoes based on external appearances such as size, shape, color, lack of defects and decay, while monitoring internal qualities such as sweetness, acidity, moisture content (MC), and nutritional content, are important for the quality and

safety of finished products [3,4]. Most instrumental techniques used to measure these properties are destructive, involve a considerable amount of manual work, and provide information in just a limited section of fruit. Therefore, it is necessary to develop a nondestructive measurement technique to obtain information on these quality attributes from whole samples, which will be very useful for producers, processors and distributors to ascertain the quality of tomatoes.

Hyperspectral imaging (HSI), a new advanced technique, provides both spectra and spatial information for samples by integrating the principles of spectroscopic and imaging technologies in a system [5–8]. The hyperspectral image is a three-dimensional (3-D) cube, normally called a hypercube, in which the spectral information is provided for each pixel in the image. HSI can also provide more detailed or complete information, including internal structure characteristics, morphological information, and chemical composition of the sample, compared with a single-machine vision technology or spectroscopy analysis technology [9]. Now defined, the HSI technique has been applied to the nondestructive measurement of agricultural products such as determination of MC [4,10–14], pH estimation [10,15–18], soluble solid content (SSC) prediction, [10,11,16,18–21] and so many other applications.

A previous study by Schmilovitch et al. [21] developed a nondestructive method to evaluate chemical contents such as total soluble solids, total chlorophyll, carotenoid and ascorbic acid content in intact bell peppers by using hyperspectral imaging in the visible and near-infrared (550–850 nm) region. In addition, Mollazade et al. [4] showed an image processing-based algorithm to evaluate the potential of HSI combined with artificial neural networks to spatial mapping of MC in tomato fruits in the spectral range of 400–1000 nm. Another study by ElMasry et al. [10] demonstrated that HSI in the visible and near-infrared (400–1000 nm) regions was tested for nondestructive determination of MC, SSC, and acidity (expressed as pH) in intact strawberries using the partial least squares (PLS) analysis, a multivariate calibration technique. However, to the best of our knowledge, up to now, there is no report on determining chemical components (MC, pH and SSC) of intact tomatoes using the hyperspectral imaging technique.

Therefore, the overall objective of this study was to develop the HSI technique using PLS regression analysis for rapid and nondestructive measurement of MC, pH, and SSC in intact tomatoes. The specific objectives of this study were to acquire the hyperspectral images of intact tomatoes; secondly, to recognize the important region of interests (ROIs) and extract the corresponding spectral data within the acquired hyperspectral images; and, finally, to establish the quantitative analysis model using PLS regression and create chemical images by applying regression coefficients resulting from the best PLS regression model.

2. Materials and Methods

2.1. Tomato Samples

Eight kilograms of tomatoes were purchased from a local supermarket in South Korea. From the original 8 kg of tomatoes purchased, tomatoes too big in size, nonuniform in color were removed, leaving a more consistent collection of ninety-five tomatoes. Tomatoes weighed 102.2 ± 12.2 g on average, and measured an average of 59.5 ± 2.7 mm in length on the major axis and 49.9 ± 3.9 mm on the minor axis. After numbering, the samples were stored at 20 °C and 85% RH. The day after tomatoes were stored was defined as the beginning of the experiment. Prior to the experiment, the samples had been equilibrated at laboratory temperature (20 ± 1 °C) for at least 2 h to avoid any effects of storage temperature on the measurements. The key steps of the experimental and data analysis procedures are outlined in Figure 1 and are explained in detail in the following sections.



Figure 1. Key steps in the full procedure for predicting the moisture content (MC), pH and soluble solid content (SSC) in intact tomatoes.

2.2. Hyperspectral Imaging System

A laboratory-based push-broom HSI system (shown in Figure 2) was used to acquire the hyperspectral images of tomatoes. The system comprised a line scan image spectrograph (Headwall Photonics, Fitchburg, MA, USA) that covered the spectral range 948–2494 nm with 7.5 nm spectral resolution, mercury cadmium telluride (MCT) detectors to detect the radiation reflected back from the samples, a high-performance camera with a detector sized 320×256 pixels (Headwall Photonics, Fitchburg, MA, USA), 1.4/25 C-mount lens (Navitar, Inc., Rochester, NY, USA), a stepper motor to move the conveyer, data acquisition software, and a display unit.



