Food Control 73 (2017) 164–174

Contents lists available at ScienceDirect

Food Control

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

Discriminant analysis for unveiling the origin of roasted coffee samples: A tool for quality control of coffee related products

Paulo R.A.B. de Toledo [a], Marcelo M.R. de Melo [b], Helena R. Pezza [a], Aline T. Toci [c], Leonardo Pezza [a], Carlos M. Silva [b, *]

[a] Institute of Chemistry, State University of São Paulo – UNESP, 14800-060, Araraquara, SP, Brazil
[b] CICECO – Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, Aveiro, 3810-193, Portugal
[c] Latin American Institute of Science of Life and Nature, Federal University of Latin American Integration – UNILA, 85867-970, Foz do Iguaçú, PR, Brazil

ARTICLE INFO

Article history:
Received 13 January 2016
Received in revised form
1 August 2016
Accepted 1 August 2016
Available online 2 August 2016

Keywords:
Chemical markers
Coffee quality
Discriminant analysis
Geographic origin
Volatiles composition

ABSTRACT

Coffee quality is highly dependent on geographical factors. Based on the chemical characterization of 25 coffee samples from worldwide provenances and same roasting degree, Discriminant Analysis (DA) was employed to develop models that are able to identify the continental or country (Brazil) provenance of blind coffee samples. These models are based on coffee composition, particularly on several key compounds either with or without significant impact on aroma, such as 2,3-butanedione, 2,3-pentanedione, 2-methylbutanal and 2-ethyl-6-methylpyrazine. All models were validated with new and independent data from literature, and also through cross validation and permutation tests. Furthermore, the robustness of the proposed models in case of incomplete characterization data was also tested, being concluded that missing data is supportable by the models. In the whole, this article provides compelling arguments for the development of DA-based tools with the purpose of controlling the quality of coffee in terms of their continental and/or national origins.

1. Introduction

Coffee is one of the most ubiquitous edible products consumed around the world, playing a central economic role in several countries where it is produced and exported. From its ancestral origins in Africa, coffee cultivation wandered east and west, eventually forming a belt roughly bounded by the Tropics of Cancer and Capricorn (Smith AW, 1985). Nowadays, the top ten coffee-producing countries are Brazil, Vietnam, Colombia, Indonesia, Ethiopia, Honduras, India, Uganda, Mexico and Guatemala. For the season of 2014/2015, Brazil was responsible for more than a third of the overall world-scale coffee production, followed by Vietnam with 19.3% share (see Fig. 1). In the whole, a group of more than twenty countries produce coffee on a regular and sizeable basis.

In light of the diversified offer and to the fact that consumers started to value products with label of origin, the confirmation of coffee authenticity through chemical/physical analysis is of great relevance. Stakeholders such as importers or sellers are interested in the development of analytical methods able to demonstrate that the imported coffee had not been adulterated along the commercial chain, or really matches the expected origin and quality specifications. Such challenges represent investigation opportunities within the coffee research field.

Several countries adopted the certification known as Protected Designation of Origin (PDO) in order to protect and control the quality and provenience of their coffee as well as to boost their added value. This certification links the product to the specific culture methods, and operating and atmospheric conditions, as well as to the raw materials. While, for the consumers, PDO products are expected to have distinctive organoleptic features (characteristic of a given provenance), the sensorial spectrum that defines the flavor of coffee may be rather complex and subjective, which complicates a clear confirmation of samples origin. Such difficulties can only be circumvented if reliable and robust analytical based methods are developed for assessing quality parameters of coffee.

The most effective way to keep track of coffee quality and provenience is through the analysis of its volatile composition, which may be directly linked or not to the final aroma experienced by the customer. The definition of quality is thus not a simple task. An

---

* Corresponding author.
E-mail address: carlos.manuel@ua.pt (C.M. Silva).

http://dx.doi.org/10.1016/j.foodcont.2016.08.001
0956-7135/© 2016 Elsevier Ltd. All rights reserved.

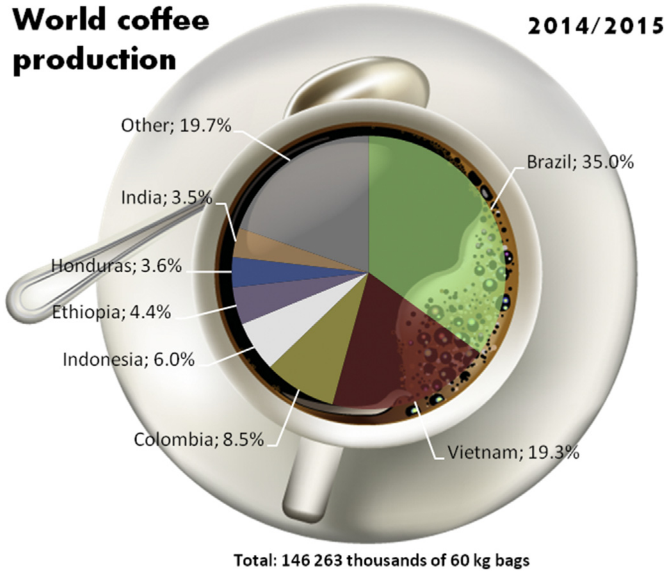

Fig. 1. World coffee production in 2014/2015 [USDA, 2015].

official definition provided by the International Organization for Standardization (ISO) refers that quality should be understood as follows: "*the extent to which a group of intrinsic features (physical, sensorial, behavioral, temporal, ergonomic, functional, etc.) satisfies the requirements, where requirement means need or expectation, which may be explicit, generally implicit, or binding*" (NBR ISO 9001, 2000). Remarkably, roasted coffee is one of the most aromatic food products, and is mainly consumed for the pleasure provided by its volatile components. The concentration of aromatic compounds in roasted coffee can reach 1 g/kg (Flament & Bessière-Thomas, 2002; Toci & Farah, 2014), and their characterization has been extensively studied over the years (Costa Freitas et al., 2001; Rocha, Maeztu, Barros, Cid, & Coimbra, 2004; Mondello et al., 2005; Yener et al., 2014).

