Trends in Food Science & Technology 50 (2016) 11–25

Contents lists available at ScienceDirect

Trends in Food Science & Technology

journal homepage: http://www.journals.elsevier.com/trends-in-food-science-and-technology

Review

# Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines

A.I. Ropodi, E.Z. Panagou, G.-J.E. Nychas*

*Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, Faculty of Foods, Biotechnology and Development, Agricultural University of Athens (AUA), Iera Odos 75, Athens 11855, Greece*

ARTICLE INFO

*Article history:*
Received 19 October 2015
Received in revised form
14 January 2016
Accepted 18 January 2016
Available online 23 January 2016

*Keywords:*
Spectroscopy
Hyper/multispectral imaging
e-nose
Chemometrics
Machine learning

ABSTRACT

*Background:* Food quality, safety and authenticity are important issues for consumers, governments, as well as the food industry. In the last decade, several researchers have attempted to go beyond traditional microbiological, DNA-based and other methods using rapid techniques. This broad term involves a variety of sensors such as hyperspectral and multispectral imaging, vibrational spectroscopy, as well as biomimetic receptors.

*Scope and approach:* The resulting data acquired from the above-mentioned sensors require the application of various case-specific data analysis methods for the purpose of simple understanding and visualization of the acquired high-dimensional dataset, but also for classification and prediction purposes.

*Key findings and conclusions:* It is evident that rapid techniques coupled with data analysis methods have given promising results in several food products with various sensors. Additionally there are several applications, new sensors and new algorithms that remain to be explored and validated in the future.

## 1. Introduction

At the dawn of the 21st century in an environment of tremendous technological progress and evolution of consumer life-styles, but also of economic problems, the European food industry is called to operate under seemingly contradictory market demands. While in general highly advanced technologies are rapidly being accepted and absorbed, the position of consumers with regard to their expectations for food products is often ambivalent. They seek food products of enhanced sensory quality, increased functional and nutritional properties combined with a traditional, wholesome image, as well as guaranteed safety but yet inexpensive foods with less processing, fewer additives and "technological" interventions. At the same time they expect extended shelf life and convenience in preparation and use.

To remedy this, the food industry and other stakeholders (*e.g.* Authorities from USA, EU and elsewhere, retailers) need to support development and application of effective quality and safety assurance systems based on controlling, monitoring, and recording the critical parameters throughout the food chain. This cannot be implemented with conventional microbiology (e.g. colony counting methods) or molecular based techniques that are considered more reliable and accurate (Velusamy, Arshak, Korostynska, Oliwa, & Adley, 2010). The fact that both are time-consuming, destructive and require highly trained personnel limits their potential to be used on-, in- or at-line (Nychas, Skandamis, Tassou, & Koutsoumanis, 2008; Papadopoulou, Panagou, Tassou, & Nychas, 2011). Furthermore, in the case of molecular tools, results may be misleading, as these techniques are focused so far on pathogenic rather than specific groups of the microbial association which contributes to spoilage and depends on storage and packaging conditions (Doulgeraki, Ercolini, Villani, & Nychas, 2012). This molecular approach is also costly, as high-tech instruments are required. In addition, due to the complexity of molecular techniques, the number of verified samples/measurements in many cases is severely limited. Therefore, efforts have been made to replace both conventional and molecular microbiological analyses with detection of biochemical changes occurring in food that could be used to assess food spoilage or safety.

Recently, some promising analytical approaches are being

---

* Corresponding author.
*E-mail address:* gjn@aua.gr (G.-J.E. Nychas).

http://dx.doi.org/10.1016/j.tifs.2016.01.011
0924-2244/© 2016 Elsevier Ltd. All rights reserved.

forwarded for the rapid and quantitative monitoring of food quality/safety. These approaches are based on (i) arrays of biomimetic sensors, (ii) vibrational spectroscopy (Fourier transform infrared -FT-IR, Raman) and (iii) surface chemistry (hyper/multispectral imaging) (Argyri, Panagou, & Nychas, 2014; Kamruzzaman, Maniko, & Oshita, 2014; Loutfi, Coradeschi, Mani, Shankar, & Rayappan, 2015; Nunes, 2014; Qin, Chao, Kim, Lu, & Burks, 2013; Sun, 2009; Teena, Manickavasagan, Mothershaw, El Hadi, & Jayas, 2013; Velusamy et al., 2010; Xiong, Sun, Zeng, & Xie, 2014). The usage of these sensors is based mainly on the principle that by-products of either metabolic activity of microorganisms, and/or the product's history itself and its origin display different biochemical profiles resulting in a characteristic sensor fingerprint that could be used for quality/safety evaluation (Ellis & Goodacre, 2001; Nychas et al., 2008). Besides, these techniques have several advantages compared to traditional methods since they are direct and non-invasive, require a minimum amount of sample or can be used on line. However, the enormous amount of data generated by this technology is very difficult to interpret. Indeed, although these rapid techniques are either non-invasive or require a minimum amount of sample, the acquired data are more complex and their analysis demands a multi-disciplinary approach. Depending on the type of sensor and data complexity, analysis involves disciplines such as computer vision/image processing, signal processing, statistical analysis/chemometrics, machine learning, computational intelligence techniques, etc. (Goodacre, 2003; Haralick & Shapiro, 1991; Hastie, Tibshirani, & Friedman, 2009; Liu, Wang, Wang, & Li, 2013; Loutfi et al., 2015; Marini, 2009; Nunes, 2014; Oliveri & Downey, 2012).

