



International Federation of Classification Societies Newsletter

President: Maurizio Vichi

Number 49

July, 2014

Editor: Paul McNicholas

The International Federation of Classification Societies, founded in 1985, is composed of:

Associação Portuguesa de Classificação e Análise de Dados, British Classification Society, Central American and Caribbean Society of Classification and Data Analysis, The Classification Society, Gesellschaft für Klassifikation, Greek Society of Data Analysis, Irish Pattern Recognition and Classification Society, Japanese Classification Society, Korean Classification Society, Moroccan Classification Society, Société Francophone de Classification, Società Italiana di Statistica, Statistical Society of Slovenia, Vereniging voor Ordiatie en Classificatie and Section of Classification and Data Analysis of the Polish Statistical Society.

The IFCS is a non-profit, non-political scientific organization, the aims of which are to further classification research. Among other activities, the IFCS organises a biennial conference, and supports two journals: *Advances in Data Analysis and Classification* and the *Journal of Classification*.

Contents

Words from the IFCS President
News from the IFCS Secretary
News from the IFCS Publication Officer
Interviews with Presidents
 Paolo Giudici (CLADAG)
 Doug Steinley (The Classification Society)
 Jeroen Vermunt (VOC)
News from the Japanese Classification Society
News from the VOC
News from CLADAG
News from The Classification Society
News from GfKI
Upcoming Conferences and Meetings

Words from the IFCS President

I hope that you are enjoying your summer. I am delighted that this newsletter, number 49, includes interviews with presidents of member societies for the first time. I think that this is an important initiative that will be of great interest to readers. I want to thank the presidents – Paolo Giudici of CLADAG, Doug Steinley of The Classification Society, and Jeroen Vermunt of VOC – for their valuable participation, and IFCS Publication Officer, Paul McNicholas, for his work in carrying out the interviews. Some recent work by the IFCS Executive Committee is summarized in the News from the IFCS Secretary.

Maurizio Vichi
IFCS President

IFCS Homepage:

<http://www.ifcs.info>

The website contains, among other things: the IFCS Constitution and By-Laws, IFCS newsletters, and links to the websites of member societies.

News from the IFCS Secretary

The IFCS Council elected Claus Weihs (GfKI) as a new member of its Election Committee with a term of four years from 01/07/2014. The Executive Committee (EC) held a Skype meeting

on 15 May 2014. On this meeting, the EC discussed the way the decisions about award winners for the Bologna conference are made. There will be an Award Committee with four subcommittees for four award categories: Chikio Hayashi Award; Student/Postdoctoral Fellow Paper Competition and Travel Award; Student Poster Session Competition Award; and IFCS Research Medal. Candidates for the IFCS Research Medal will need to have nominators, whereas anybody can apply for the other awards.

The EC approved a Protocol Agreement about collaboration with the International Association of Statistical Computing (IASC), which is a Section of the International Statistical Institute. The Moroccan Classification Society nominated Ahmed Moussa as its representative in the IFCS Council. The Sociedad Centroamericana y del Caribe de Clasificación y Análisis de Datos re-elected Carlos Cuevas Covarrubias as president and IFCS Council representative. Please inform me of any current elections and nominations in the member societies so that I can keep the IFCS directory up to date. I wish all readers a pleasant summer!

Christian Hennig
IFCS Secretary

News from the IFCS Publication Officer

I would like to begin by thanking all of the member societies who have contributed to this newsletter. This newsletter contains the first of a series of interviews with presidents of member societies. I enjoyed the process of producing the interviews, and I trust you will enjoy reading them. Thanks to Paolo Giudici (CLADAG), Doug Steinley (The Classification Society), and Jeroen Vermunt (VOC) for their enthusiastic participation in the interview process!

Paul McNicholas
IFCS Publication Officer

Interviews with Presidents

Paul McNicholas, IFCS Publication Officer, conducted all interviews. Note that, for clarity, direct quotations from interviewees are typed in *italics*.

