

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Models Fitting to Pattern Recognition in Hyperspectral Images

Uziel Francisco Grajeda-González,
Alejandro Isabel Luna-Maldonado,
Humberto Rodriguez-Fuentes,
Juan Antonio Vidales-Contreras,
Ernesto Alonso Contreras-Salazar and
Héctor Flores-Breceda

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.73159>

Abstract

Worldwide, the concern on food safety, for example, on agriculture products, has become a topic with huge relevance. Nowadays, hyperspectral imaging systems for rapid detection of dangerous agents have emerged in response to these needs. In this research project, we proposed a new algorithm for *Salmonella typhimurium* detection on tomato surfaces in visible range (400–1000 nm). Gaussian model was used as a way to take out a model that could be calculated its definite integral; the final result of this algorithm is the area under curve (AUC), which gives a quantitative approach of spectral signatures. Three doses (5, 10, and 15 μL) and a control response (0 μL) were spread out on 20 tomatoes' surface. Subsequently, it was observed that some decrease responses with higher dose; also, numerically this pattern was seen with the help of AUC value. As a last step, a single factor analysis of variance showed no significance due to doses. Despite this outcome, the algorithm provides to be a good methodology for pathogen detection.

Keywords: food safety, hyperspectral imaging system, AUC, Gaussian model, ANOVA

1. Introduction

Hyperspectral imaging technology has been well developed in different areas of the industry, such as mining, quality assessment in food processes and detection of diseases that affect crops

and fruits, among others [1]; it is also important to mention that nowadays the reduction cost of sensors and electrical circuits has allowed the gradual immersion of hyperspectral imaging systems. Supervised learning is characterized by the need to know the expected responses based on human knowledge or the characteristics of the system; these responses are known as target function; then the system tries to compare our inputs (in our case the set of pixels) with this function, to the process of comparison, and testing of inputs with expected responses is called learning; the learning process ends when the algorithm has an acceptable level of performance; supervised learning can be grouped into two approaches such as classification and regression [2].

Classification approach: the data should be grouped into “categories,” for example, “infected,” “uninfected,” “damaged,” “undamaged,” and “mature.”

Regression approach: data are treated as continuous function that can be modeled with mathematical functions that predict behavior. Some examples of supervised machine learning algorithms are:

- Linear regression
- Support vector machines
- Supervised neuronal network

On the other hand, the unsupervised processes try to model the distribution of the data and thus to obtain conclusions; this type of algorithm group has similar characteristics along the data by itself, without the help of expected knowledge [3]. Accordingly, there are two approaches:

Clustering: in this type of analysis, the result is groups of data that share characteristics associated with certain trends, for example, the economy of a country with respect to the level of education of its population.

Association: in this type of analysis you want to find rules that describe a large portion of the data, such as “people who buy X also tend to buy Y.”

Examples of these unsupervised learning algorithms are:

- PCA
- PLSR
- Fisher’s discriminant analysis
- K-means clustering
- Unsupervised neuronal network

Choosing one of these methodologies to work depends on conditions of the experiment, that is, if into the experiment, the possibility to calibrate the algorithm with an expected response exists, for example, if an expert is able to detect damaged areas in a crop, before starting the research, this information could generate the target function and can be used to train the

algorithm [4]. Besides, data provided for hypercubes are usually analyzed by statistical pattern recognition approaches in three-dimensional space; these analyses come across from the simplest to the most complex; an additional way to getting relevant information from spectral data is analyzing its shape with curve fitting. Also, hyperspectral curve fitting methodology has the advantage of modeling multiple overlapping absorption, transmittance, and reflectance, with substantial less bands [5].

Moreover, wavelet is another technique that has impacted the way to analyze hyperspectral data. Due to its application on fields such as signal and image processing, pattern recognition, and data compression, wavelet transform has been an alternative for data analysis and dimensionality reduction [6]; the main idea of processing with wavelet transform is to decompose a signal into a series of shifted and scaled sub-representation of the mother wavelet function. This decomposition provides a hierarchical framework for interpreting the spectral information; some researchers have utilized wavelet transform for feature extraction, for example, classification of health and damage areas in leaves [7]. Other researchers [8] have studied the combination between PCA and wavelet coefficient to improve dimension reduction, and also they could highlight the small variations contained in spatial information. Another interesting application performed with wavelets was the fusion between hyperspectral and multispectral data [9]; the fusion image that proved to have more relevant information due to wavelets could be considered as a low-pass and high-pass filters that allow separate information which is not found with the naked eye.

