

## Chapter 8

# Application of Support Vector Machines to Melissopalynological Data for Honey Classification

**Giovanna Aronne**

*University of Naples Federico II, Italy*

**Veronica De Micco**

*University of Naples Federico II, Italy*

**Mario R. Guarracino**

*Italian National Research Council, Italy*

### ABSTRACT

*In this paper, the authors address the problem of the discrimination of geographical origin and the selection of marker species of honeys using Support Vector Machines and z-scores. The methodology is based on the elaboration of palynological data with statistical learning methodologies. This innovative solution provides a simple yet powerful tool to detect the origin of honey samples. In case of honeys from Sorrento Peninsula, the discrimination from other Italian honeys is obtained with high accuracy.*

### 1. INTRODUCTION

In recent years, an increasing attention has been devoted in Europe to the determination of the botanical and geographical origin of honey. This is partially due to allow European product to compete more effectively with cheap honey appeared on the EU market. To cover the high

production costs, honey should be sold at much higher prices than those accepted by international markets. High quality standard, traceability and link to the geographical origin by means of strict regulation, such as Protected Designation of Origin (PDO), can give extra value to European honey. These laws protect a regional food, vegetable or fruit defining specific and objective procedures to determine the quality standard of the product and to verify the link to its geographical origin.

DOI: 10.4018/978-1-4666-0333-2.ch008

At present, the procedure to verify geographical origin of honey is not well established; some attempts have been done to overcome this situation, and a promising field of research relies on the application of statistical methods on data about pollen content of honeys (Aronne, 2010).

Floral honey always contains numerous pollen grains mainly from the plant species foraged by honey bees. Pollen analysis of honey, namely melissopalynology, is of great utility to determine and control both botanical and geographical origin. The determination of geographical origin is based on the entire pollen spectrum being consistent with the flora of a particular region and with any reference spectra or descriptions in the literature (Louveau, 1978; Herrero, 2001; Persano Oddo, 2004; Persano Oddo, 2007; Aronne, 2010).

Analysis of the geographical origin of honey is based on the assumption that the selection of honey species made by bees is influenced by the peculiarity of local vegetation. Therefore, the palynological component of a honey, if correctly analysed, should provide information on the foraging area. This research topic assumes considerable importance when its aim is to safeguard consumer interests and to protect honest producers of honeys labelled with the indication of geographical origin (Aronne, 2008).

Although some computer-aided methods for the classification of honeys have been developed (Battesti, 1992; Scala, 2004a; Scala, 2004b; Aronne, 2008a; Aronne, 2008b; Aronne, 2008c; Aronne, 2010) at the moment the identification of geographical origin depends mainly on the experience and the knowledge of the palynologists who are asked to compare results from specific samples with a hypothetical pollen spectrum from honey producible in the same geographical area. It is therefore evident that elaboration of melissopalynological data requires precise, sensitive analytical tools which go beyond the subjective evaluation, providing the means to correlate data and information which are otherwise elusive.

Starting from the assumption that palynological data contain valuable information on the vegetation characteristics of the area in which bees have foraged, we used analysis tools and techniques that have been successfully applied to other fields, such as genetics, agriculture, economics and computer science (Guarracino, 2008; Mucherino, 2009). We believe that *data mining* techniques and concepts from *statistical learning theory* could provide the methodology enabling analysis of the pollen content of honeys. This analysis is determined by data models, their analysis, implementation and the use of specific algorithms.

In this paper we report the application of these tools to melissopalynological data; our specific purpose was to test and define a new methodological proposal to trace the geographical origin of honeys. To this extent, we used the results of palynological analysis of chestnut honeys (honey produced from the nectar of *Castanea sativa* Mill.) from the Sorrento Peninsula (Southern Italy) and from other areas in Italy. According to Von der Ohe et al. (2004), a honey can be defined “chestnut honey” if at least 86% of its pollen grains are from *Castanea sativa*. As a consequence, in this kind of honey only the remaining 14% of the pollen component is responsible for differences between samples and it is quite difficult to evaluate the geographical origin of the samples. The aim of our work was to find a mathematical model able to distinguish the chestnut honeys produced in the Sorrento Peninsula from those produced elsewhere. In the following sections, starting from initial statistical evaluation of the data, we report a synthesis of the working phases. In next section, data used for experiments are described. In section 3, data preparation is detailed. In section 4, methods used for classification are given. In sections 5 and 6, results regarding classification and variable selection are discussed. Finally, in section 7, conclusions are drawn.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/application-support-vector-machines-melissopalynological/63759](http://www.igi-global.com/chapter/application-support-vector-machines-melissopalynological/63759)

## Related Content

---

### Modeling of current and future state of biodiversity in Central America using GLOBIO3 methodology

Denisse McLean R. (2011). *Land Use, Climate Change and Biodiversity Modeling: Perspectives and Applications* (pp. 349-375).

[www.irma-international.org/chapter/modeling-current-future-state-biodiversity/53760](http://www.irma-international.org/chapter/modeling-current-future-state-biodiversity/53760)

### Big Data Challenges and Opportunities in Agriculture

Maya Gopal P.S. and Bhargavi Renta Chintala (2020). *International Journal of Agricultural and Environmental Information Systems* (pp. 48-66).

[www.irma-international.org/article/big-data-challenges-and-opportunities-in-agriculture/244147](http://www.irma-international.org/article/big-data-challenges-and-opportunities-in-agriculture/244147)

### Sustainable Waste Management Challenges in Sri Lanka

B. F. A. Basnayake, R. T. K. Ariyawansa, A. K. Karunaratna, S. M. Werahera and N. Mannapperuma (2020). *Sustainable Waste Management Challenges in Developing Countries* (pp. 352-381).

[www.irma-international.org/chapter/sustainable-waste-management-challenges-in-sri-lanka/240084](http://www.irma-international.org/chapter/sustainable-waste-management-challenges-in-sri-lanka/240084)

### Potential Nitrogen Load from Crop-Livestock Systems: A Spatial Database for a Multi-Scale Assessment and Mapping

Marco Vizzari, Sara Antognelli, Mariano Pauselli, Paolo Benincasa, Michela Farneselli, Luciano Morbidini, Piero Borghi, Giacomo Bodo and Alessandra Santucci (2016). *International Journal of Agricultural and Environmental Information Systems* (pp. 21-40).

[www.irma-international.org/article/potential-nitrogen-load-from-crop-livestock-systems/163317](http://www.irma-international.org/article/potential-nitrogen-load-from-crop-livestock-systems/163317)

### Using Spatial Statistics Tools on Remote-Sensing Data to Identify Fire Regime Linked with Savanna Vegetation Degradation

Anne Jacquin and Michel Goulard (2013). *International Journal of Agricultural and Environmental Information Systems* (pp. 68-82).

[www.irma-international.org/article/using-spatial-statistics-tools-remote/76653](http://www.irma-international.org/article/using-spatial-statistics-tools-remote/76653)