Figure 2. Schematic of the hyperspectral imaging system.

The tomatoes were illuminated with diffused light provided by four tungsten-halogen lamps (Light Bank, Ushio Inc., Tokyo, Japan) with fiber optics to illuminate the samples (each 100 W, 12 V) located at the circumference of a diffusing dome, at equal distances from one another. To improve

the signal-to-noise ratio, the samples were set to move at 0.2 mm/scan through the conveyor unit, and to be able to cover the spatial shape of the samples, the exposure time was set at 7500 μ s, and the distance of the samples from the camera was set at 40 cm. Tomato samples were placed on a cup and transferred to the conveyer belt to be scanned line by line. The movement of a sample fixed to the translation stage was controlled by the step interval and the number of steps. Both spectral and spatial information were obtained from the samples when it was in the range of the camera during movement through the conveyor unit. The hyperspectral images of tomato samples were saved in a raw format as a three-dimensional (3-D) hypercube consisting of two spatial dimensions and one spectral dimension. The dimensions of the hypercube were 320 pixels in the *x*-direction, *n* pixels in the *y*-direction (based on the length of the sample), and 208 bands in the *k*-direction. The system components and image acquisition process were controlled by software developed using Microsoft (MS) Visual Basic (Version 6.0) operated in the MS Windows operating system.

2.3. Image Correction

The hyperspectral images of tomatoes were transformed into hyperspectral reflectance images using Equation (1) in order to remove the noise generated by the device and the effects of uneven light source intensities. To correct the raw images, white and dark reference images were acquired. A white reference was obtained using a white Teflon tile with >99% reflectance, whereas the dark reference (>0% reflectance) was obtained with the light source turned off and the camera lens completely covered with its opaque cap. In order to extract the actual infrared response of the samples, the influence from both the white reference image and dark current image is removed. The calibration image X_{cal} was calculated by using the raw hyperspectral image X_{raw} , white reference image X_{ref} , and dark reference image X_{dark} by

$$X_{cal} = \frac{X_{raw} - X_{dark}}{X_{ref} - X_{dark}}$$
(1)

2.4. Spectra Data Extraction

The calibrated hyperspectral image was segmented by using a simple threshold value as the average value of the background and tomato pixels to remove the effect of the background and to visualize only the tomato pixels. A 3×4 median filter was applied to each of the tomato images separately for removing the dead pixels, which is caused by a camera detector. Furthermore, the region-of-interest (ROI) step was performed on the segmented image to extract the spectral signature and statistical information. An ROI was manually selected of each tomato image. The pixel spectra were averaged for each band within the entire ROI of each tomato separately before further analysis. In total, 95 average spectra (wavelength range = 948–2494 nm) representing the 95 scanned tomatoes were calculated and saved. Because of low signal-to-noise ratio performance due to inefficiencies of the lighting system at certain wavelength regions, e.g., low light output in the >1600 nm, the spectra were limited to the range 1000–1550 nm for further analysis of the relationship between the results of the reference measurement (MC, pH, and SSC) and the HSI data using multivariate analytical methods in combination with preprocessing techniques. The image correction and ROI selection steps were incorporated using MATLAB software (Version 8, The Mathworks Inc., Natick, MA, USA).

2.5. Reference Measurements

Three attributes (MC, pH, and SSC) of each tomato were measured and used as indicators of chemical components in intact tomatoes. After acquiring the spectral images, each fruit was divided into two equal halves. One-half was used for MC determination and the other half was juiced using a juicing machine to determine pH and SSC. The MC, expressed in percent wet basis (% w.b.), was measured by the gravimetric method using a drying oven (HST-502M, Hanbaek Co. Ltd., Gwangju, Korea) at 70 °C until it reached a constant weight. The weight was measured using an analytical balance (EK-1200i, A & D Co. Ltd., Tokyo, Japan). Finally, MC was calculated. The SSC and pH of

the tomato samples were determined using a digital refractometer (model: PR- 32α , Atago Co., Tokyo, Japan) and a pH meter (SX-620 pH tester, San-Xin Instrumentation Inc., Shanghai, China), respectively. Figure 3 shows the distribution of MC, pH, and SSC measurement of intact tomatoes in this study.



