Editorial of the special issue on the occasion of the 12th European Symposium on Statistical Methods for the Food

Title: Editorial du Numéro Spécial à l'occasion des 12èmes Journées Européennes "Agro-Industrie et Méthodes Statistiques"

Jean-Pierre Gauchi[1]

This special issue brings together representative works featured at the 12th European Symposium "Statistical Methods for the Food Industry", which took place from 29 February to 2 March 2012 at the AgroParisTech facilities in Paris.

The objective of this conference, typically focusing on French achievements in the past, has been fully opened to our European colleagues in the last few years. For this 2012 event, this initiative was witnessed by the presence of five speakers, all European and non-French.

In recent years, the scope of the historical and founding themes of this conference - statistical methods in agriculture, strictly speaking, and sensometrics – has been extended to include statistical methods in food (processes, microbiology and risk assessment) and in chemometrics (analysis and control methods, experimental design). The title of this special issue covers all of these aspects.

This special issue includes nine articles on applied statistics. The first three articles describe methods applied in the field of agronomy. The fourth is devoted to sensometrics, the fifth is typical of the methods used in chemometrics (but taking examples from agronomy). The sixth and seventh, concerning the field of food microbiology, even if they deal with very different topics, are intended to ultimately improve the quantitative assessment of dietary risks. The eighth provides a contribution to the control of industrial processes that can be applied to the food industry. Finally, the ninth proposes an extension of PLS regression for Beta distribution errors, with examples of applications.

The first article, written by Jukka Ranta, Antti Mikkela, PirkkoTuominen and Helene Wahlström, is entitled "Bayesian risk assessment for Salmonella in egg laying flocks under zero apparent prevalence and dynamic test sensitivity". This is a very in-depth article on the crucial issue of food safety risks, since salmonellosis is one of the most common diseases encountered in industrial poultry production in the world. To describe the prevalence, the authors use a hidden Markov process model with two states, in continuous time. The infection status of a herd is treated as a hidden binary variable that can be detected as positive for the presence of salmonella only by

---

[1] INRA, Unité de Mathématiques et Informatique Appliquées (MIA), Jouy-en-Josas, France.
E-mail : Jean-Pierre.Gauchi@jouy.inra.fr

imperfect microbiological tests. The sensitivity of the test depends on the type of sampling and analysis used, as well as on the unknown phase of the epidemic among the poultry population.

The second article, written by Robert Sabatier, Myrtille Vivien and Christelle Reynès, is entitled "A new proposal, Multiway Discriminant Analysis : STATIS-LDA". The general framework is one where groups of observations are the same for all tables. The method is based on the joint use of discriminant analysis (FDA), common (in the sense of Fisher), and the well-known STATIS approach, thus the name of the method. Its principle is based on the inter/intra decomposition (specific to FDA) of each block, and the quantification of the importance of each block for discrimination. The authors emphasize the robustness of the method as illustrated by two convincing examples, the first based on simulations and the second related to Loire wines.

The third article, written by Aida Eslami, El Mostafa Qannari, Achim Kohler and Stéphanie Bougeard, is entitled "Multivariate data analysis of multi-group datasets". The specificity of these factor analyses is that they take individuals in structured groups into account a priori. The basic method is the multi-group principal component analysis of Krzanowski (1984), and the contribution of the publication is essentially a new optimization criterion to implement this method. The method is then extended to the case of structured data in both multi-groups and multi-tables. The special case where the different tables relate to the same variables is also considered. The two real cases covered include a set of data on Portuguese wines for the purpose of highlighting the typology of the two categories of wines, and a set of olive oils with the aim of exploring relationships between physico-chemical and sensory variables conditional on the geographic origin of three groups of olive oils (Greece, Spain, Italy).

The fourth article, concerning sensometrics and written by Rune Haubo Bojesen Christensen and Per Bruun Brockhoff, is entitled "Analysis of sensory ratings data with cumulative link models". Instead of analyzing data from hedonic studies using ordinal regression, as is customary in the sensometrics community, the authors propose to improve these analyses using cumulative link models. They also show the relationship of these models with Pearson omnibus tests and how they can lead to more powerful tests in situations without repetitions. Several examples from the sensometrics literature are given.