Several researches that worked with modeling fitting and wavelet approach can be found in scientific literature; we are going to mention some of them: in Ref. [10], they investigated anomaly detection on a test data cube taken from a part of San Diego International Airport; in this research they proposed to use a Gauss-Markov algorithm to detect and classify statistical parameters within the data, that is, covariance matrix; as a result they show two binary images with 100% of target detection. It was developed a new algorithm [11] based on index total chlorophyll (C_{ab}) content; they proposed a new index called area under curve normalized to maximal band depth between 650 and 725 nm ($ANMB_{650-725}$); as a preliminary step, the area under the continuum-removed reflectance curve in the range of 650–725 nm ($AUC_{650-725}$) was computed. As an outcome, using area under curve (AUC) divided by a maximal band depth could predict chlorophyll content with good accuracy. It should be noted that despite the fact most of the current equipment operates between 400 and 2500 nm (visible and near infrared), it is important to select correctly the bands which contain the data where the area of interest is located; due to this fact, numerous works that focus on algorithms for band selection exist [12–14]. In addition, Ref. [15] compared different mathematical models for describing the hyperspectral scattering data in order to predict fruit firmness and soluble solid content (SSC) of Golden Delicious apples; the model utilized in the research was the Lorentzian distribution function, which gave a high fitting with an average correlation coefficient (r) greater than 0.995, owing to the oval shape of apples; it was necessary to calculate the integral of the measurement reflectance as a function of the area covered by the lens of the camera and the reflectance intensity I over the acceptance angle. As a conclusion, they mention that mathematical modeling of scattering data to obtain the total light reflectance, using an appropriate Lorentzian function, can provide a good way to predict apple fruit firmness and soluble solid content.

2. *Salmonella typhimurium* detection using hyperspectral imaging system

Foodborne detection has been a topic of interest in recent decades, due to food industry and government regulations. Traditional techniques based on agar culture media have huge shortcomings in rapid confirmation response and the inability to analyze a large number of samples; another disadvantage is the need to destroy the fruits in order to carry out the planting on the culture media. Moreover, hyperspectral imaging system has emerged as tool to detect bacteria in a considerable reduced time [16].

Specifically, *S. typhimurium* infection is usually transmitted by consumption of contaminated fruits, vegetables, fresh beef, or pork. Outbreaks caused by these bacteria have been reported in Canada, Europe, and the United States [17]; the symptoms of these bacteria are gastrointestinal problems, fevers, and in some cases death.

On the other hand, Mexican tomato production faces the challenge of complying with regulations imposed by the United States (USA) and Canada, where agricultural products must comply with safety features for sale in the foreign market; as well as economic losses in recent years due to the waiting for a long period of time and doubtful detection on infectious agents have caused the need for faster and more efficient detection methods [18].

This research project was focused on the obtaining of the hyperspectral signatures and Gaussian prediction models with high fitting to calculate the AUC and with this information detect *S. typhimurium* on tomato surface. Hyperspectral imaging system promises to be a good technique for worldwide food safety.

3. Materials and methods

3.1. Biotechnological material

S. typhimurium was used, suspended in a media culture (broth in cryopreservation state) necessary for its survival. The experiment utilized commercial selective media *Salmonella-Shigella* (SS) agar, Hektoen enteric, and xylose lysine deoxycholate (XLD) agar. To isolate the bacteria, the streak plate isolation method was used; to display and select the suspect colonies more easily, this procedure was performed in triplicate. Assay tubes with 5 mL of tetrathionate broth were inoculated with *S. typhimurium* strain. This culture medium contains peptone and sodium carbonate, and the selectivity is the result of the presence of sodium thiosulfate that generates tetrathionate when added at a ratio of 0.2% iodine-iodide solution and 0.1% of bright green to each tube, allowing the growth of bacteria containing the reductase enzyme tetrathionate, and inhibits the development of other accompanying microorganisms. The incubation time was 24 h at a temperature of 37°C under aerobic conditions.

3.2. Tomato samples

Twenty tomatoes (*Solanum lycopersicum* L.) variety “Roma” were selected in a state of postharvest; they were purchased at a local supermarket in the municipality of General Escobedo, Nuevo Leon, Mexico. The tomatoes complied with high visual quality.

3.3. Hypercube of contaminated tomato

3.3.1. Hyperspectral system

The hyperspectral equipment utilized for this research was the PIKE F210b (Alliend Vision Technologies, GmbH); the camera is coupled to a Spectograph ImSpector V10E (Specim, Spectral Imaging Ltd.); the hyperspectral system is attached to a linear translation structure, which is, essentially, a band, a motor, and a speed regulation stage. This is necessary due to the push-broom operation; besides, the spectral range of the equipment goes from 400 to 1000 nm. Finally, the system works with two halogen-tungsten bulbs with a power of 60 W.

3.3.2. Sample inoculation

In order to start the research, the first step was inoculating the surface of 20 tomatoes with *Salmonella typhimurium* bacteria at three different amounts of dosification; these were 5, 10, and 15 μL and a zone with no contamination (0 μL), as we can see in **Figure 1**. The spread of a little drop on the tomato surface it was carry out with the help a micropipette.