Figure 3. Distribution of (a) moisture content (b) pH and (c) soluble solid content measurement of intact tomatoes.

2.6. Data Analysis

2.6.1. Spectral Preprocessing

In general, the spectral data (of the selected ROI) contain random noise and spectral variations generated by the camera or instrument; therefore, pretreatment of spectral data is essential before subjecting it to multivariate analysis [22]. The average spectra were processed independently using different preprocessing methods including moving weighted average smoothing, normalization, S–G first- and second-order derivative, multiplicative scatter correction (MSC), and standard normal variate (SNV) in order to remove any irrelevant information such as high-frequency random noise, baseline drift, signal-to-background ratio, and others [23]. The averaging technique is used to reduce the number of wavelengths or to smooth the spectrum of tomatoes. It is also used to optimize the signal-to-noise ratio [24]. A normalization method is employed to remove the scattering effect in the tomatoes spectra [25]. Derivative spectra are needed to remove baseline shift and super-imposed peaks [26]. The MSC transformation method is applied to compensate additive and multiplicative scatter effects included in the spectra [27]. SNV is another spectra correction method that removes the slope variation from the spectra generated by the scatter [28,29].

2.6.2. Development of Calibration Model

After preprocessing of the spectra, the samples were randomly divided into two sample sets. A calibration sample set consisted of 60 samples. This set was used for developing the calibration model using a leave-one-out cross-validation method. A prediction sample set, consisting of 35 samples, was used for prediction purposes. In this study, partial least squares (PLS) regression multivariate analysis was used to develop a calibration model for MC, pH, and SSC of tomatoes prediction. The PLS regression approach is one of the most popular chemometric algorithms for calibration model development due to its simplicity and small volume of calculations. The preprocessed spectral data were linked to MC, pH, and SSC of tomatoes using a PLS regression analysis to develop a calibration model. In the development of all the calibration and prediction models, 20 PLS factors were set up as a maximum. All types of spectral preprocessing and the development of the calibration and prediction models for MC, pH, and SSC of tomatoes prediction were performed using The Unscrambler®software (version 9.8, CAMO, Oslo, Norway).

2.6.3. Evaluation of the Calibration and Prediction Models

Performance of the calibration and prediction models were evaluated using several statistical parameters, including coefficient of correlation (*r*) between measured and predicted value [10], root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and number

of latent variables [23]. A good calibration model should have a high correlation coefficient (r_{cal}) and a low RMSEC. In addition, the difference between RMSEC and RMSEP should be small for a good calibration model [30]. Another important parameter is the number of latent variables used to explain the correlation of the model [21].

the complexity of the model [31]. The quality of the final model was evaluated according to the correlation coefficient of prediction (r_{pred}) and root mean square error of prediction (RMSEP) in the prediction set. A Student's *t*-test and prediction interval were also performed in order to evaluate the significance level of the developed model.

2.6.4. Chemical Images of MC, pH, and SSC in Intact Tomatoes

In this study, the beta coefficients yielded by the best PLS regression model were used as feature wavelengths for MC, pH, and SSC of tomatoes [32]. The PLS regression model was developed between the standardized spectral data and the reference MC, pH, and SSC of tomatoes and the resulting regression coefficient is called the beta coefficient. Because each pixel in a hyperspectral image has a spectrum, using HSI, the concentration of composition can be calculated for each pixel to visualize the distribution of the components in the sample [33]. In this study, the hyperspectral image was first unfolded into a 2-D matrix and then multiplied by the beta coefficient obtained from the calibration model. After multiplication, the resultant vector was folded back to form a 2-D image in which a 3×5 median filter was employed to enhance the visual display. The 2-D image is usually called a chemical image or prediction map, in which the spatial distribution of the predicted attribute is easily interpretable. The equation is used to develop the chemical image as follows:

Chemical image =
$$\sum_{i=1}^{n} I_i R_i + C$$
 (2)

where I_i is the ith image of *n* reflectance spectral images, R_i is the beta coefficient derived from the PLS regression model, and *C* is the constant of the PLS regression model. All the image-processing steps involved for visualization purposes were performed with the program developed using MATLAB software (Version 8, The Mathworks Inc., MA, USA).