Interview 1: Paolo Giudici, CLADAG President

Paolo Giudici is (full) Professor of Statistics in the Department of Economics and Management at the University of Pavia in Italy. He currently teaches courses in statistics (undergraduate level), financial risk management (Master's level), and data science (PhD level). He categorizes his research interests and publications into three classes: *data analysis and data mining, credit risk measurement, and operational risk measurement*. At present, Paolo is President of the Classification and Data Analysis group (CLADAG) of the Italian Statistical Society, and President of the scientific committee of the Italian Financial Risk Association. Today, he writes from Pavia.

Paolo has interesting thoughts on current big challenges in classification and related methods. He lists these as *classification and data analysis for big data, particularly of those coming from social networks; classification and data analysis for data coming from different sources, available in different formats; positive communication and research feedback from research fields where classification and data analysis is being used (such as biomedical sciences and economical/social sciences)*. The last journal article he read appeared in *Statistics and Science: A report on the London workshop on the future of statistical sciences*. This report sees *big data as the main challenge for future statisticians* and, not surprisingly, Paolo's PhD students work in *big data analysis and classification methods for financial risk management*. His own research currently focuses on *data analysis for classification and prediction of financial risks, using both conventional data (financial markets data, balance sheet data, macroeconomic variables) and big data (analysts' opinions, social network posts and tweets)*.

I have had the pleasure of meeting Paolo at conferences on several occasions, including at the CLADAG meeting that was held at his historic university in Pavia in 2011. This was my first time in Pavia and I can vividly remember the historic beauty of the university. I wonder whether the appreciation of such beautiful surroundings wanes over time? *My appreciation of the University of Pavia has increased over time, as I become involved in more networking*

activities, facilitated by the surrounding environment. Paolo regularly attends conferences, and he is looking forward to three major conferences in Italy in the next year or so. In December, he will attend the Computational and Financial Econometrics meeting in Pisa. Next year, he is looking forward to IFCS 2015 in Bologna in July and the CLADAG meeting in Cagliari in October. For the latter meeting, Paolo is President of the Scientific Committee and he adds that *I hope our readers will join!*

I asked Paolo to describe the research funding environment for classification researchers in Italy. He answered that it is *very difficult*, explaining that *funding has shrunk, both from the public and the private sector*. He added that *recruiting of new (young) researchers is almost impossible*; however, *on a positive note, there is increasing cross-border collaboration, for example at the European level*. In a difficult funding environment, societies can be an important source of help and encouragement for early career researchers. Paolo tells me that CLADAG provides them a place for networking and envisaging research directions, particularly through the conferences we directly organize or the ones we actively support. His advice for early career researchers is to *consider research projects that involve data analysis and classification in a highly interdisciplinary setting, involving, besides statistics, knowledge of the applied field where the methods are really applied*.

In terms of the IFCS, Paolo sees its main role as *to coordinate and promote developments in classification and data analysis*. He suggests that the IFCS should strive to *give visibility to new fields of research, such as those mentioned above, with specialized sessions in organized conferences or in promoting existing conferences*. Outside of work, Paolo likes to partake in sports like running or cycling, especially with my family members.

A final word from Paolo. *Please consider coming to our next Cladag conference in Cagliari, in October 2015. It will be located in a beautiful sea resort and will focus on data analysis methods for large and big data, with special reference to socio-economic applications*.

Interview 2: Doug Steinley, President of The Classification Society

Doug Steinley is (full) Professor of Psychological Sciences at the University of Missouri. At present, he is President of The Classification Society and an Associate Editor of the Journal of Classification. Today, he writes from Ann Arbor, Michigan, where he is *teaching a 5-day workshop on finding patterns in data for the Interuniversity Consortium for Political and Social Research at the University of Michigan*.