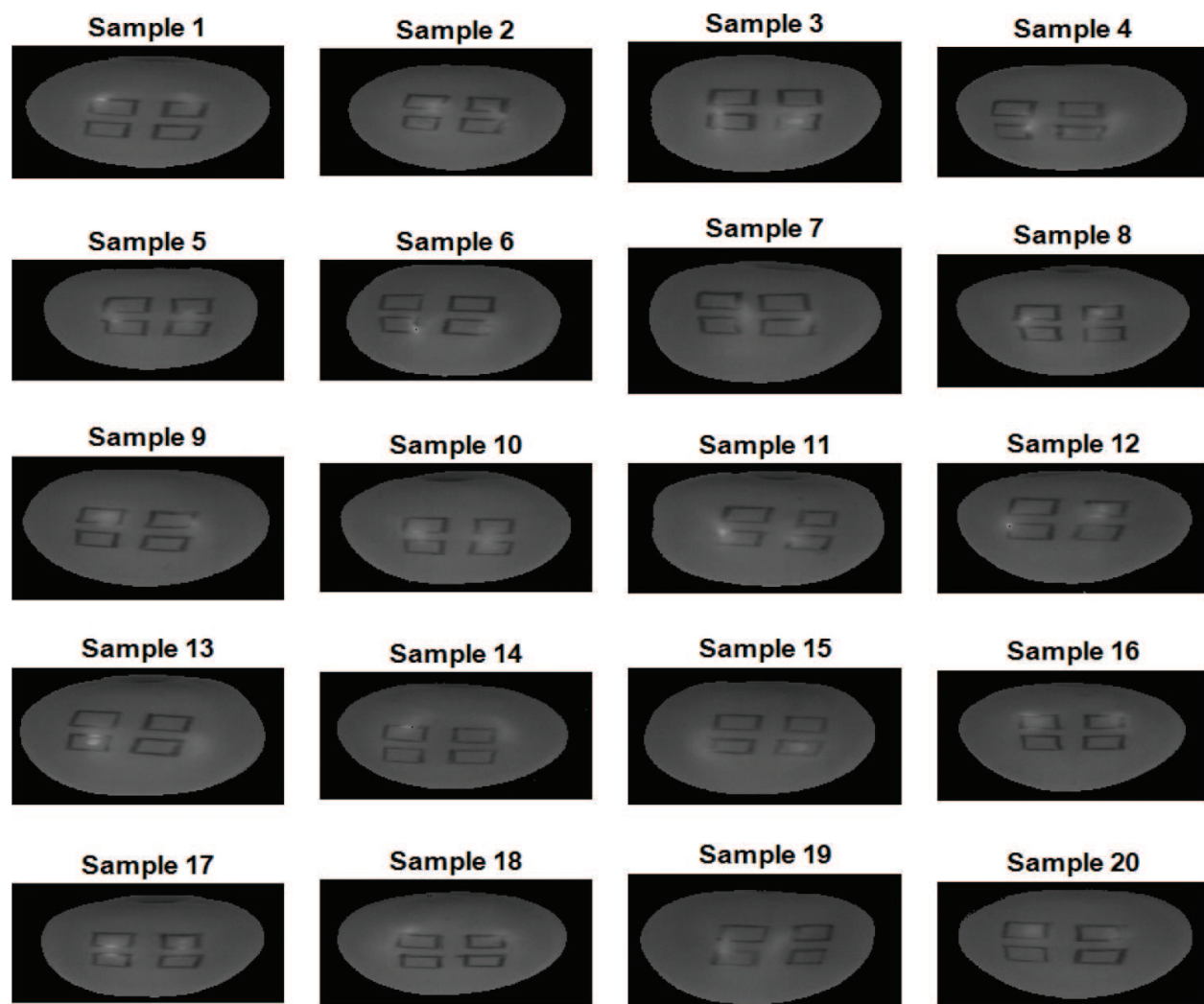


Figure 1. Tomato hypercubes shown in wavelength 692 nm; each tomato was labeled with four square-shaped zones which was spread with 0, 5, 10, and 15 μL of *S. typhimurium*.

20 hypercubes were obtained with a 600×1920 spatial resolution and 1080 bands with 12 bits of resolution.

3.3.3. *Preprocessing and data preparation*

Hyperspectral imaging processing usually has a pre-stage called preprocessing, necessary to remove the effect of death pixels, noisy signals, errors caused by analog to digital process conversion, etc. Additionally, due to high abundance of data, it required a calibration process and test hypercubes for correcting data [19]. A general workflow is shown in **Figure 2**, and its subsequent analysis is discussed in the next section.

3.3.4. *Normalization*

The analysis of hypercubes involves huge amount of data, thence one of the main reasons for be adapt the hypercubes to more manageable sizes and with this to improve the computer processing time.

