

# AgroStat



Marseille, 14-16 March 2018

Due to the increasing quantity of data in agrosociences, there is a need for specific tools which place statistics and data science at the heart of challenges of the contemporary world. The AGROSTAT conference gives statisticians, engineers and users of statistical methods a unique opportunity to exchange around topics, such as sensometrics, chemometrics, experimental designs, risk analysis, process control or big data.

This event brings together internationally recognized academic and industrial organizations representatives, to take stock of advances in statistics, express their needs and to anticipate future challenges.

This conference, which is held every two years, is organized this year by **Aix-Marseille University**, the "Mediterranean Institute of Biodiversity and Marine and Continental Ecology", UMR CNRS 7263 / IRD 237, team Toxicology & Environmental Health (TSE), under the auspices of the Agro-Industry Group of the French Statistical Society (SFdS). The SFdS is a non-profit organization bringing together researchers, engineers, teachers and statistics users.

### **SCIENTIFIC COMMITTEE:**

- El Mostafa QANNARI (Oniris, France – President)
- Michelle SERGENT (Aix-Marseille University, France - Co-presidente)
- Stéphanie BOUGEARD (Anses Ploufragan, France)
- Damien BREMAUD (L'Oréal, France)
- Per BROCKHOFF (Technical University of Denmark, Denmark)
- Philippe COURCOUX (Oniris, France)
- Bernadette GOVAERTS (UC Louvain, Belgium)
- El Mestafa HADRAMI (FST, Fes, Morocco)
- Sébastien LE (AgroCampus Ouest, France)
- Sébastien MARQUE (Capionis, France)
- Hal MCFIE (Hal MacFie Sensory Training Ltd, Great Britain)
- Nicolas PINEAU (Nestlé, Switzerland)
- Michael MEYNER (P&G, Germany)
- Pascal SCHLICH (INRA Dijon, France)
- Douglas RUTLEDGE (AgroParisTech, France)

### **ORGANIZING COMMITTEE :**

- Michelle SERGENT (Aix-Marseille University, IMBE, France)
- Magalie CLAEYS-BRUNO (Aix-Marseille University, IMBE, France)
- Valérie DEYRIS (Aix-Marseille University, ISM2, France)
- Franck TORRE (Aix-Marseille University, IMBE, France)
- Marine BERRO (Aix-Marseille University, IMBE, France)

## Wednesday 14 March

9h00	Welcome speech - M. SERGENT, M. QANNARI		
<i>Inaugural conference</i>			
9h15	PL1	B. K. Ersbøll	Big Data from Farm to Fork, advantages and challenges
<u>Session 1: BIG DATA/MACHINE LEARNING/DEEP LEARNING - Chair: S. Marque</u>			
10h20	O01	P. Rebenaque	Automated analysis of tasting comments in sensory analysis
10h40	O02	M.-B Blanquart	Impact of the questionnaire structure on overall results in preference mapping: a meta-analysis on 285 consumer studies
11h00	O03	S. Bougeard	Current multiblock methods: competition or complementarity? A comparative study in a unified framework
11h20	<i>Coffee break</i>		
<u>Session 2: DEVELOPMENT TOOLS - Chair: D. Brémaud</u>			
11h50	O04	N. Pineau	Use of R-Shiny apps to communicate sensory and consumer modeling tools outputs
12h10	O05	I. Rebhi	An interactive shiny tool for sensory and consumer data mapping : sensmapui
12h30	<i>Lunch</i>		
<u>Session 3: CHEMOMETRICS - Chairs: D. Rutledge/ E. Vigneau</u>			
14h00	PL2	P. Bastien	Use of sparse methods in cosmetics
15h00	O06	B. Jaillais	Random forests for the prediction of water content by near-infrared hyperspectral imaging spectroscopy in biscuits
15h20	O07	C. Peltier	What is the better test to detect multivariate differences in large dimensional data?
15h40	O08	D.N. Rutledge	Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis
16h00	<i>Coffee break</i>		
16h30	O09	E. Vigneau	Analyse des relations entre plusieurs blocs de données par l'approche Path-Comdim: une application pour évaluer la qualité environnementale sur le littoral atlantique français
16h50	Poster presentations		
17h15	POSTER SESSION		
18h00	<i>Welcome Reception: Les Halles de la Major</i>		

## Thurs day 15 March

### Session 4: SENSOMETRICS - Chairs : Ph. Courcoux / P. Schlich

8h45	PL3	J. Castura	Consumer diversity in sensory evaluation data
9h30	O10	M. Brard	A latent class regression model for the clustering of multivariate binary ratings
9h50	O11	E. Qannari	One thousand and one ways to analyze free sorting data
10h10	O12	N. Pineau	CATA as an alternative method to free sorting

10h30 *Coffee break*

11h00	O13	F. Llobell	Clustatis: a cluster analysis of multiblock datasets. application to sensometrics
11h20	O14	G. Lecuelle	Modeling temporal dominance of sensations data with stochastic processes
11h40	PL4	B. Boulanger	Round table: The world beyond p-values: how to make research in the 21 <sup>st</sup> ?

