Editorial

This special issue is a collection of papers specially prepared for the 2014 ENBIS-SFdS Spring Meeting on graphical causality models, trees, Bayesian networks and big data. The objective of the meeting was to investigate new methodological advances in the area of big data analytics, causality mechanisms and robust modeling. These state of the art areas of theoretical and applied research concern a large spectrum of statisticians in various fields of specialization and application. The event was jointly organized by the European Network for Business and Industrial Statistics (ENBIS) and the Société Française de Statistique (SFdS). It took place on April 9th to 11th, 2014, at the Institut Henri Poincaré in Paris (http://springmeeting2014.sfds.asso.fr).

The French Statistical Society (SFdS) aim is to promote the use and understanding of statistics, and to foster its methodological development. To accomplish this goal, it facilitates the flow of information between statisticians working in different areas (research, teaching, administration, business, industry) and organizes meetings on a regular basis, allowing statisticians and professionals who use statistics to enrich and update their knowledge. The European Network for Business and Industrial Statistics (ENBIS) is a forum for the exchange of ideas and networking among individuals and organizations interested in theoretical developments and practical applications of Statistics. The organization, formed in 2000, is holding an annual conference and special topic spring meetings in collaboration with other organizations (for more details see www.enbis.org). Specifically, in 2007 the ENBIS-DEINDE meeting held in Turin focused on Physical versus Computer Experiments (http://web.econ.unito.it/deinde07/), in 2009 the ENBIS-EMSE meeting held in St Etienne was on the Design and Analysis of Experiments (http://www.emse.fr/enbis-emse2009/), and a second ENBIS-DEINDE meeting was held in Turin in 2011 (http://calvino.polito.it/enbis_deinde_2011/). The ENBIS-SFdS 2014 spring meeting is an outcome of many years of successful collaboration between ENBIS and SFdS (http://www.sfds.asso.fr/). Together with Quality Technology and Quantitative Management (QTQM), the ENBIS-SFdS 2014 spring meeting organizing committee has put a call for papers, a year ahead of the meeting, so that a special issue with a selection of papers to be presented at the meeting was ready for the conference itself. This special issue of QTQM has been made available to all conference participants and is accessible on the QTQM open access website (http://web.it.nctu.edu.tw/qtqm/).

The paper by Markus Kalisch and Peter Bühlmann provides a review of recent causal inference methods and covers several topics such as methods for causal structure learning from observational data or a mix of observational and interventional data, models with arbitrarily complex structures of hidden variables, approaches for estimating the interventional distribution and causal effects and a note on available software. Peter Bühlmann was one of the keynote speakers of the conference and he opened it on its first day. Two papers are focused on network inference. Gregory Nuel, Andrea Rau and Florence Jaffrèzic address the inference of causality by estimating causal effects in gene expression from a mixture of observational and intervention experiments. Inferring networks from samples obtained in different experimental conditions is the topic of the paper authored by Nathalie Villa-Vialaneix, Magali SanCristobal, Matthieu Vignes and Laurence Liaubet. The LASSO-type method proposed in this paper is based upon maximizing a penalized log likelihood, a first penalty aims at inferring a sparse solution and a second penalty is used to make the networks obtained in different conditions
consistent with a consensual network. Patrick Meyer proposes a new relevance measure that aims at improving the detection of relevant variables based on a pairwise measure called the rank minrelation/majrelation coefficient because of its connection to the rank correlation coefficient. After a formal definition of this new bivariate dependency, several examples and experiments illustrate its properties in order to select relevant variables. The paper of Giovanni Perucca and Silvia Salini shows an application of Bayesian networks to the analysis of data coming for railway users' satisfaction survey. This kind of study is important as it provides inputs to policy makers who define priorities for years to come. The Bayesian networks approach is compared with the ordered logistic analysis which is the common statistical tool applied within this context. An in depth investigation of dependencies among variables in a thermal spraying process is provided by Nikolaus Rudak, Sonja Kuhnt and Eva Riccomagno. They consider and compare different methods for the selection of Gaussian chain graph models and provide access to free software for the semi-automatic model selection of such graph models. The paper is focused on multivariate regression chain graphs for Gaussian random variables. James Cussens's paper on Integer Programming for Bayesian Network Structure Learning is addressing the problem of learning a Bayesian network using integer programming using SCIP (Solving Constraint Integer Programming). The paper shows that set packing constraints and SCIP's generic cutting planes provide a much better approximation than the cluster constraints alone which leads to the improved solving times. The paper by Luciana Dalla Valle proposes a novel approach to integrate financial information from various sources by incorporating the dependence structure among the variables. The approach is based on two types of graphical models: vines and nonparametric Bayesian belief nets (NPBBNs). Vines are undirected graphs, representing pair copula constructions, which are used to model the dependence structure of a set of variables. NPBBNs are directed graphs that use pair copulas to model the dependencies, and allow for diagnosis and prediction via conditionalization. The final paper in this special issue is by Irad Ben-Gal, Alexandra Dana, Niv Shkolnik and Gonen Singer. We list this paper at the end but Prof Ben-Gal was in fact one of the keynote speakers of the conference and opened it on its first day. The paper on Efficient Construction of Decision Trees by the Dual Information Distance Method is dealing with a fundamental task in Big Data analytics, the construction of efficient decision and classification trees. Ben-Gal et al present a novel approach for efficient construction of decision trees, the dual information distance (DID) method, which they prove to be both computationally attractive and relatively robust to noise.

In conclusion we would like to acknowledge the help of a distinguished list of reviewers who helped improve the papers by providing general and focused feedback to the authors. The reviewers of this special issue are: Anne Dutfoy, Marie-Laure Martin-Magniette, Christophe Ambroise, Fanny Villers, Philippe Leray, Antoine Chambaz, Saharon Rosset, Aviv Gruber, Shirley Coleman, David Steinberg, Rony Lidor, Julien Chiquet, Fabrice Rossi, Avner Bar-Hen, Catherine Matias, Rainer Goeb, Olivier Roustant and Eitan Greenspan. We thank them again for their work, dedication and professional expertise.

As editors, we were very pleased by the breadth and depth of applications and methods presented in these papers. We present these as a contribution to the growing role of Statistics in big data analytics, causality mechanisms and robust modeling.

Guest Editors
Jean-Michel Poggi (Past President of SFdS), E-mail: Jean-Michel.Poggi@parisdescartes.fr
Ron S. Kenett (Past President of ENBIS), E-mail: ron@kpa-group.com
Alberto Pasanisi (EDF, Electricité de France), E-mail: alberto.pasanisi@edf.fr