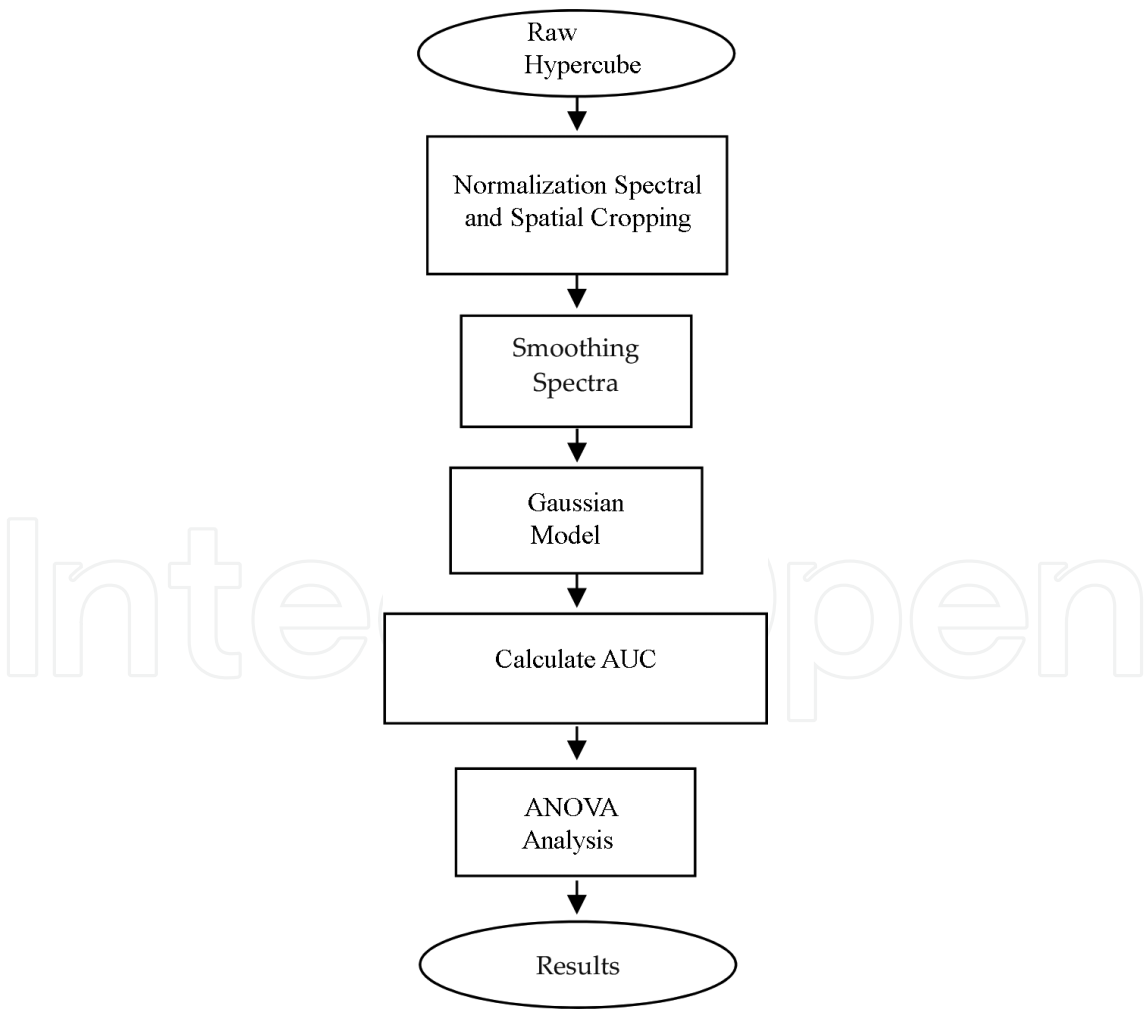


Figure 2. Workflow of proposed algorithm for pathogen detection, using area under curve.

As the first step, normalization of all hypercube was carried out using Eq. (1):

$$Hypercube_{Normalized} = \frac{RawHypercube - Dark_reference}{White_reference - Dark_reference} \quad (1)$$

where $Hypercube_{Normalized}$ is the calibrated hypercube, $RawHypercube$ is the total data without any type of process, $Dark_reference$ was taken with the absence of illumination and camera lens covered, and $White_reference$ data cube was generated with a high reflectance white mosaic and the lights on.

3.3.5. Spatial and spectral crop

As was mentioned before, each cube of data had a spatial dimension of 600×1920 . It should be noted that most of this information is merged with the background, which is not necessary to analyze, from there that a spatial cropping was necessary. Each cube was reduced to an average cube of 280×565 spatial dimension. On the other hand, sometimes it is not necessary to keep all data corresponding to start and end of the spectra, thereby a spectral crop was conducted in order to reduce no essential data.

3.3.6. Smoothing spectra

In hyperspectral preprocessing, the use of smoothing methods to remove high-frequency noise signal on the reflectance spectra is regular; a quite common smoothing method used in remote sensing is the Savitzky-Golay filter [20], which is based on least-squares polynomial approach applied on the short steps of wavelengths. In this procedure, a window of 11 steps, with a polynomial degree 2, was used. **Figure 3** shows all spectra after preprocessing mentioned above. Each spectrum is the result of a region of interest (ROI) averaging a contiguous quadratic shape of nine pixels.

