Analytica Chimica Acta 891 (2015) 1–14

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Review

Data fusion methodologies for food and beverage authentication and quality assessment – A review

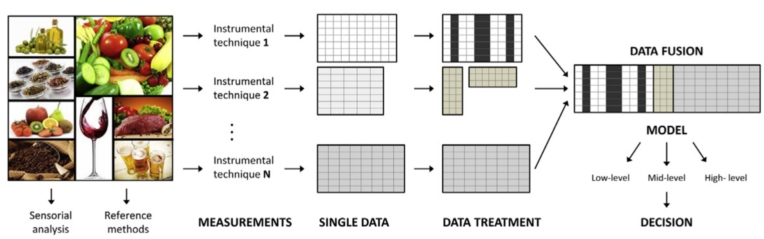
Eva Borràs [a], Joan Ferré [b,*], Ricard Boqué [b], Montserrat Mestres [a], Laura Aceña [a], Olga Busto [a]

[a] iSens Group, Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Campus Sescelades, 43007 Tarragona, Spain
[b] Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Campus Sescelades, 43007 Tarragona, Spain

HIGHLIGHTS

- Multivariate data fusion is used in food authentication and quality assessment.
- Data fusion approaches and their applications are reviewed.
- Data preprocessing, variable selection and feature extraction are considered.
- Model selection and validation are also considered.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 24 November 2014
Received in revised form
9 March 2015
Accepted 20 April 2015
Available online 24 April 2015

ABSTRACT

The ever increasing interest of consumers for safety, authenticity and quality of food commodities has driven the attention towards the analytical techniques used for analyzing these commodities. In recent years, rapid and reliable sensor, spectroscopic and chromatographic techniques have emerged that, together with multivariate and multiway chemometrics, have improved the whole control process by reducing the time of analysis and providing more informative results. In this progression of more and better information, the combination (fusion) of outputs of different instrumental techniques has

Abbreviiatons: AFS, acoustic firmness sensing; AIF, acoustic impulse resonance frequency; ANNs, artificial neural networks; ANOVA, analysis of variance; ATR, attenuated total reflection; BPNN, back-propagation neural networks; BY, bioyield test; CA, cluster analysis; CART, classification and regression trees; CIELab, CIE 1976 L*a*b* color space; CLV, cluster latent variables; CV, computer vision; CVA, canonical variate analysis; DAD, diode-array detection; DFA, discriminant function analysis; EEM, excitation–emission matrix spectroscopy; EU, European Union; EVOO, extra-virgin olive oil; FTIR, Fourier-transform infrared spectroscopy; GAs, genetic algorithms; GC, gas chromatography; GEP, generalized evidence processing theory; HCA, hierarchical cluster analysis; HL, high-level data fusion; HSI, hyperspectral scattering image; H-PCA/PLS, hierarchical PCA/PLS; HPLC, high-performance liquid chromatography; HS, head-space; ICA, independent component analysis; ICP, inductively coupled plasma; iPLS, interval PLS; IRMS, isotope ratio mass spectrometry; *k*NN, *k*-nearest neighbors; LC, liquid chromatography; LDA, linear discriminant analysis; LL, low-level data fusion; MB-PCA/PLS, multiblock PCA/PLS; MCR, multivariate curve resolution; MIR, mid-infrared spectroscopy; ML, mid-level data fusion; MLR, multiple linear regression; MOS, metal oxide semiconductors; MOSFET, metal oxide semiconductors field/effect transistors; MS, mass spectrometry; MSC, multiplicative signal correction; NIR, near-infrared spectroscopy; NMR, nuclear magnetic resonance; OPA, orthogonal projection analysis; OTR, oxygen transfer rate; PARAFAC, parallel factor analysis; PCA, principal component analysis; PCs, principal components; PCR, principal component regression; PDO, protected designation of origin; PGI, protected geographical indication; PLSR, partial-least squares regression analysis; PLS-DA, partial-least squares discriminant analysis; PNN, probabilistic neural network; POFNN, pseudo outer-product fuzzy neural network; QDA, quadratic discriminant analysis; RBF, radial basis function; REE, rare earth elements; RGB, red–green–blue color model; SBSE, stir bar sorptive extraction; SIMCA, soft independent modeling of class analogy; SNV, standard normal variate; S-PLS, serial PLS; SPME, solid-phase micro-extraction; SSC, soluble solid content; SVM, support vector machine; TOF, time of flight detector; TVB-N, total volatile basic nitrogen; UHPLC, ultra-high performance liquid chromatography; UNEQ, unequal class models; UV–vis, , ultraviolet–visible spectroscopy; VIP, variable importance in projection; VOO, virgin olive oil.

* Corresponding author. Tel.: +34 977 559 564.
E-mail address: joan.ferre@urv.cat (J. Ferré).

http://dx.doi.org/10.1016/j.aca.2015.04.042

Keywords:
Data fusion
Chemometrics
Authentication
Food
Beverage
Classification

emerged as a means for increasing the reliability of classification or prediction of foodstuff specifications as compared to using a single analytical technique. Although promising results have been obtained in food and beverage authentication and quality assessment, the combination of data from several techniques is not straightforward and represents an important challenge for chemometricians. This review provides a general overview of data fusion strategies that have been used in the field of food and beverage authentication and quality assessment.

Contents

1. Introduction

The consumer's demand for high quality food products is steadily increasing. The quality, especially of agricultural products, is specified in terms of a traceable origin, known chemical composition, adequate physical properties, satisfactory sensory evaluation, safety and health safeguards with respect to microbiological and toxic contamination and is influenced by the processing and storage of the products.

Fraudulent acts such as the adulteration with cheaper ingredients decrease the quality of the products, mislead the consumer and may imply a health risk. This is even more economically relevant for products that must comply with special laws of Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI) as stated in the European Union (EU) product quality policy of agricultural and rural development. Because of their much appreciated characteristics, such products have a high added value and are more prone to deceptive practices. In this sense, authorities are required to be able to assess the authenticity of a suspect product regarding the legal product description, detect fraudulent processing practices, prevent adulteration and control any other practices which may mislead the consumer such as mislabeling of geographical origin or composition of the product [1]. In order to protect producers and consumers from these fraudulent activities, the EU has established regulations with quality schemes determining specific origin and production processes of high valued food products [2], protecting consumers rights to receive truthful information about the food [3], assuring quality policy measures of specific products [4–6] and protecting geographical indications and designations of origin [7–9]. In addition, European research projects such as the TRACE project (TRACE: 'Tracing Food Commodities in Europe', No. FP6-2003-FOOD-2-A 6942 (2005–2009)) were funded to advance in the practice of verifying the origin of food products. One of the outputs of the TRACE project was that the combination of classical analytical methodologies and chemometric methods could determine characteristic patterns of compounds or parameters related to a geographical origin, the adulteration of samples or some specific conditions (e.g., processes, storage, harvest or variety).

Once authentication has been granted the main basic technique for food quality assessment from the consumer point of view is sensory analysis, which has a degree of subjectivity inherent to human perception. In recent years, much research has been performed to substitute the perception of human senses with 'artificial sensors', instruments providing signals related to the sensory attributes together with suitable multivariate pattern recognition techniques. Given the complexity of the foodstuff matrices, food quality derives from a complex combination of characteristics so analytical measurements for a single analyte or technique can rarely be correlated with the quality fulfillment. It is necessary to switch to multivariate