We begin by chatting about the IFCS and the field of classification in general. Doug sees *the role of the IFCS as the facilitator of interaction between the member organization, promoting collaboration and mutual sharing of knowledge to broadly promote the primary goals of classification and related methods*. What current challenges within classification are particularly tough? *One of the bigger challenges is related to the enormous quantity of techniques that are continually being published that propose “new” and “improved” methods for classification and clustering*. Generally, there should be a *standardized approach to conceptualize the integration of this plethora of innovation*. *I have been fortunate enough to be involved in the Benchmarking Initiative being spearheaded by Iven Van Mechelen (at the behest of the IFCS) as an attempt to begin to get a handle on some of these issues*. Doug also has interesting ideas to strengthen collaboration within the IFCS. *Eventually, I would like to see an initiative that looked to enhance collaboration across disciplines in addition to within the existing member organizations*. This problem is related to the above problem of the vast quantity of classification and related methods. *A common problem is to have numerous disciplines invent and re-invent the same methods*. Part of the difficulty is *there exists no common literature for the variety of disciplines that employ classification methods*. *The more often a common thread can be created as a “connector” between disciplines and societies, the quicker progress on outstanding problems can be achieved*.

The conversation turns to research funding and PhD students. Where do classification researchers in the U.S. apply for funding? *Most funding for classification researchers, and other*

methodologists, falls under the guise of the National Science Foundation and the National Institute of Health. What kind of work is funded? While it is possible to get funded for “pure” methodological development, it is much, much easier to get funded if the methodological work is being done in support of a more substantive question. For instance, much of my own funding has been for developing new/different approaches for classification and clustering in the context of alcohol abuse and dependence. Doug is currently advising three PhD students and each is working on slightly different areas. One student is working on developing methods for predicting missing links in the context of social network data. My second student is evaluating the performance of fit indices in mixture modeling and exploring the relationship between fit indices and cluster recovery in mixture modeling. My newest student is working on a project that is using k-median clustering as an internal mechanism inside of sliced inverse regression as a robust method for choosing the slices. Does any of this tie-in with the aforementioned work on alcohol abuse and dependence? Yes. Currently, we are comparing the performance of sliced inverse regression as a competitor to traditional methods in psychology (traditionally, factor analysis) of data reduction in predictive modeling contexts. Additionally, for the mixture modeling component, we are finding that traditional data structures in the alcohol abuse and dependence literature are highly skewed; consequently, it is beginning to appear as though mixture models with skewed distributions (such as the many worked on by yourself) are more appropriate than the standard multivariate normal distribution.

I have had the pleasure of meeting Doug at a few conferences, most recently at the IFCS meeting in Tilburg last July. He tells me that the next conference he will attend is the meeting of the Society for Multivariate Experimental Psychology in Nashville, Tennessee. I asked Doug what advice he would give to early career researchers. He emphasized that it is important to remain resilient in the face of the adversity that is often encountered while beginning a research program. This often takes the form of rejected journal submissions and the difficulty of obtaining funding for research. Never giving up on yourself or your ideas is an important part of becoming successful. Success is just as much about perseverance and tenacity as it is

intelligence. Good advice indeed, and applicable to us all, I think. What does The Classification Society do for early career researchers? The Classification Society sponsors sessions for new students. Generally, there are some travel awards given for outstanding dissertation research to be presented at the annual conference.

This year is the 50th anniversary of The Classification Society. Are there any plans to mark this anniversary? Yes, the 50th anniversary will be celebrated next year at McMaster University in Hamilton, Ontario, Canada.

