



International Federation of Classification Societies Newsletter

President: Maurizio Vichi

Number 52

October, 2015

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, Multivariate Data Analysis Group of the South African Statistical Association, Multivariate Statistics and Classification Group of the Spanish Society of Statistics and Operations Research, Società Italiana di Statistica, Statistical Society of Slovenia, Vereniging voor Ordinatie 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*.

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.

Contents

Words from the IFCS President
News from the IFCS Secretary
News from the IFCS Publication Officer
Report on IFCS 2015
IFCS Award Winners
Call for Submissions: BRCA Data Analysis Award
Call for Nominations: CS Distinguished Dissertation Award
News from CLAD
News from GfKI
News from the JCS
News from MDAG
News from the VOC
News from The Classification Society
News from the Greek Society of Data Analysis
News from the SFC
News from SKAD
IFCS Cluster Benchmark Data Repository: Call for Contributions
Upcoming Conferences and Meetings

Words from the IFCS President

The IFCS 2015 conference was held July 6-8 in Bologna. It was a clear success: 200 papers presented, 220 participants, 5 plenary talks with excellent international scientists, the initial session celebrating the 30th anniversary of the foundation of the Federation, 6 awards assigned to outstanding researchers, including the first IFCS research medal. I wish to thank the SPC and the LOC for this successful conference; in particular, the presidents Francesco Palumbo and Angela Montanari for the excellent program and the wonderful hospitality in the beautiful city of Bologna. I also wish to thank the IFCS Awards Committee and, in particular, the President Salvatore Ingrassia, for the intense and excellent work done in assigning the awards. The next IFCS Conference will be organized 8-10 August,

in Tokyo. The Chair of the SPC is Tadashi Imaizumi and the Vice-Chair of the SPC is Sadaaki Miyamoto. Let me take this opportunity to thank the Japanese Classification Society, and the President Tadashi Imaizumi, for their willingness to organize the next IFCS conference and for their excellent work in preparing the bid together with IFCS President-Elect Akinori Okada.

This year, the IFCS celebrated the 30th anniversary of its foundation. The Federation was established 4th July 1985 by 6 associations. After 20 years, IFCS doubled to 12 societies in 2005. In 2006, IFCS has admitted the 13th member and, in 2009, the 14th. By the end of 2013, the IFCS had reached 15 members with the Moroccan Classification Society. In the last two years, two new associations became members of IFCS: the *Multivariate Statistics and Classification Group of the Spanish Society of Statistics and Operation Research* SEIO-AMyC (<http://amyc.seio.es/>) and the *Multivariate Data Analysis Group of the South African Statistical Association* (SASA-MDAG) (<http://www.sastat.org.za/mdag-0>). Now, the *Classification and Multivariate Statistics Group (CMSG-HSA) of the Hungarian Statistical Association* (HSA) has just applied to become 18th member of the Federation. Finally, a protocol of collaboration was recently signed between IFCS and IASC. Therefore, after 30 years, the Federation has more than tripled its membership and I feel that this is a testament to the good health of IFCS.

I am also proud to say that, after 30 years, the scope and purpose of the Federation – to promote mutual communication, co-operation and interchange of views among all those interested in scientific principles, numerical methods, and practice of data analysis and classification in a broad sense and in as wide a range of applications as possible – is still very innovative. The growing needs of data, expressed by the new social and institutional attitudes of the Information Society as well as the development of global processes with almost infinite sources of data production, have generated the need for new methodologies to deal with the so called ‘data deluge’. Phenomena are now described, in a richness of detail, by means of multivariate and complex information that vary over time and space. The data collected on these phenomena need new and advanced methodologies to give realistic and

appropriate results to users. The IFCS has always been a melting pot of ideas and new methodologies for exploring, modelling and analysing data and for extracting the relevant knowledge from them. The Federation, in Kobe in 1996 and, again, Rome in 1998, was the first scientific society to use the now trendy term Data Science as a title of a conference. This qualifies the pioneering research of IFCS on data analysis presented already at that time. The IFCS community of scientists has a multidisciplinary connotation and uses techniques, theories, skills and abilities that are specific to different disciplines such as, statistics, probability, mathematics, computer science, operations research and applied engineering, just to cite a few. This mixing of different scientific attitudes creates the richness of ideas and competences that are now most useful to give appropriate responses to the needs of the Information Society. In this year of celebration, I would also like to stress the high quality of papers presented at the IFCS conferences and published in *Journal of Classification*, *Advances in Data Analysis and Classification* and *Studies in Classification, Data Analysis, and Knowledge Organization*. Thus, my final wish for all the IFCS community is simply to continue to explore innovative methodologies and applications as it has done very successfully in the last 30 years.