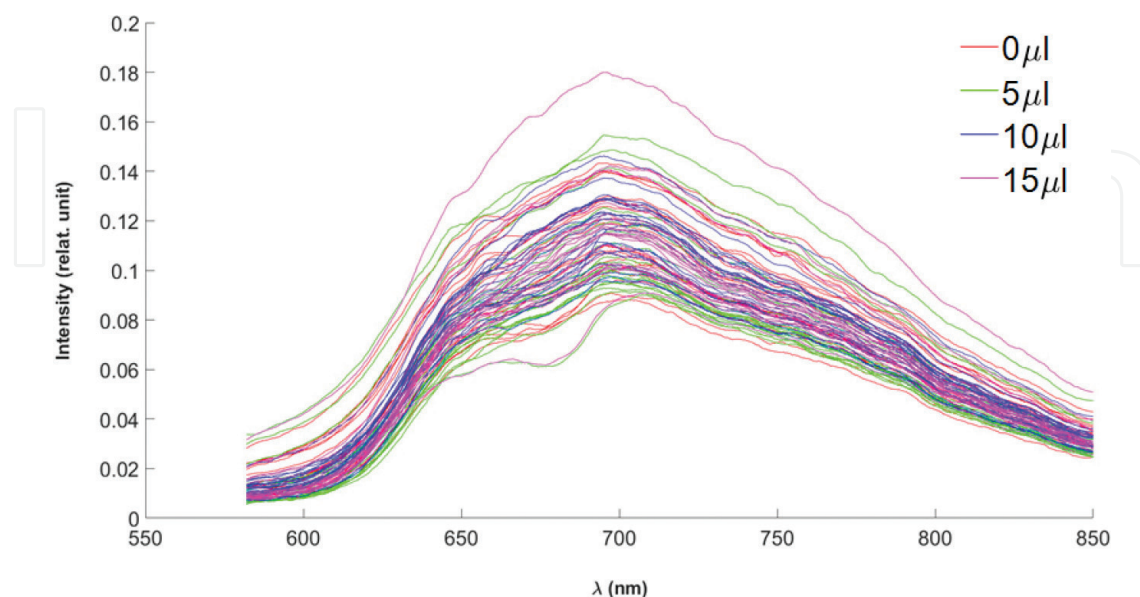


Figure 3. Spectra of 80 pixels extracted from 20 tomatoes; a ROI was selected in each contaminated zone.

3.3.7. Obtaining modeling of spectral signatures

On the other hand, in nature a bunch of data distribution is frequently located as a Gaussian or normal distribution (as it is shown in **Figure 2**), so that this model relates directly the behavior of the datasets. Gaussian curve fitting is still investigated as an algorithm for detecting patterns in biological, social, and physical sciences [21].

In order to compute Gaussian models for each spectrum, MATLAB 2016a and curve fitting tool (cftool) were used. A total of 80 models were obtained; after several tests and errors, the best combination found for modeling was Gaussian polynomial model with five terms as the form of Eq. (2):

$$f(x) = a_1 e^{-\left(\frac{x-b_1}{c_1}\right)^2} + a_2 e^{-\left(\frac{x-b_2}{c_2}\right)^2} + a_3 e^{-\left(\frac{x-b_3}{c_3}\right)^2} + a_4 e^{-\left(\frac{x-b_4}{c_4}\right)^2} + a_5 e^{-\left(\frac{x-b_5}{c_5}\right)^2} \quad (2)$$

where $f(x)$ is the Gaussian model, x is the wavelength independent variable, and $a_1, a_2, a_3, a_4, a_5, b_1, b_2, b_3, b_4, b_5, c_1, c_2, c_3, c_4, c_5$ are the coefficients to be calculated.

3.3.8. Computing area under curve and statistical analysis

The calculation of all areas was carried out, by calculating the definite integral (Eq. (3)). MATLAB 2016a provides an effective command called “quad” which numerically evaluates the integral, with an adaptive Simpson quadrature [22]:

$$A = \int_{wl_0}^{wl_1} f(x) dx \quad (3)$$

where A is the AUC; wl_0, wl_1 are the lower and upper limits, respectively, of wavelength; and $f(x)$ is a Gaussian model. Besides, the range between 582 and 850 nm was utilized, distributed into 482 bands.

After the areas under curves were obtained, a single factor analysis of variance (ANOVA) was performed in EXCEL 2016; the reason for this was to find out if any relationship between the dosage amount (every 5 μ L) and the decrease of the spectral signature response exists; a total of 20 tomatoes and 80 areas were analyzed.