3. Results and Discussions

3.1. Overview of Spectral Features and Statistics of Reference Analysis

Figure 4 shows the mean raw spectra of the intact tomato in the spectral range of 1000–1550 nm with resulting second-derivative preprocessed spectral profiles, which is similar with previous studies [34]. The near-infrared region (NIR) was sensitive to the concentrations of organic materials, which involved the response of molecular bonds C–H, O–H, and N–H [35]. The MC, pH, and SSC contain bonds of C–H, O–H, C–O, and C–C. Thus, it is possible to use this region for determination of MC, pH, and SSC in tomatoes. However, their absorption peaks overlap in several parts of the spectral region, resulting in the spectral profiles of tomatoes in the whole spectral region being quite even with some broadband peaks. There are peaks at 1100–1200 nm, and to a lesser extent at 1350–1500 nm, which may be associated with the second overtone of band C–H and the stretching first overtone of bond O–H in H₂O, respectively [35]. However, these peaks are usually located in wide spectral bands such that those key wavelengths that are helpful for predicting the MC, pH, and SSC of tomatoes cannot be directly identified. The proposed spectral preprocessing is first applied to raw spectral data and the preprocessed spectra were used to develop the PLS regression model.

An overview of MC, pH, and SSC distributions of tomatoes in the calibration and prediction sets is presented in Table 1. These statistic values include number of samples, range, mean, and standard deviation (SD). In this study, 95 samples were divided into the calibration and prediction sets (60:35). The range of the calibration set was from 91% to 95.9% for MC, 3.9 to 4.4 for pH and 2.7% to 5.5% Brix for SSC, and the range of the prediction set was from 91.2% to 94.4% for MC, 3.9 to 4.3 for pH, and 3.4% to 4.9% Brix for SSC. The range of the calibration set is bigger than that of the prediction set, which is helpful when developing a good model.



Figure 4. Typical reflectance spectra of intact tomatoes.

Table 1. Characteristics of calibration and prediction sample sets.

Items	MC (%)		рН		SSC (% Brix)	
	Calibration	Prediction	Calibration	Prediction	Calibration	Prediction
No. of samples	60	35	60	35	60	35
Maximum value	95.9	94.4	4.4	4.3	5.5	4.9
Minimum value	91.0	91.2	3.9	3.9	2.7	3.4
$\text{Mean}\pm\text{SD}$	93.6 ± 0.94	93.4 ± 0.92	4.1 ± 0.1	4.1 ± 0.08	3.9 ± 0.47	3.9 ± 0.44

3.2. PLS Regression Models

Using the PLS regression method, calibration and prediction models were developed for the various preprocessed spectra (Table 2). Among all of these calibration models, the S-G first-derivative preprocessed spectra-based model is better for MC, pH, and the smoothing preprocessed spectra-based model is better for SSC in intact tomatoes because of high correlation coefficient, minimal difference between RMSEC and RMSEP, and the minimal number of latent variables. PLS regression prediction results for MC, pH, and SSC are presented in the scatter plots shown in Figure 5. In all figures, the ordinate and abscissa axes represent the predicted and measured fitted values, respectively, of the corresponding parameters. The calibration correlation between the spectra and the MC of tomatoes was high, with r_{cal} from 0.81 to 0.88 and RMSEC from 0.44 to 0.54 (see Table 2). When the calibrated model was applied to the prediction set (35 samples), the results were applicable with $r_{vred} = 0.81$, RMSEP = 0.63% using the S–G first-derivative preprocessed spectra (see Figure 5a). This calibration model was better than that reported for MC in intact tomatoes using HSI and artificial neural networks in the range of 400–1000 nm, with a correlation coefficient (r_{pred}) of 0.773 [4]. For an online application, a smaller number of variables are important in order to develop a simple calibration model. In this study, the PLS model appeared to be acceptable since six factors (LVs) were used in the calibration model (see Table 2).