The fifth article, concerning chemometrics and written by Lidwine Grosmaire, Christelle Reynès and Robert Sabatier, is entitled "Joint selection of wavenumber regions for MidIR and RAMAN spectra and variables in PLS regression using Genetic Algorithms". With this type of selection, few techniques exist to select intervals for spectra (Raman and MidIR). The authors' proposal is based on a genetic algorithm for PLS1 regression, especially suited to this type of data that are represented by a multi-table (both types of spectra and a table of physico-chemical variables). The method is applied to cassava processing.

The sixth article focuses on optimal designs for experiments and is written by Dries Telen, Filip Logist, Eva Van Derlinden and Jan Van Impe. It is entitled "On the trade-off between experimental effort and information content in optimal experimental design for calibrating a predictive microbiology model". In predictive microbiology, mathematics and dynamic models are developed to describe microbial evolution as a function of environmental conditions. It is known that optimal experimental design techniques make it possible to obtain an "optimized" parameter estimation of these models, but an important additional question concerns the optimal sampling scheme, i.e., the experimental effort to be provided. In this article, an optimized sampling scheme based on a bang-bang approach is first implemented. This leads to a compromise between

the experimental effort and the information content in the optimal experimental designs. A multi-objective optimization type is then implemented to take both an optimality criterion of the experimental design and the quantification of the level of the experimental effort into account at the same time. Simulations are carried out using a model known to the food microbiology research community.

The seventh article, also dealing with food microbiology but oriented towards quantitative risk analysis, was written by Laurent Guillier, Jean-Marc Kabunda, Jean-Baptiste Denis and Isabelle Albert, and is entitled "Elicitation for food microbial risk assessment : a probabilistic approach extending Risk Ranger proposal". Using Risk Ranger software implemented by an Excel spreadsheet, the authors propose to develop this software into a probabilistic version. This improvement is based on an elicitation process variability using two quantiles of the underlying distribution of the variable of interest for each question proposed by the experts. The experts are also asked about their level of confidence for each quantile, leading to the incorporation of a level of uncertainty into the interest variable analyzed. These new additions make the Excel sheet more efficient. The improved software is then applied to the quantitative assessment of the risk of intoxication by histamine, the most frequent type of poisoning involved in the consumption of fishery products in France.

The eighth article, written by Ndèye Niang, Flavio Fogliatto and Gilbert Saporta, is entitled "Non parametric on-line control of batch processes based on STATIS and clustering". Batch processes are widely used in the industrial sector, particularly in the food, chemical and pharmaceutical industries. The authors propose an approach to the quality control of batch processes based on the STATIS method and non-parametric confidence regions obtained from convex hulls. This general approach can be used to control batch processes at the end of production, as well as during production after a classification step under constraints, based on a multivariate extension of the algorithm of W.D. Fisher. The proposed method is illustrated using real data from a batch process at a fixed time, and concerns an industrial chemistry problem.

The ninth article, written by Frédéric Bertrand, Nicolas Meyer, Michèle Beau-Faller, Karim El Bayed, Izzie-Jacques Namer and Myriam Maumy-Bertrand, is entitled "PLS Beta Regression". In the general context of regression, when the analyzed response is naturally expressed in the form of rate, proportion or index, with values that are necessarily included between zero and one or, more generally, with two fixed values known in advance, the Gaussian error framework is no longer appropriate. The context of the Beta distributions seems more conducive, especially since the density functions of Beta distributions can take on a wide variety of forms, thus providing good flexibility to model the data. In addition, if the predictors present multicollinearity problems (or worse, when they are more numerous than the observations), the PLS (Partial Least Squares) regression will be very efficient. The authors propose to combine the advantages of Beta distributions and PLS regression to develop a new method known as Beta PLS regression. The application of this regression method could be very useful in many fields, from agronomy to chemistry, and including medicine and econometrics.