Interview 3: Jeroen Vermunt, VOC President

Jeroen Vermunt is (full) Professor of Methodology and Statistics and Head of the Department of Methodology and Statistics at Tilburg University in The Netherlands. At present, Jeroen is President of the Dutch and Flemish Classification Society (VOC). Today, he writes from Tilburg, the venue of the very successful IFCS 2013 meeting. More specifically, he tells me: *I am sitting behind my university desktop computer at this moment. It is beginning of August and extremely quiet at the university.*

Perhaps unnecessarily, I asked Jeroen to provide a short bio. *After finalizing my Master's in 1988 at Tilburg University, I worked for a few years at Statistics Netherlands in De Hague. In 1992, I returned to Tilburg University, where I obtained my PhD in 1996. Since then, I stayed at this university, where I became a full professor in 2003. Currently, I am the head of the Department of Methodology and Statistics. Next, I asked about his current research, which focuses on mixture and latent class models. With my group, I work on tools that make the correct application of these methods easier for applied researchers. These are among other tools for assessing model fit and checking model assumptions in simple and complex mixture models, tools for relating cluster membership to external variables in a stepwise manner, power computation methods for mixture models, and methods for applying mixture models in a way similar to hierarchical clustering techniques. This sort of research sounds perfect for PhD students. I am an advisor of about 10 PhD students. Most of them work on the kind of topics I mentioned when answering the previous question. I also encourage them to work on*

“modern” topics connected to increased computational power and availability of big amounts of data, such as specific types of Bayesian calculations for model fit assessment or methods for model parameter estimation with streaming data.

What about the current major challenges in classification and related methods? In my field of social and behavioral sciences, applied researchers typically do their statistical analysis on their own, that is, without the help of statisticians. At the same time, the available data sets and the statistical methods they use are becoming more complex. For example, it is getting more and more common for social and behavioral scientists to use techniques such as latent class and mixture models. To make sure that this is done in the right way, good statistical training is needed as well as good software. What I see as the main challenges in classification and clustering is the development of tools that make it easier to use the currently available advanced methods in the right way by non-experts, and to implement these tools in easy to use software.

Where does the IFCS fit in? In my opinion, the main role of IFCS is to serve as a scientific community for statisticians working on classification and clustering themes but coming from rather diverse applied fields. To me, IFCS is a very useful complement to mono-disciplinary methods and statistics societies from my own applied field, such as the Psychometric Society and the European Association of Methodology. Jeroen also has interesting ideas to strengthen collaboration within the IFCS. Collaboration between the European societies within IFCS could and should be strengthened. This can be achieved via joint expert meetings dealing with well-defined topics. Another option is to collaborate in the context of our PhD training, for example, by jointly organizing courses on advanced topics in classification and clustering and by stimulating students to make more use of the existing international exchange programs.

How does a classification researcher in The Netherlands get funding? There is no specific funding for classification researchers in the Netherlands. We get our funding via applications at the social and behavioral science section of the Dutch Science Foundation. Crucial in these applications is that we demonstrate the relevance

of the proposed classification research for social and behavioral sciences, which in general is not very difficult. The last five years our group obtained several millions of Euros for research on classification related topics. It must be a good sign that Jeroen thinks that the choice for a PhD in a classification related topic is a very good one. There is still a lot of scientific work to be done in this field. For some students, the PhD training will be the start of an academic career, for others it will be an excellent starting point for a research job at a profit or non-profit research institute. It is clear to me that there is huge demand for experts in classification. Societies can be a valuable resource for PhD students and early career researchers, and the VOC encourages PhD students to present their work at our conferences. This year we introduced a best presentation award for PhD students. Membership of our society is free for PhD students.

Finding time to read journal articles can be very difficult. Most of what I read is related to articles that I review. When reviewing, I often go back to the original sources that the paper refers to. Apart from this reviewing, the last paper I read was the article “Auxiliary Variables in Mixture Modeling: Three-Step Approaches Using Mplus” by Tihomir Asparouhov and Bengt Muthén which appeared in the last issue of the journal Structural Equation Modeling. This article caught my attention because it elaborates on methods I proposed in 2010, and which our group extended recently in various ways. It is nice to see how others implement these approaches in their mixture modeling software. Finding time for hobbies is also difficult and Jeroen’s main hobby is cycling. Though Netherlands is flat, I like most cycling in the mountains.