In the case of pH, a good regression correlation coefficient was obtained in the calibration set, with r_{cal} from 0.32 to 0.76 and RMSEC from 0.06 to 0.09 respectively (see Table 2). When the model was used to predict the samples, the best results were found with r_{pred} = 0.69, RMSEP = 0.06 using S–G first derivative (see Figure 5b). The PLS model appeared to be acceptable due to the two factors (latent variables, LVs) used in the calibration model (see Table 2). The pH content of intact tomatoes in the prediction set ranges from 3.9 to 4.3; the lack of large data variation was considered to be influential on the regression results. Although another study of pH prediction in strawberry using HSI showed

disparate findings, with standard error of prediction (SEP) values of 0.129, the models developed here for predicting pH displayed adequate predictive capacity for an online application [10].

Demonstra	Dream and in a Mathad	No. of LVs	Calibration		Prediction	
Parameter	Preprocessing Method		r _{cal}	RMSEC	r _{pred}	RMSEP
МС	Raw	7	0.81	0.44	0.79	0.56
	Smoothing (moving average) ^a	8	0.87	0.45	0.79	0.65
	Max normalization	7	0.88	0.44	0.71	5.74
	Mean normalization	6	0.84	0.51	0.76	0.64
	Range normalization	7	0.86	0.48	0.77	0.62
	S–G first derivatives ^b	6	0.85	0.50	0.81	0.63 *
	S–G second derivatives ^b	9	0.88	0.45	0.71	0.69
	MSC ^c	5	0.83	0.52	0.74	0.66
	SNV ^d	5	0.81	0.54	0.73	0.67
рН	Raw	3	0.72	0.06	0.65	0.06
	Smoothing (moving average) ^a	3	0.75	0.06	0.71	0.06
	Max normalization	2	0.33	0.09	0.50	0.08
	Mean normalization	2	0.32	0.09	0.50	0.08
	Range normalization	2	0.32	0.09	0.49	0.08
	S–G first derivatives ^b	2	0.75	0.06	0.69	0.06 *
	S–G second derivatives ^b	2	0.76	0.06	0.67	0.06
	MSC ^c	2	0.32	0.09	0.37	0.08
	SNV ^d	2	0.32	0.09	0.37	0.08
SSC	Raw	7	0.82	0.24	0.63	0.33
	Smoothing (moving average) ^a	5	0.80	0.28	0.74	0.33 *
	Max normalization	6	0.80	0.28	0.53	0.41
	Mean normalization	6	0.80	0.28	0.55	0.40
	Range normalization	6	0.79	0.29	0.54	0.39
	S– first derivatives ^b	7	0.78	0.30	0.65	0.36
	S–G second derivatives ^b	6	0.64	0.36	0.50	0.39
	MSC ^c	5	0.78	0.29	0.55	0.40
	SNV ^d	5	0.77	0.30	0.47	0.42

Table 2. Partial least squares regression results for the prediction of moisture content (MC), pH, and soluble solid content (SSC) for intact tomatoes with different preprocessed spectra.

Notes: ^a Number of segments: 13; ^b Number of segments: 13 and polynomial order: 2; ^c Multiple Scatter Correction; ^d Standard Normal Variate; RMSEC, and RMSEP are the root mean square error of calibration and prediction respectively, unit: % for MC, a.u. for pH and % Brix for SSC; *r_{cal}* and *r_{pred}* are the correlation coefficient of calibration and prediction respectively; S–G: Savitzky–Golay; LVs: Latent variables; * Best model.

For SSC in intact tomato measurements, the calibration correlation between the spectra and the SSC was as adequately high as 0.64–0.82, with the RMSEC ranging from 0.24% to 0.36% Brix (see Table 2). When the model was used to predict the samples, the prediction results were also desirable, with a correlation coefficient (r_{pred}) of 0.74 between the measured and the predicted values; the RMSEP was 0.33% Brix with smoothing preprocessed spectra (see Figure 5c). The PLS regression model appeared to be robust since only five factors (LVs) were used in the calibration model (see Table 2). Our results are consistent with the findings of Li et al. [35], who found a correlation coefficient (r) of 0.88 and RMSEP of 0.35% Brix for the prediction of SSC in pear using HSI with a spectral range of 930–2548 nm. By a 95% confidence paired t-test, there were no significant differences between the experimental values of MC, pH, and soluble solid content (SSC) and those predicted by HSI. These results demonstrate that a calibration model for prediction of internal qualities of intact tomatoes using HSI has been successfully developed and validated.