12h30 *Lunch & posters*

14h30 *SOCIAL EVENT*

19h30 *Gala diner : Reverso - Les Terrasses du port*

## Friday 16 March

### Session 5: EXPERIMENTAL DESIGNS - Chairs: M. Claeys/M. Sergent

<b>9h00</b>	<b>PL5</b>	<b>J-P Gauchi</b>	<b>Metamodeling and global sensitivity analysis for computer models with correlated input</b>
<b>9h45</b>	<b>O15</b>	S. Marque	Plan d'expériences et simulations sur le contrôle qualité des contaminants microbiologiques de produits finaux
<b>10h05</b>	<b>O16</b>	Q. Carboué	Experimental design and solid state fermentation: a holistic approach to improve cultural medium for the production of fungal metabolites
<b>10h25</b>	<i>Coffee break</i>		
<b>10h55</b>	<b>O17</b>	V. Rodrigues	Food source attribution of human campylobacteriosis by meta-analysis of case-control studies
<b>11h15</b>	<b>O18</b>	U. Gonzales-Barron	An extended bigelow-type meta-regression model describing the heat resistance of neosartorya spores
<b>11h35</b>	<b>O19</b>	V. Cadavez	Dynamic determination of optimum growth rate of listeria monocytogenes in minas soft cheese during cold shelf-life
<b>11h55</b>		P. Schlich	Statistical analysis of chocolate tasting data obtained from participants
<b>12h15</b>	<i>Closing of the conference, Awards</i> <i>Lunch</i>		

## PATHOGENS-IN-FOODS: A DATABASE OF OCCURRENCE OF MICROBIOLOGICAL HAZARDS IN FOODS COMMERCIALISED IN EUROPE

Sérgio Deusdado<sup>1</sup>, Vasco Cadavez<sup>1</sup>, Vânia Rodrigues<sup>1</sup>, Pauline Kooh<sup>2</sup>, Moez Sanaa<sup>2</sup>, Ursula Gonzales-Barron<sup>1</sup>

<sup>1</sup>CIMO Mountain Research Centre, School of Agriculture (ESA), Polytechnic Institute of Braganza, Portugal

<sup>2</sup>French Agency for Food, Environmental and Occupational Health and Safety (ANSES), France

E-mail: sergiod@ipb.pt

### Abstract

The objective of this study was to build a database of the occurrence (both prevalence and counts) of the most important biological hazards in foods commercialised in Europe. For this, systematic literature searches were first conducted for every pathogen; namely, *Salmonella*, *Campylobacter*, Shiga toxin-producing *Escherichia coli*, *Listeria monocytogenes*, *Yersinia enterocolitica*, *Bacillus cereus*, *Clostridium perfringens*, *Staphylococcus aureus*, *Toxoplasma gondii*, norovirus, Hepatitis A virus, Hepatitis E virus, *Cryptosporidium* and *Giardia duodenalis*; and after screening for relevance and methodological quality assessment, data were carefully extracted from the primary studies into a harmonised arrangement consisting of primary study characteristics, food characteristics and stage within the food chain, microbiological methods, prevalence results, enumeration results and potential for bias. Based on the microbiological survey results extracted from 977 primary studies, the database *Pathogens-In-Foods* has been constructed to facilitate data access and retrieval according to hazard, food class, country or any other relevant variable, with the ability to execute simple statistical calculations.

### INTRODUCTION

In the literature, there are many investigations addressing the quantification of the occurrence of biological hazards in foods. Having access to this information has become extremely relevant in the development of pathogens' risk assessment models and risk management by both researchers and food safety authorities. Nevertheless, this information is disperse, disharmonised and not always accessible. Thus, the objective of this study was to bring together, under a harmonised arrangement, all good-quality data on the occurrence of pathogens in European foodstuffs. Thus, the work was carried out in two stages: (i) first, a systematic review and data extraction from available literature on occurrence (i.e., prevalence and concentration) of the most important biological hazards according to food matrix category (viz. meat and meat products, eggs and egg products, milk and dairy products, seafood, produce (fruits and vegetables) and composite products) commercialised in Europe; and (ii) second, the construction of a database that facilitates data access and retrieval according to bacterium, food class, country or any other relevant variable, with the ability to execute simple statistical calculations.