I thoroughly enjoyed the IFCS meeting in Tilburg last year, and Jeroen tells me that he will be at IFCS 2015 in Bologna. I asked Jeroen whether he has any advice for organizers of future IFCS meetings. The Bologna statistics department that organizes IFCS 2015 is a very strong group with lots of experience in organizing these kinds of meetings. So, I don’t think they need much advice from me. What we focused on Tilburg is on having interesting invited speakers, tutorials, and organized sessions, and moreover on making the conference attractive and affordable for PhD

students. A final word from Jeroen. *I appreciate enormously the efforts of the IFCS board to increase the number and diversify of member societies. Recently, this resulted in the admission of the Moroccan Classification Society. In this way, IFCS becomes more and more International.*

News from JCS

(I) The Japanese Classification Society Symposium on New Developments in Extended Principal Component Analysis

The symposium will be held on September 2014 at the University of Tokyo, Hongo Campus in Tokyo under the 2014 Japanese Joint Statistical Meeting. The organizer is Kouhei Adachi (Osaka University). The official Web site of the 2014 Japanese Joint Statistical Meeting is <http://www.jfssa.jp/taikai/2014/>

(II) Information about the 33rd Annual Meeting of the Japanese Classification Society

The 33rd Annual Meeting will be held on March 2nd and 3rd, 2015 at Teikyo University, Kasumigaseki Campus in Tokyo. The chairperson of the organizing committee is Satoru YOKOYAMA (Teikyo University). For additional information, please visit our website at <http://www.bunrui.jp/>.

Satoru Yokoyama

News from the VOC

The Dutch/Flemish classification society (VOC) celebrates its 25th anniversary with a *Jubilee Meeting* on November 6-7, 2014 in the Abbey Hotel Rolduc at Kerkrade (The Netherlands). We are happy and proud to announce the following speakers: Denny Borsboom (University of Amsterdam), Lieven De Lathauwer (KU Leuven), Paul Eilers (Erasmus MC Rotterdam), Jelle Goeman (Radboud University Nijmegen), Patrick Groenen (Erasmus University Rotterdam), Willem Heiser (Leiden University), Christian Hennig (University College London), Jeroen Jansen (Radboud University Nijmegen), Daniel Oberski (Tilburg University), Marco Riani (University of Parma), Marieke Timmerman (Groningen University), and, the winner of the VOC student presentation award at the third VOC conference, Lianne Ippel (Tilburg University).

We would like to invite you for this Jubilee meeting. The meeting is open to both VOC-members and non-members. For more information on the program and registration see the VOC website: <http://www.voc.ac>. We hope to meet you in Kerkrade.

On the 9th of May, 2014, the VOC had its *third conference* with a key note contribution from Arthur Tenenhaus (Supélec, France) on ‘Regularized Generalized Canonical Correlation Analysis for Multiblock or Multigroup Analysis’ and many excellent contributions on a range of topics (including latent class analysis, PLS path models, and Texture Analysis and Multivariate Classification models in CT Images from dogs with Pulmonary Thrombosis, to name a few). At this occasion, a best student presentation award was given to Lianne Ippel (Tilburg University) who presented a paper ‘On the Implementation and Empirical Evaluation of Streaming Shrinkage Factors with Applications in Predicting Non Response’.

Katrijn Van Deun

News from CLADAG

The next CLADAG meeting will be held at the Flamingo Hotel in Santa Margherita di Pula, Cagliari, on October 8-10, 2015. The meeting will take place under the auspices of the IFCS and of the Italian Statistical Society (SIS). The chosen location will facilitate exchange of ideas and networking, as well as an easier access for young researchers. The conference themes include, but are not limited to, the following:

A - Classification Theory

Bayesian Classification – Biplots – Clustering models – Consensus of Classifications – Correspondence Analysis – Discrimination and Classification – Factor Analysis and Dimension Reduction Methods - Fuzzy Methods – Genetic Algorithms – Hierarchical Classification – Multidimensional Scaling – Multiway Scaling – Multiway Methods – Neural Networks for Classification – Non Hierarchical Classification– Similarities and Dissimilarities –Software algorithms for classification – Unfolding and Related Scaling Methods

