Chemometrics and Intelligent Laboratory Systems 124 (2013) 14–20

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Soil sampling planning in traceability studies by means of Experimental Design approaches

Sara Totaro [a], Paola Coratza [b], Caterina Durante [a,*], Giorgia Foca [c], Mario Li Vigni [a], Andrea Marchetti [a], Mauro Marchetti [b], Marina Cocchi [a]

[a] University of Modena and Reggio Emilia, Department of Chemistry, via Campi 183, Modena, Italy
[b] University of Modena and Reggio Emilia, Department of Earth Sciences, Largo S. Eufemia 19, Modena, Italy
[c] University of Modena and Reggio Emilia, Department of Agricultural Science, Via Amendola 2 — Padiglione BESTA — 42122 Reggio Emilia, Italy

ARTICLE INFO

Article history:
Received 28 February 2012
Received in revised form 27 February 2013
Accepted 2 March 2013
Available online 16 March 2013

Keywords:
Traceability
Food
Soil sampling
D-onion optimal design

ABSTRACT

The present research is part of a project dealing with the development of analytical methodologies mainly based on primary indicators, such as isotopic ratios of stable and radiogenic elements, for the authenticity and geographical traceability of oenological food, typical of the Modena district. In particular, considering the objective of establishing a food-territory link by means of these analytical indicators, it is straightforward that the representativeness of sampling, for both food and soils, covers a primary role in the robustness of the traceability models.

With this aim, the issue of selecting a set of representative, informative and different soil samples is tackled. In this case, the goal is not obtaining a set of soil samples uniformly spanning the territory, since the planning of a punctual sampling of the whole district of Modena is not feasible considering the total number of affordable samples, but rather choosing a representative set of vineyards where the soil samples can be located. Thus, all the vineyard-registered producers of the district of Modena were considered and different variables (geological features of the soils, winegrowing coverage, grape varieties, yearly production of the farms, etc.) were handled with Experimental Design (DoE) techniques, to simultaneously take into account the different information in order to achieve a sustainable and rational site sampling. In particular, D-optimal onion design was chosen since it is widely used for mapping and planning purposes, hence it permits to achieve the maximum coverage and uniformity of the selected samples in the whole domain. An efficient mapping of the geographical region has been obtained ensuring coverage of farms characterized by the main grape production and insisting on soils with different geological features.

1. Introduction

Food characterised by high quality, safety and healthiness is a requirement for consumers and a primary objective for producers as well as industries and control authorities, which are in charge to take actions in order to verify food authenticity and traceability [1]. Moreover, consumers also associate high quality to a food as a function of its geographical origin. According to European Regulation 178/2000, the geographical origin of food should be guaranteed by means of traceability [2]. Unfortunately, the tracing and/or tracking procedures, while monitoring the flow of incoming and outgoing materials, are often limited since they are not referred to objective data but mostly based on "declarations" supported by paper documents. Hence, it is of the utmost importance to develop analytical tools able to 'certify' the provenience of food in order to accomplish food control and quality valorisation. In particular, information on geographical origin is required in the case of products awarded with Protected Designation of Origin (PDO). Therefore, in this context, one of the aims is to directly link the food to its territory of provenance using proper parameters, namely traceability indicators. These may be distinguished in *primary indicators*, i.e. elemental composition, stable isotope ratios etc., that can be directly linked to the same determinations in soil samples, and/or *secondary indicators*, i.e. variables related to food composition/making procedure, which can be regarded as foodstuff fingerprints and may allow discrimination from food products of different geographic origin, but cannot be directly linked to the soil [3]. Actually, in traceability studies, in order to link the food commodity to the territory of provenance, there are mainly three different analytical approaches that are adopted. In the first one, measurements of traceability indicators are carried out on a representative number of food samples and a possible relationship with its geographical origin is found by using multivariate classification models based on a calibration set of samples for which the origin is certain, while geographical, geological, and climatic reference information are taken from paper documentation [4–8]. In the second approach, the primary indicators are

* Corresponding author. Tel.: +39 059 2055028; fax: +39 059 373543.
E-mail address: caterina.durante@unimore.it (C. Durante).

0169-7439/$ – see front matter
http://dx.doi.org/10.1016/j.chemolab.2013.03.001

directly measured both in food as well as in some reference samples of the soil of origin [9–11]. In this case, no systematic selection/sampling of investigated soils is used. Finally, in the third approach, a representative soil sampling is considered as well, taking into account climate, geographical and geological features referred to extended macro-areas (i.e. 100 ÷ 200 km^2) [12]. The choice of one of these approaches mainly depends on the aim of the research. For example, when the extension of the investigated area is important and cannot be considered geologically homogeneous, it is necessary to map it, especially if information about the monitored indicators is not already available. In fact, since trace element composition as well as stable isotope signature in food are also affected by the geology of the territory, it is important to obtain an a priori knowledge of the investigated area. Obviously, in this case, the investigation of the whole 'population' of a soil is not feasible, and a systematic method of sub sampling has to be used.

Furthermore, before any soil sampling planning is devised, different parameters have to be considered and established such as the sampling procedures, the number of samples, the sampling sites, the time of sampling, and so on. Besides, the *productivity* variables of a farm (area, geographical positions and, for oenological productions, grape-varieties) should also be taken into account in order to evaluate the influence of a given production site on the whole density production.