4. Results and discussion

4.1. Gauss model results

Table 1 shows corresponding results of goodness of fit curve with Gaussians models. A low value in sum of squares error (SSE) is notorious, meaning that the model has a smaller random error component, since they are closer to zero [23]; as well as the coefficient of determination (R^2) has values higher than 0.9986; this proves high matching between the Gaussian model and the spectra responses to a certain dose. Besides, the other two parametric models for goodness of fit are adjusted R-sq and root-mean-square error (RMSE) which shows values higher than 0.99920 and less than 0.001, respectively.

Another mathematical approach to know the good fitting of one predicted model is known as residuals, defined as the differences between the response of original data and the response to predicted model (Eq. (4)) with regard to recognizing if the model was

Dose (μL)	SSE	R^2	Adj R-sq	RMSE
0	0.00045494	0.99908	0.99905	0.00091
5	0.00026017	0.99939	0.99937	0.00071
10	0.0006778	0.99864	0.99860	0.00109
15	0.00037805	0.99922	0.99920	0.00085

Table 1. Results of fit models; every value is the media of fit result of each dose.

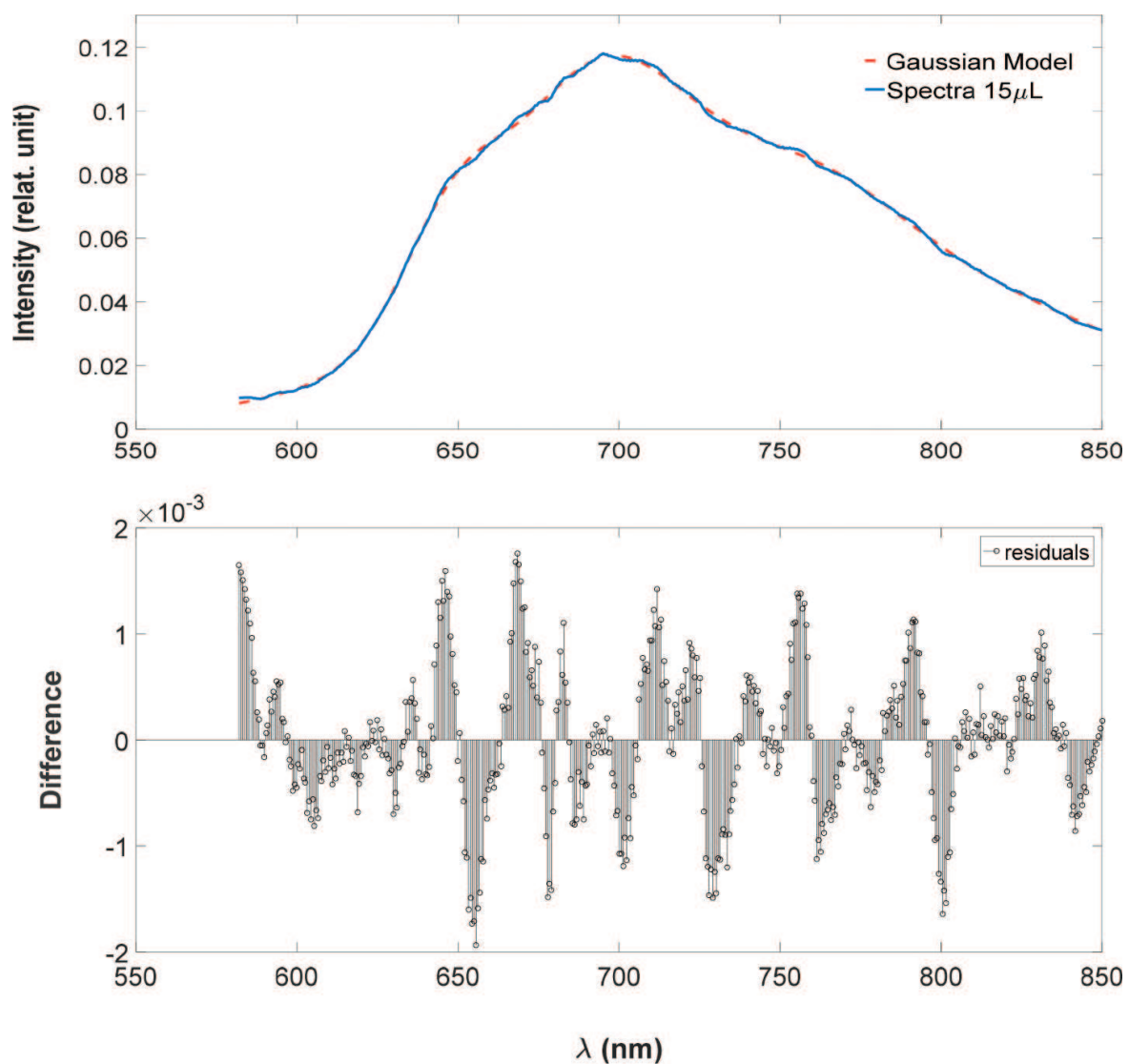


Figure 4. Upper graph shows an example of smoothed spectra (blue) and the predicted model (red); lower plot shows a random behavior on residuals, which means good prediction.

$$r = y - \hat{y}$$

(4)

where r are the residuals, y are the spectra of contaminated zones, and \hat{y} is the predicted model. An example of residual response is shown in **Figure 4**. Whether the plot of residuals seems to behave in a random way, it means that the model fits the data well; otherwise, if residuals appear to behave in a systematic pattern, then it is a clear case of mismatch between data and model [24]. In this research, the whole 80 models showed random residuals.