In the above PLS regression results, individual wavelength contributions by MC, pH, and SSC contents in tomatoes was not considered in the prediction results. This was because the PLS regression method first applied linear transform to the entire individual wavelength data [34]. As a result, it was

difficult to determine how individual wavelengths were directly related to the MC, pH, and SSC contents in tomatoes to be predicted. However, it would be helpful to examine how MC, pH, and SSC in tomatoes were simply related to individual wavelengths so that a better understanding of their correlated spectra might be achieved.



Figure 5. Prediction results of the established PLS regression models for chemical contents in intact tomatoes (**a**) moisture content (**b**) pH and (**c**) soluble solid content (SSC).

3.3. Chemical Images of MC, pH and SSC in Intact Tomatoes

Figure 6 showed a sequence of representative processed images, illustrating the application of the hyperspectral image processing, single-band and threshold methods for the prediction of the MC, pH, and SSC in intact tomatoes. The 1082 nm waveband image was used as a representative image for visualization purposes because it showed the higher contrast among the other bands. The background regions of the non-fluorescence black cup were eliminated from the image by using a 0.1 value of a threshold. The resultant image reveals the major area of the tomato from the background. It shows that the HSI technique is allowed to acquire multiple samples at a time that contains a complete spectrum for every pixel in each sample. It also allows for the visualization of the different chemical constituents in each sample based on their spectral signatures because regions of similar spectral properties have similar chemical composition.



Figure 6. (a) Tomato samples; (b) 3-D hypercube of tomato image; (c1) Corrected grayscale image at 1082 nm; (c2) Binary image at 0.1 threshold; (c3) Resultant grayscale image after masking with the previous image; (d1–d3) Chemical image of MC, pH and SSC respectively.

Figure 7 shows the chemical images/prediction map of the MC, pH and SSC of the intact tomato. The images were constructed by multiplying the obtained beta coefficient (regression coefficient) from the best preprocessed PLS regression model with the spectra of each pixel in the image. The power of these distribution maps resides in the rapid and easy access they afford to the spatial distribution of MC, pH and SSC in the tomato and their relative concentrations.

By including all the pixels, this approach has the advantage of displaying more detailed and accurate information. The difference in MC, pH and SSC within the same sample was very interesting and easily visualized in the concentration maps. The MC of the tomato showed uniform distribution along the fruits, while the pH was almost doubled in certain areas of the tomato. In the case of pH,

the tomato shows pH variation (red-yellow-blue color variation) with lower pH (more blue) in the central areas of the fruit compared to the peripheral areas of the fruit. In the case of SSC content, lower SSC showed uniform distribution in their peripheries and, to a greater extent, SSC towards their central parts in tomato. The whole-tomato differences in internal qualities such as MC, pH, and SSC may be caused by differences in sunlight exposure for the fruit surface during cultivation.



Figure 7. Prediction maps for tomato quality parameter (**a**) moisture content, (**b**) pH and (**c**) soluble solid content (SSC).

4. Conclusions

The development of the HSI system in the spectral range of 1000–1550 nm for rapid and nondestructive prediction of the MC, pH and SSC of intact tomatoes was investigated. The quantitative PLS regression model with appropriate preprocessed technique was established using full wavelengths and showed reasonable performance with r_{pred} of 0.81, 0.69, 0.74, and RMSEP of 0.63%, 0.06%, and 0.33% Brix. By applying beta coefficients yielded by the best PLS regression model, prediction maps were generated to visualize the levels of MC, pH and SSC to each pixel in the image. This pixel-wise chemical map was useful for interpreting the distribution of MS, pH and SSC in intact tomatoes in a simple and easy way. These findings indicate that the HSI technique combined with PLS regression has the potential to predict the chemical components of intact tomatoes. Although this result is promising, further study is needed to develop a robust model that is more broadly based on a wider range of different conditions such as sizes, storage, different maturity stages, seasons, etc.