In literature, there are several strategies, which have been employed for the planning of representative soil sampling [13–17]. For instance, systematic sampling can be obtained by using regular and circular grids, systematic and non-systematic patterns, and unaligned random sampling [13–16]. In general, these approaches take into account only few variables, e.g. geographical positions, and are usually aimed at uniformly sampling the studied areas. As an example, they are used to suggest a scheme of where to collect samples within a field or how to retrieve soil samples to acquire geological information in order to build the geological maps of a given area [14]. On the other hand, some authors developed a sampling design, mainly for food but which could be extended to soil, for the selection of a representative subset of sites with uniform distribution with respect to the production area and representative of the production density by means of two algorithms, i.e. Kennard–Stone and Potential Function [17].

Nevertheless, all these approaches do not take simultaneously into account all the relevant variables considering both soil features, in terms of geology, etc. and the productive context of the studied food commodity, in order to chose a representative set of soil sampling sites. This aspect, necessary to gain a clear and complete description of the investigated area and a proper selection of sampling sites, is the main focus of this study, which is part of a project dealing with the development of analytical methodologies for the authenticity and geographical traceability of typical food of a 'restricted' region. In particular, the present research is focused on the selection of soil sampling locations in productive sites, considering *productivity* variables as well as geological information as a selection criterion, by means of a multivariate approach. The food commodities that are objects of the project are two oenological-derived products typical of the district of Modena, which are awarded with PDO label, namely Aceto Balsamico Tradizionale di Modena (ABTM) and Lambrusco wines. In accordance with their production regulations [18,19], the raw materials, together with all their productive steps, are allowed only within the district of Modena, whose area is estimated in 2700 km^2, of which 90 km^2 are grape-cultivated. In particular, Lambrusco wines are divided in four additional geographical denominations, namely *Lambrusco di Sorbara*, *Lambrusco di Salamino di S. Croce*, *Lambrusco di Grasparossa* and *Lambrusco di Modena*. The growing of the grapes used for their production is located on an alluvial plan for the first two wines, on a hilly and alluvial plan area for *Lambrusco di Grasparossa* and on the whole Modena district for *Lambrusco di Modena* as well as for ABTM.

Therefore, with the aim of building traceability models for these matrices, here the issue of selecting a set of representative, informative and different soil samples is tackled. In this case, the goal is not obtaining a set of soil samples uniformly spanning the territory, since the planning of a punctual sampling of the whole district of Modena is not feasible considering the total number of affordable samples, but rather choosing a representative set of vineyards where the soil samples can be located. Thus, all the vineyard-registered producers of the district of Modena were considered and different variables were selected to characterize each of them, such as: i) the different geological, pedological and geo-morphological characteristics of the soils of their vineyards, ii) their winegrowing coverage, iii) the different grape variety and iv) their yearly productions. Experimental Design (DoE) techniques were employed, after having extracted and codified all the available information, in order to take into account simultaneously the different sources of information while achieving a sustainable and rational site sampling. In particular, D-optimal onion design [20] was chosen since it is widely used for mapping and planning purposes, hence it consents to achieve the maximum coverage and uniformity of the selected samples in the whole domain.

To our knowledge, this approach, based on DoE technique, constitutes a novelty with respect to established sampling protocols in literature.

2. Experimental and methods

As mentioned, the extension of Modena district is 2700 km^2, of which 90 km^2 is grape-cultivated, and two main areas can be defined on the basis of the geology: the alluvial plane zone (referred to as A) and the Apennine margin (referred to as B). For each of these main areas, all the informative variables, able to describe the soil geological, geo-morphological and pedological features, the winegrowing coverage and productivity of the wine-farms were considered. Then, each variable, when needed, was codified according to a binary coding (0/1: presence/absence) and the data were arranged in two bi-dimensional matrices, one for each area, **A** and **B**, and finally, a D-optimal based design of experiments was carried out on the score matrices obtained by means of Principal Component Analysis. All the procedure is described in detail in the following sessions.

2.1. Identification of candidate farms set and definition of "production variables"

The soils to be considered as candidate set for the DoE procedure were chosen among the fields of the farms (around 4600) belonging to the list of all the grape-vine producers enrolled in the Wine-Vine Register of the Modena district.

Information on these farms was obtained thanks to the willingness of the Agricultural Productions Office of the Province of Modena. In particular, the total extension of the farms (which can be located in one or more municipalities), their relative area in each municipality and the cultivated grape varieties were considered. This information could be directly related to the productivity of the farms. Indeed, it can be assumed that the farm production is directly related to its area.

After collecting these data, a first pruning of the farms was performed, limiting the selection to the farms where only the grape varieties listed in the production regulations of both ABTM [18] and Lambrusco wines [19] are cultivated, namely Lambrusco di Sorbara, di Grasparossa, di Salamino, Ancellotta, Fortana, Malbo Gentile, Trebbiano Modenese, Trebbiano Romagnolo, Trebbiano Toscano, Trebbiano Giallo and Sangiovese. Then, the four widest farms for each municipality were always taken into account; moreover, the ones differing at least for one of the cultivated grape varieties, with respect to the previously selected, were also included.

The areas of the fields of each farm (total and relative to each municipality, when a farm extends over more than one) together