B- Multivariate Data Analysis

Bayesian Data Analysis – Big Data Analysis – Categorical Data Analysis – Covariance Structure Analysis – Data Mining – Data Science – Data Visualization – Decision Trees – Functional data analysis – Mixture and Latent Class Models – Multilevel data Analysis – Non Linear Data Analysis – Nonparametric and Semiparametric Regression – Partial Least Squares – Pattern recognition – Robustness and Data Diagnostics – Social networks – Software algorithms for multivariate analysis – Spatial Data Analysis – Symbolic Data Analysis.

More details on the participation at the conference will be given in the next issue.

Paolo Guidici

News from The Classification Society

The Classification Society is pleased to announce the 2015 annual meeting will be held in Hamilton, Ontario next June. Further, we are particularly excited to note that there will be a special session to celebrate a half-century of the Classification society. Specific dates and additional details will follow in the next issue of the newsletter.

Additionally, after 12 years of exceptional service, Willem Heiser will be stepping down as editor of the Journal of Classification. A committee has been formed to recommend a new editor to the society's board of directors.

Doug Steinley

News from GfKI

Brief report on 50th Anniversary Celebration Meeting Classification Society, Wednesday, July 9, 2014

The meeting took place at FirstSite, Colchester hosted by British Classification Society (BCS) and Department of Mathematical Sciences, University of Essex in cooperation with: Big Data and Science Week hosted by Faculty of Science & Health, University of Essex; Gesellschaft für Klassifikation (GfKI); and European Conference on Data Analysis (ECDA2014).

Scientific programme:

David Hand, Imperial College London, UK: "*Classifying the world: how classification methods have revolutionised understanding*"

Fionn Murtagh, De Montfort University, Leicester, UK: "*Data science in psychoanalysis: a short review of Matte Blanco's bi-logic, based on metric space and ultrametric or hierarchical topology*"

Maurizio Vichi, University of Rome, Italy: "*Advances in multimode-clustering, clustering and dimensionality reduction*"

Hans Kestler, University of Ulm, Germany: "*Boolean networks for modeling biological signaling processes*"

Peter Flach, University of Bristol, UK: "*Recent advances in threshold selection and classifier calibration*"

David Wishart, Hans-Hermann Bock, David Hand, Maurizio Vichi, Peter Flach, Sabine Krolak-Schwerdt, and Berthold Lausen: "*50th Anniversary resume, outlook and discussion: Classification society and data science*"

Berthold Lausen

Upcoming Conferences and Meetings

August 2014

Irish Machine Vision and Image Processing (IMVIP) Conference, August 27-29, Derry-Londonderry, Northern Ireland: sites.google.com/site/imvip2014/

September 2014

Second International Workshop on Model-Based Clustering and Classification (MBC²), September 3-5, Catania, Italy: www.dei.unict.it/mbc2

November 2014

VOC 25th Anniversary Jubilee Meeting, November 6-7, Abbey Hotel Rolduc, Kerkrade, The Netherlands: <http://www.voc.ac>

July 2015

IFCS-2015, July 6-8, Bologna, Italy.

March 2015

33rd Annual Meeting of the Japanese Classification Society, March 2-3, Teikyo University, Kasumigaseki Campus, Tokyo, Japan: <http://www.bunrui.jp/>

September 2015

European Conference on Data Analysis
(ECDA2015), September 2-4, Colchester UK.

October 2015

CLADAG 2015, October 8-10, Flamingo Hotel,
Santa Margherita di Pula, Cagliari, Italy.

**Looking Ahead to IFCS Newsletter
Number 50**

Contributions for the forthcoming issue of the
IFCS newsletter should be sent to

mcnicholas@math.mcmaster.ca

before November 21, 2014.

Plain text or Word files are preferred.

Thanks!

Paul McNicholas

IFCS Publication Officer