Acknowledgments: This research was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ01125602), Rural Development Administration, Korea.

Author Contributions: A.R., M.S.K., H.L. and B.-K.C. conceived the structure of the paper; A.R., L.M.K. and S.L. collected the data; A.R., H.L. and C.M. analyzed the data; A.R., H.L. and B.-K.C. collected the references and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. FAOSTAT: Food and Agriculture Organization of the United Nations Statistics Division. Available online: http://www.fao.org/faostat/en/#data/QC/visualize (accessed on 10 November 2016).
- 2. Bhowmik, D.; Sampath Kumar, K.P.; Paswan, S.; Srivastava, S. Tomato—A natural medicine and its health benefits. *J. Pharmacogn. Phytochem.* **2012**, *1*, 33–43.
- 3. Petcharaporn, K.; Kumchoo, S. Calibration model of %titratable acidity (citric acid) for intact tomato by transmittance SW-NIR spectroscopy. *Int. J. Biol. Biomol. Agric. Food Biotechnol. Eng.* **2014**, *8*, 825–828.
- Mollazade, K.; Omid, M.; Tab, F.A.; Mohtasebi, S.S.; Zude, M. Spatial mapping of noisture content in tomato fruits using hyperspectral imaging and artificial neural networks. In Proceedings of the CIGR-AgEng2012: IV International workshop on Computer Image Analysis in Agriculture, Valencia, Spain, 8–12 July 2012.
- 5. Amigo, J.M.; Martí, I.; Gowen, A. Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food Structure, composition and quality. *Data Handl. Sci. Technol.* **2013**, *28*, 343–370.