Dear readers, this is my final contribution to the Newsletter as President of the Federation because I am at the end of my term (2014-2015). Let me say, without rhetoric, that I was honoured to serve as IFCS President. It was a real pleasure because my research and scientific activities were always close to the IFCS. I wish to take this occasion, once again to warmly thank the Executive Committee: Iven van Mechelen, Akinori Okada, Christian Hennig, Paul McNicholas, and Nema Dean who have worked hard during this period. I also wish to thank the generous collaborations from many volunteers within the IFCS community that certainly improved the activities of the Federation. After 30 years, IFCS still has a number of relevant challenging issues ahead and I am sure the Federation will continue to solve them in excellent ways under the capable leadership of the incoming President, Akinori Okada, to whom I wish all the best success.

Maurizio Vichi
IFCS President

News from the IFCS Secretary

The most important event after the last IFCS Newsletter was the very successful and enjoyable IFCS Conference in Bologna with almost 200 presentations and more than 200 participants. In Bologna, an IFCS Council meeting took place, attended by 28 representatives, representing 15 societies including all the new IFCS member societies MCS, SASA-MDAG and AMyC-SEIO. Core items were the report on the conference in Bologna and the presentation of the IFCS conference in Tokyo 2017. The IFCS now welcomes bids from member societies for hosting the IFCS conference 2019. It is advisable to bid a long time before the conference, and societies who consider bidding should get in contact with the IFCS Executive Committee, the President or the President-Elect. Furthermore it was proposed and decided that there should from now on be only two IFCS Newsletters per year again. The Council decided to explore whether information about IFCS representatives and member society presidents could be put on the IFCS website in such a way that the societies can edit this information. However, in consultation with the IFCS website administrator it emerged that this would be a lot of effort and raise security concerns, so it seems better to only put the information on the website, but to leave the editing to the IFCS Secretary, as before. A further news item is that the IFCS has received an application for membership from the Classification and Multivariate Statistics Group of the Hungarian Statistical Society, on which there will soon be an IFCS Council ballot.

Since the last IFCS Newsletter, the IFCS Executive Committee has met three times, on 18 June (via Skype), 5 July (in Bologna) and 24 September (via Skype). Much time on the first two meetings was spent with preparations for the conference in Bologna and the Council meeting. The election of a new IFCS President-Elect was prepared, and there was a discussion on the IFCS contribution to conferences with which the IFCS has an agreement. The next such conference that comes up is the COMPSTAT conference in Oviedo, 23-26 August 2016. As IFCS Secretary, I will be in the Scientific Programme Committee and there will be one or two IFCS invited sessions. The IFCS will also contribute to the 61st World Statistics Conference in Marrakech, 2017. Please continue to inform me of any current

elections and nominations in the member societies so that I can keep the IFCS directory up to date. My best wishes go to all readers for the time until the next newsletter.

Christian Hennig
IFCS Secretary

News from the IFCS Publication Officer

Thanks to everyone who contributed to this newsletter – there have been many contributions this time and I hope this enthusiasm will continue into future newsletters! As Maurizio Vichi's term as IFCS President comes to an end, I want to thank him for the wonderful work he has done on behalf of the Federation; I thoroughly enjoyed working with him. I am very much looking forward to working with Akinori Okada as his term as IFCS President begins shortly.

Paul McNicholas
IFCS Publication Officer

Report on IFCS 2015

The last IFCS conference was hosted by the Dipartimento di Statistica Paolo Fortunati at the Alma Mater Studiorum in Bologna, it was held from July 5th to July 8th. Angela Montanari chaired the Local Organizing Committee and Francesco Palumbo (University of Naples Federico II) chaired the Scientific Program Committee. Both of them have served the Italian Classification and Data Analysis society as President. A very special plenary session was organized to celebrate the IFCS 30th anniversary. The session saw the participation of two eminent IFCS past Presidents: H. H. Bock and professor J. Gower. It was an excellent opportunity to appreciate the long road made by the Federation during these 30 years.