The aroma of coffee is intrinsically related to the chemical composition of the beans, which undergo innumerable chemical modifications during roasting, generating a wide variety of volatile compounds. On the other hand, the chemical composition of the beans, and consequently their quality, is directly affected by a wide range of parameters, including species and variety of coffee, climate, soil, bean quality, blend type, post-harvest processing, type of roast, and storage. More than 800 volatile compounds have been identified in roasted coffee so far. These can be divided into different classes, including (in order of abundance) furans, pyrazines, ketones, pyrroles, phenols, hydrocarbons, acids and anhydrides, aldehydes, esters, alcohols, sulfur compounds, and others (Flament & Bessière-Thomas, 2002). Nonetheless, the desirable coffee aroma is produced by a delicate balance in the composition of volatiles, and it is believed that only about 5% of these compounds are actually odorous and capable of impacting coffee flavor (Yeretzian, Jordan, & Lindinger, 2003) (see Table 1). Among these compounds, pyrazines stand out, followed by furans, aldehydes, ketones, phenols, and sulfur compounds, among others (Akiyama et al., 2005; Czerny, Mayer, & Grosch, 1999; Maeztu et al., 2001; Sanz, Czerny, Cid, & Schieberle, 2002).

Arabica coffee is known to have a favorable growth at medium to high altitudes (1000–2100 m) and daily average temperatures around 18–22 °C, typical of equatorial regions. In addition, annual rainfall levels of 1500–2500 mm seem to favor this variety (Illy, 2005). On the other hand, Bertrand et al. (2012) noticed that pluviometric indices ranging from 807 to 1918 mm/year led to higher levels of volatile compounds known to impart negative notes to the coffee, such as 2-ethylhexan-1-ol (heavy, earthy, and slightly floral), and 3-methyl-2-butenoate (overripe fruity/ethereal) (Flament & Bessière-Thomas, 2002). Nevertheless, for coffees grown at high altitudes under annual rainfall levels of 1500–2500 mm, an increase of 5-methyldihydrofuran-2(3H)-one (γ-valerolactone) was noticed, which confers positive sweet/vanilla notes.

By evaluating the effect of temperature, Bertrand et al. (2012) noticed that, in comparison to coffee samples grown at lower temperatures, those cultivated under hot conditions evidenced notable increases of the concentrations of certain alcohols, such as 2-butoxyethanol, 2,3-butanediol and 1,3-butanediol. The last two compounds, which impart earthy and green flavors, have been associated with lower aromatic quality of coffee (Flament and Bessière-Thomas, 2002). In contrast, molecules like 2-methylfuran (caramel/nutty notes), 2-butanone (raspberry ketone/sweet-fruity odor) and methylthiomethane (dimethylsulfide-cabbage, sulfurous) suffer significant concentration reduction as temperature is increased (Flament and Bessière-Thomas, 2002). As a general statement, it is admitted the quality of Arabica coffee can be improved under fresher climatic conditions. In turn, Robusta coffee benefits from a hot and humid climate, lower altitudes (100–1000 m), and an average daily temperature of 22–26 °C, found in tropical regions (Illy, 2005).

The present article proposes the discrimination of coffee samples (mostly *Coffea arabica)* from different countries and continents, setting their volatiles composition as assessment criterion. Upon application of Discriminant Analysis (DA) to data, valid equations for provenance labeling are sought as tools to validate coffee samples origin. By compiling and using a database containing 25 coffee samples, this is the first attempt in the literature to reach such comprehensive classification through DA methods.

The document is structured in the following way: Section 2 is devoted to modeling; Section 3 comprises the database information, namely coffee samples (3.1) and coffee characterization (3.2). Results are presented in Section 4, in the following sequence: preliminary normalization of results (4.1), discrimination of samples according to their geographical origins (4.2), specific differentiation of Brazilian coffee samples (4.3), and geographic and environmental factors and coffee aromas (4.4). Finally, the main conclusions are drawn in Section 5.

2. Modeling

In view of the large amount and variety of volatile compounds found in roasted coffee samples, statistical approaches may be of special usefulness to treat and interpret experimental data. Within this context, multivariate analysis is a powerful tool, since not only considers individual direct impact of the volatile compounds, but also takes into account eventual correlations between them. Two of the most popular methods are Principal Component Analysis (PCA) and Discriminant Analysis (DA).

As far as PCA is concerned, it comprises a technique for shortening the size of a given collection of data without loss of their variance. For this the number of variables is reduced to a minimum, called principal components, that keep the information of the original data set. One particularity of PCA relies on the fact that it is a fully automated method that identifies itself the principal components without human specification of the groups (components) that should be formed. In this sense, PCA is rather suitable for the analysis of multidimensional data where crossed correlation (redundancy) may be present (Jackson, 1991).

Taking into account the data set of this article is not