- 6. Qin, J.; Chao, K.; Kim, M.S.; Lu, R.; Burks, T.F. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* **2013**, *118*, 157–171. [CrossRef]
- Mohammadi Moghaddam, T.; Razavi, S.M.A.; Taghizadeh, M. Applications of hyperspectral imaging in grains and nuts quality and safety assessment: A review. J. Food Meas. Charact. 2013, 7, 129–140. [CrossRef]
- Wu, D.; Sun, D.-W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals. *Innov. Food Sci. Emerg. Technol.* 2013, 19, 1–14. [CrossRef]
- 9. Huang, M.; Wan, X.; Zhang, M.; Zhu, Q. Detection of insect-damaged vegetable soybeans using hyperspectral transmittance image. *J. Food Eng.* **2013**, *116*, 45–49. [CrossRef]
- 10. ElMasry, G.; Wang, N.; ElSayed, A.; Ngadi, M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* **2007**, *81*, 98–107. [CrossRef]
- 11. Rajkumar, P.; Wang, N.; Elmasry, G.; Raghavan, G.S.V.; Gariepy, Y. Studies on banana fruit quality and maturity stages using hyperspectral imaging. *J. Food Eng.* **2012**, *108*, 194–200. [CrossRef]
- 12. Huang, M.; Wang, Q.; Zhang, M.; Zhu, Q. Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. *J. Food Eng.* **2014**, *128*, 24–30. [CrossRef]
- Gowen, A.A.; O'Donnell, C.P.; Taghizadeh, M.; Gaston, E.; O'Gorman, A.; Cullen, P.J.; Frias, J.M.; Esquerre, C.; Downey, G. Hyperspectral imaging for the investigation of quality deterioration in sliced mushrooms (Agaricus bisporus) during storage. *Sens. Instrum. Food Qual. Saf.* 2008, *2*, 133–143. [CrossRef]
- 14. Sivakumar, S.S. Potential Applications of Hyperspectral Imaging for the Determination of Total Soluble Solids, Water Content and Firmness in Mango. Master's Thesis, McGill University, Montreal, QC, Canada, 2006.
- 15. Li, X.; He, Y. Non-destructive measurement of acidity of Chinese bayberry using Vis/NIRS techniques. *Eur. Food Res. Technol.* **2006**, 223, 731–736. [CrossRef]
- Baiano, A.; Terracone, C.; Peri, G.; Romaniello, R. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Comput. Electron. Agric.* 2012, *87*, 142–151. [CrossRef]
- Nogales-Bueno, J.; Hernández-Hierro, J.M.; Rodríguez-Pulido, F.J.; Heredia, F.J. Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: A preliminary approach. *Food Chem.* 2014, 152, 586–591. [CrossRef] [PubMed]
- 18. Noh, H.K.; Lu, R. Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. *Postharvest Biol. Technol.* **2007**, *43*, 193–201. [CrossRef]
- 19. Leiva-Valenzuela, G.A.; Lu, R.; Aguilera, J.M. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. *J. Food Eng.* **2013**, *115*, 91–98. [CrossRef]
- Mendoza, F.; Lu, R.; Ariana, D.; Cen, H.; Bailey, B. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. *Postharvest Biol. Technol.* 2011, 62, 149–160. [CrossRef]
- 21. Schmilovitch, Z.; Ignat, T.; Alchanatis, V.; Gatker, J.; Ostrovsky, V.; Felföldi, J. Hyperspectral imaging of intact bell peppers. *Biosyst. Eng.* **2014**, *117*, 83–93. [CrossRef]
- 22. Kandpal, L.M.; Lee, S.; Kim, M.S.; Bae, H.; Cho, B.-K. Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B1 (AFB1) on corn kernels. *Food Control* **2015**, *51*, 171–176. [CrossRef]
- Rahman, A.; Kondo, N.; Ogawa, Y.; Suzuki, T.; Shirataki, Y.; Wakita, Y. Prediction of K value for fish flesh based on ultraviolet-visible spectroscopy of fish eye fluid using partial least squares regression. *Comput. Electron. Agric.* 2015, 117, 149–153. [CrossRef]
- 24. Cen, H.; He, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* **2007**, *18*, 72–83. [CrossRef]
- Downey, G.; Robert, P.; Bertrand, D.; Kelly, P.M. Classification of commercial skim milk powders according to heat treatment using factorial discriminant analysis of near-infrared reflectance spectra. *Appl. Spectrosc.* 1990, 44, 150–155. [CrossRef]
- 26. Swierenga, H.; de Weijer, A.P.; van Wijk, R.J.; Buydens, L.M.C. Strategy for constructing robust multivariate calibration models. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 1–17. [CrossRef]
- 27. Chen, H.; Song, Q.; Tang, G.; Feng, Q. The combined optimization of Savitzky-Golay smoothing and multiplicative scatter correction for FT-NIR PLS models. *ISRN Spectrosc.* **2013**, 2013, 642190. [CrossRef]

- 28. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]
- 29. Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P.A.; Massart, D.L. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *J. Pharm. Biomed. Anal.* **1999**, *21*, 115–132. [CrossRef]
- 30. Camps, C.; Christen, D. On-tree follow-up of apricot fruit development using a hand-held NIR instrument. *J. Food Agric. Environ.* **2009**, *7*, 394–400.
- 31. Gómez, A.H.; He, Y.; Pereira, A.G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *J. Food Eng.* **2006**, 77, 313–319. [CrossRef]
- 32. Kamruzzaman, M.; Makino, Y.; Oshita, S. Hyperspectral imaging for real-time monitoring of water holding capacity in red meat. *LWT Food Sci. Technol.* **2016**, *66*, 685–691. [CrossRef]
- Kamruzzaman, M.; ElMasry, G.; Sun, D.-W.; Allen, P. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innov. Food Sci. Emerg. Technol.* 2012, 16, 218–226. [CrossRef]
- 34. He, Y.; Zhang, Y.; Pereira, A.G.; Gómez, A.H.; Wang, J. Nondestructive determination of tomato fruit quality characteristics using VIS/NIR spectroscopy technique. *Int. J. Inf. Technol.* **2005**, *11*, 97–108.
- 35. Li, J.; Tian, X.; Huang, W.; Zhang, B.; Fan, S. Application of long-wave near infrared hyperspectral imaging for measurement of soluble solid content (SSC) in pear. *Food Anal. Methods* **2016**, *9*, 3087–3098. [CrossRef]



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).