4.2. Areas under curve and their analysis

Areas extracted from all spectral signatures are shown in **Figure 5**. The trend in this dataset seems to decrease with higher dose in most subsamples; the meaning of this is greater absorbance on the infected surface; as an exception, tomato surfaces 3, 4, 5, 6, 10, 11, 13, and 15 do not seem to have this behavior; one possible explanation is closely related with orientation and position at the time of hypercube acquisition, that is, little light saturation zones.

As a last step, the results of calculation for a single factor ANOVA are shown in **Table 2**. Because P-value <0.05 means that there is no significance between doses and spectra response, a similar methodology was conducted by [25].

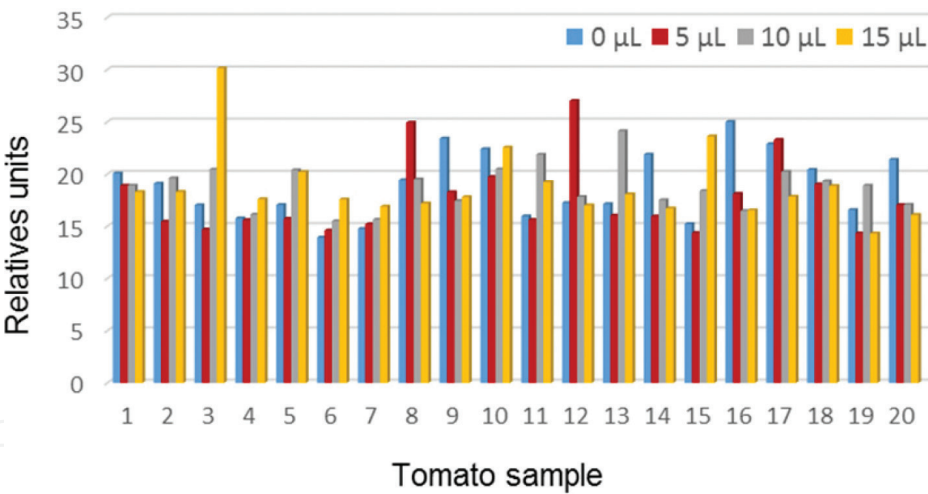


Figure 5. Total areas under curve of 20 tomatoes.

ANOVA						
Source of variation	SS	df	MS	F	P-value	F crit
Between groups	17.8996928	3	5.96656427	0.59978245	0.6171155	2.72494392
Within groups	756.038935	76	9.94788072			
Total	773.938627	79				

Table 2. Results of single factor ANOVA, taking AUC for data analysis.

5. Conclusion

Up to now, hyperspectral dataset analysis is carried out by different methodologies, algorithms, and techniques; in this research, we proposed to calculate AUC as an alternative for hypercubes; after AUC calculation, a single factor ANOVA would be enough for data analysis.

Despite results set down, it seems like visible range is not a good band for *S. typhimurium* detection. Secondly, sample orientation could improve results, since only a little inclination degree generated zone with high saturation because of the shiny nature of the tomato surface.

The novelty in this work was that there is little information related to the modeling of spectral signatures and their subsequent calculation of AUC as method to determine factors such as degree of contamination on fruits surface. Moreover, this methodology tries to quantify a spectral signature assigning it a value for understanding phenomenon that interacts with hyperspectral image systems. Future works could be related to improving AUC with different spectral responses in using variates fruit surfaces. Although there could be other variables to consider, which would affect the results as such, the scope of this work could be said to be a preliminar research.

Author details

Uziel Francisco Grajeda-González¹, Alejandro Isabel Luna-Maldonado^{1*},
Humberto Rodriguez-Fuentes¹, Juan Antonio Vidales-Contreras¹,
Ernesto Alonso Contreras-Salazar² and Héctor Flores-Breceda¹

*Address all correspondence to: alejandro.lunaml@uanl.edu.mx

1 Department of Agricultural and Food Engineering, Faculty of Agriculture, Autonomous University of Nuevo León, General Escobedo, Mexico