### METHODOLOGY

Systematic review searches using a combination of appropriate keywords were conducted using the bibliographic engines Science Direct, Pubmed, Scielo, Scopus and Web of Science using key terms in English, Spanish, French and Portuguese that involved food (meat, meat products, eggs, egg products, milk, dairy products, seafood, produce, fruits, legumes, vegetables, ready-to-eat, composite or multi-ingredient), outcome (occurrence, contamination, prevalence or concentration) and biological hazard (*Salmonella*, *Campylobacter*, Shiga toxin *Escherichia coli*, STEC, VTEC, O157:H7, O26:H11, O145:H28, O103:H2, O111:H8, O104:H4, *Listeria monocytogenes*, *Yersinia enterocolitica*, *Bacillus cereus*, *Clostridium perfringens*, *Staphylococcus aureus*, *Toxoplasma gondii*, norovirus, Hepatitis A virus, Hepatitis E virus, *Cryptosporidium* or *Giardia duodenalis*). Searches were restricted by year of publication (only primary studies published since 2000 were considered) and country (only European countries). Primary studies that passed the screening for relevance were then assessed by two researchers on their suitability for inclusion in the database, according to the following specific criteria: (i) detection and/or quantification of the biological hazard using approved microbiological methods; (ii) sufficient data on prevalence (number of contaminated samples and sample size) and concentration (mean, standard

deviation and sample size) given; and (iii) food chain stage where samples were extracted clearly specified (viz. end of production or retail). Outcomes from primary studies were marked as potentially biased if (i) microbiological methods were not well described; or (ii) the limit of detection/quantification or some measure of test accuracy was not given. The large volume of references were then documented and managed in the Zotero software with direct links to the sources in PDF. Reference databases were separately built for the microbial pathogens, parasites and viruses. All the bibliographic search process was carried out in two stages and by two different researchers, at the beginning of the project, and at the end of the data extraction process to ensure that any new reference was added. The data extracted from the primary studies were: study ID, country, year and duration of survey, pathogen, serotype/serovar/phage when applicable, packed food status (yes/no), stage in the food chain (retail, farm, mid processing, end processing, restauration), temperature at retail (chilled, ambient, frozen), food category (beverages, composite, dairy, grains, eggs, fruits, legumes, meat, seafood, vegetables, sugars and oils), a sub-hierarchy for every food category, ready-to-eat (RTE) status, sample weight, detection method, limit of detection (LoD), sample size, number of enriched positive samples, isolation (yes/no), enumeration method, limit of quantification (LoQ), number of samples above LoQ, histogram of frequencies for counts, mean microbial concentration, standard deviation, mechanism to handle substitution and potential-for-bias status (yes/no).

## RESULTS

Circa. 6000 measures of prevalence and counts were extracted from 734 primary studies reporting on microbial pathogens, 71 primary studies on parasites and 172 primary studies on viruses in foods. The database was created using MySQL [1]. MySQL is the world's most popular open source database management system, mainly because of its reliability, high availability and robustness. It completely fulfils the needs for an online relational database, available for the majority of platforms and easy to develop and maintain. In order to manage the database, phpMyAdmin [2] was chosen as it allows to administrate all the development of the project over the web, as well as supporting a wide range of operations on MySQL, namely operations to manage databases, tables, columns, relations, indexes, users, permissions, etc. Those operations can be performed via the user interface, keeping the ability to directly execute SQL statements. Moreover, phpMyAdmin is a free/publicly available platform. All the coding was developed using PHP [3]. PHP (recursive acronym for PHP: Hypertext Pre-processor) is a widely-used open source general-purpose scripting language that is especially suited for web development and can be embedded into HTML. PHP language provides vast resources to implement the functionality we need to provide a web service robust and scalable, allowing the researchers to obtain the data they need, properly filtered, specifically organized and easily accessible from the web to all hardware platforms and operative systems. We choose JOOMLA [4] as the CMS (Content Management System) to implement and maintain our web site because of its flexibility to congregate all the components of our system's architecture. Furthermore, JOOMLA allows us to perform tasks like security, user administrations and third part plugins in a transparent manner. The URL for the website is [www.esa.ipb.pt/pathogensinfood](http://www.esa.ipb.pt/pathogensinfood). The server, hosting our system (i.e., database and web site) is located in the CIESA, acronym for Informatics Centre from the Superior Agrarian School of the Polytechnic Institute of Braganza, Portugal. The web interface is still under development, and it will provide an easier and usable interface to access the pathogens data in the database, combining the data retrieval and organisation with the ability to execute simple statistical calculations that will help the researchers to enhance their productivity regarding the data analysis.

## REFERENCES

- [1] [www.mysql.com](http://www.mysql.com)
- [2] [www.phpmyadmin.net](http://www.phpmyadmin.net)
- [3] [www.php.net](http://www.php.net)
- [4] [www.joomla.org](http://www.joomla.org)