There are currently some reviews available on rapid methods applied in food commodities; however, they usually focus on a specific instrument, type of food, and application or data analysis methodology (Cheng & Sun, 2014; Cozzolino, Roumeliotis, & Eglinton, 2014; Domingo, Tirelli, Nunes, Guerreiro, & Pinto, 2014; Duchesne, Liu, & MacGregor, 2012; Kamruzzaman, Makino, & Oshita, 2015; Loutfi et al., 2015; Marini, 2009; Sankaran, Khot, & Panigrahi, 2012; Teena et al., 2013; Woodcock, O'Donnell, & Downey, 2008; Zhang et al., 2014). This article will focus on (1) data mining and data analysis derived from non-destructive, non-invasive instruments reported in the literature in recent years, regardless of sensor, food type and application, and (2) the aspects related to the implementation of these techniques to the food industry and other stakeholders, such as food managers who are not familiar with these approaches.

## 2. Instrumental techniques and resulting data

Data mining and data analysis from sensors, based on spectroscopy, hyperspectral and/or multispectral imaging, has been included in this review and the basic principles of each type of sensor and resulting data will be described briefly in this section.

### 2.1. Vibrational spectroscopy

*Vibrational Spectroscopy **(VS)*** is a collective term used to describe two analytical techniques – infrared and Raman spectroscopy. Infrared (IR) and Raman spectroscopy are non-destructive, non-invasive tools that provide information about the molecular composition, structure and interactions within a sample. These techniques measure vibrational energy levels, which are associated with the chemical bonds in the sample. The sample spectrum is characteristic, like a fingerprint, and vibrational spectroscopy is used for identification, characterization, structure elucidation, reaction monitoring, quality control, and quality assurance. Infrared and Raman spectroscopy provide complementary information about molecular structure. It is important to stress that IR and Raman spectroscopy have found a wide variety of applications in food quality analysis. In the case of IR, several studies have been applied in the evaluation of food spoilage, including animal origin foods such as meat, milk and cheese, as well as plant origin foods like wheat, fruit spirits and beer (Argyri et al., 2014; Damez & Clerjon, 2013). On the other hand, the studies reported for evaluating food spoilage through the use of Raman spectroscopy are rather limited, including food products such as meat and milk.

The resulting data for both types of spectroscopy are two dimensional, i.e. spectra consisting of a wavenumber and the value of the measured parameter. Fourier Transformation (FT-IR spectroscopy) may be used before exploiting the spectra. In Fig. 1(a) and (b), examples of FT-IR and Raman spectra of minced beef samples are presented, where for a specific wavenumber a single value of absorbance and intensity is acquired. Several methods for pre-processing before data analysis, and their combinations, have been proposed, as they are affected by noise and sometimes display baseline and scatter effects (Engel et al., 2013; Jarvis & Goodacre, 2005). While pre-processing methods tend to be case-specific, some of these methods include first or second derivatives, Wavelet Transform (WT), detrending, Multiplicative Scatter Correction (MSC) or its extended form (EMSC), and Standard Normal Variate (SNV) transformation (Argyri et al., 2014; Engel et al., 2013; Jarvis & Goodacre, 2005). In the available literature, usually one or more pre-processing methods are applied and -depending on the application-different methods provide different results (Biancolillo, Bucci, Magrì, Magrì, & Marini, 2014; Coppa et al., 2014). Lastly, a genetic algorithm was employed in order to decide the appropriate method or combination of methods for FT-IR spectra pretreatment (Jarvis & Goodacre, 2005).

### 2.2. Multispectral (MSI) and hyperspectral imaging (HSI)

*Hyperspectral and Multispectral imaging **(HSI – MSI)*** is a technique with which both spectral and spatial information from chemical targets can be obtained. This chemical imaging technique combines *Vibrational spectroscopy*, and *Computer vision*. The former technology is an optical technology that depends on the interaction between incident light and molecules in matter. As spectrometers analyze only a small portion of the food sample (therefore, the spectra, strictly speaking, are sometimes not representative of the whole sample), technologies that take into account the whole sample or a large part of it may provide more representative and detailed measurements. In order to obtain spatial information, another technology, namely, *Computer Vision* is available. This discipline imitates the principle of human vision, using three bands (red, green and blue) to acquire the characteristics of objects. Working in the visible range, the features obtained by computer vision include shape, color, size, and texture. However, only occasionally is this method reported to be sufficient for detecting chemical and biological parameters. Both spectroscopy and computer vision techniques have found a wide range of applications in the food industry. However, both techniques have their own disadvantages. The merits in spectroscopy and computer vision are both combined in ***hyperspectral imaging***, which can also be used to generate *chemical maps* to show distributions of parameters of interest. However, the rich information in hyperspectral imaging also results in difficulties in data processing, which makes it hard for industrial online applications. To overcome this problem, a simplified version called ***multispectral imaging*** (MSI) is available. The difference between the two lies only in the number of bands involved. For HSI, there are normally more than 100 bands, while for MSI, it is usually less than 20. The success of MSI deeply relies on the efficiency of HSI for providing the important wavelengths. In