2 Department of Industrial Engineering, Culiacán Technological Institute, Culiacán, Mexico

References

- [1] Wu D, Sun D-W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part II: Applications. *Innovative Food Science and Emerging Technologies*. 2013;**19**:15-28. DOI: 10.1016/j.ifset.2013.04.016
- [2] Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag; 2006
- [3] Ben-David S, Shalev-Shwartz S. *Understanding Machine Learning: From Theory to Algorithms*; 2014. DOI: 10.1017/CBO9781107298019
- [4] Zhang YQ, Rajapakse JC. *Machine Learning in Bioinformatics*. John Wiley and Sons; 2009

- [5] Brown AJ. Spectral curve fitting for automatic hyperspectral data analysis. *IEEE Transactions on Geoscience and Remote Sensing*. 2006;**44**(6):1601-1607. DOI: 10.1109/TGRS.2006.870435
- [6] Hsu P-H. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2007;**62**(2):78-92. DOI: 10.1016/j.isprsjprs.2006.12.004
- [7] Hsu PH, Tseng YH, Gong P. Dimension reduction of hyperspectral images for classification applications. *Geographic Information Science*. 2002;**8**(1):1-8. DOI: 10.1080/10824000209480567
- [8] Gupta M, Jacobson N. Wavelet principal component analysis and its application to hyperspectral images. In: 2006 Int Conf Image Process. 2006. pp. 1585-1588. DOI: 10.1109/ICIP.2006.312611
- [9] Gomez RB, Jazaeri A, Kafatos M. Wavelet-based hyperspectral multi-spectral image fusion. In: 2001 Geo-Spatial image and data exploitation II. International Society for Optics and Photonics. Vol. 4383. pp. 36-43
- [10] Wang L, Gao K, Cheng X, Wang M, Miu X. A hyperspectral imagery anomaly detection algorithm based on gauss-Markov model. In: 2012 Fourth International Conference on Computational and Information Sciences. 2012. pp. 135-138. DOI: 10.1109/ICCIS.2012.21
- [11] Malenovský Z, Ufer C, Lhotáková Z, et al. A new hyperspectral index for chlorophyll estimation of a forest canopy: Area under curve normalised to maximal band depth between 650-725 nm. *EARSel eProceedings*. 2006;**5**:161-172. DOI: 10.5167/uzh-62112
- [12] Yang H, Du Q, Su H, Sheng Y. An efficient method for supervised hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*. 2011;**8**(1):138-142. DOI: 10.1109/LGRS.2010.2053516
- [13] Chang CI, Wang S. Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 2006;**44**(6):1575-1585. DOI: 10.1109/TGRS.2006.864389
- [14] Li S, Zhu Y, Wan D, Feng J. Hyperspectral band selection from the spectral similarity perspective. In: International Geoscience and Remote Sensing Symposium (IGARSS); 2013. pp. 410-413. DOI: 10.1109/IGARSS.2013.6721179
- [15] Peng Y, Lu R. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biology and Technology*. 2008;**48**(1):52-62. DOI: 10.1016/j.postharvbio.2007.09.019
- [16] Yoon S-C, Shin T-S, Lawrence KC, Heitschmidt GW, Park B, Gamble GR. Hyperspectral imaging using RGB color for foodborne pathogen detection. *Journal of Electronic Imaging*. 2015;**24**(4):43008. DOI: 10.1117/1.JEI.24.4.043008
- [17] Cito F, Baldinelli F, Calistri P, et al. Outbreak of unusual *Salmonella enterica* serovar Typhimurium monophasic variant 1, 4[5], 12:I-, Italy, June 2013 to September 2014. *Eurosurveillance*. 2016;**21**(15):1-10. DOI: 10.2807/1560-7917.ES.2016.21.15.30194

- [18] Ribera LA, Palma MA, Paggi M, Knutson R, Masabni JG, Anciso J. Economic analysis of food safety compliance costs and foodborne illness outbreaks in the United States. *HortTechnology*. 2012;**22**(2):150-156
- [19] James B, Geladi P. Hyperspectral NIR image regression part II: Dataset preprocessing diagnostics. *Journal of Chemometrics*. 2006;**20**(3-4):106-109. DOI: 10.1002/cem
- [20] Schmidt KS, Skidmore AK. Smoothing vegetation spectra with wavelets. *International Journal of Remote Sensing*. 2004;**25**(6):1167-1184. DOI: 10.1080/0143116031000115085
- [21] Gauch HGJ, Chase GB. Fitting the Gaussian curve to ecological data. *Ecology*. 1974;**55**(6): 1377-1381. DOI: 10.2307/1935465
- [22] MATLAB. Version 9.10.0 (R2016a). Natick, Massachusetts: The MathWorks Inc.; 2016
- [23] Steinley D. Validating clusters with the lower bound for sum-of-squares error. *Psychometrika*. 2007;**72**(1):93-106. DOI: 10.1007/s11336-003-1272-1
- [24] NIST/SEMATECH. e-Handbook of Statistical Methods. 2012. <http://www.itl.nist.gov/div898/handbook/>
- [25] Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*. 2003;**28**(7):916-931. DOI: 10.1016/S0306-4530(02)00108-7