This Conference saw over 220 participants from across the country and from around the world. As usual the Conference program was organized over plenary invited sessions, specialized sessions and contributed paper sessions. Trevor Hastie (USA), Marco Riani (Italy), Gerda Claeskens (Belgium) and Terry Speed (Australia) were invited to give a plenary talk; one more plenary session was the Presidential Address of IFCS

from Maurizio Vichi (Italy). The whole number of contributions, including 11 posters, was 200. The effort of our colleagues at the University of Bologna made possible this conference such a success. The high quality of the contributions was assured by the work of all members of the Scientific Program Committee. Our deepest gratitude goes foremost to all of them. A special thank must be addressed to Laura Anderlucci who has supported the participants via the conference website.

A post-proceedings volume will be edited by Angela Montanari, Francesco Palumbo, and Maurizio Vichi and will be published by Springer in the prestigious series on *Studies in Classification, Data Analysis and Knowledge Organization* (ISSN: 1431-8814). Authors have been invited to submit their contributions via the conference website <http://ifcs.baku.ac.at/ifcs2015>. The deadline for submitting the 10/12 pages full paper is December 18th.

Francesco Palumbo

IFCS Award Winners

The winners of the IFCS awards, presented at IFCS 2015, are:

- IFCS Research Medal: Hans-Hermann Bock.
- Chikio Hayashi Award: Ryan Browne.
- Student / Post-Doctoral Fellow Paper: Silke Janitza and Brian Franczak
- Poster Prize: Francesca Fortunato.

I would like to underline that this has been the first time for many of these prizes and this attracted the interest of many researchers.

Salvatore Ingrassia

Call for Submissions: BRCA Data Analysis Award

Under the auspices of:

- Data Science Society GfKI
- IFCS
- International Biometric Society – Deutsche Region
- Deutsches Konsortium Familiärer Brust- und Eierstockkrebs

- HeritX: Sparking Solutions for Inherited Cancer

The *BRCA Data Analysis Award* will be conferred for original and innovative work that uses data-analytic and statistical methods in order to support the development of strategies for the prevention of Breast Cancer (BRCA) related inherited cancer. Data analysis techniques, combined with computational algorithms, are broadly and routinely applied in many fields like medicine, bioinformatics, economics, marketing, pattern recognition, sociology, etc. Typical approaches include statistical modeling, classification and clustering, factorial analysis, exploratory tools, and visualization. Representative scientific societies are, e.g., the International Federation of Classification Societies, the German Classification Society, the Biometric Society, and others.

This award concentrates on the domain of inherited cancer research. In fact, inherited cancer is affecting millions of people worldwide. It devastates entire families, from generation to generation. The most prominent type is caused by a genetic change (mutation) in the so-called BRCA gene (with subtypes BRCA1, BRCA2 etc.). A carrier of this gene defect will, not only and with a large probability, develop breast, ovarian or other cancers, while fifty percent of her children will inherit the defective gene and the implied cancer risk as well. In contrast to other cancer types, inherited cancer can and should be intelligently prevented already before it starts rather than being intensively treated after it has occurred. This seems to be achievable because the first step in the development of BRCA-related cancer, the inherited gene defect, is known from the outset. However, there is a lack of research and innovation in transforming this knowledge into a practical treatment and strategy. To change this situation, new approaches are necessary and might be supported by data-analytic methods. This goal underlies the *BRCA Data Analysis Award* and also is followed by various medical, scientific, and commercial institutions, including, e.g., the global nonprofit research initiative HeritX in USA.

The BRCA Data Analysis Award has been introduced to recognize a compelling investigation that is related to inherited BRCA genes and may benefit the early diagnosis and/or medical prevention of BRCA-related cancer by

using data-analytic methods. The award includes a honorarium of 1000 € (for personal use) and the presentation (in person) of the author's work at one of the major conferences on data analysis and its applications. The award not only honors relevant research in the BRCA field, but also is intended to motivate young researchers across the data analysis community to address problems in this field. Qualified research needs to be completed in 2014-2016, and its author must be ≤ 40 years. Among the list of submitting candidates and works the winner will be selected through an expert committee across BRCA research and data analysis, orchestrated by Hans-Hermann Bock.

Interested researchers are invited to submit a nomination package including:

1. a brief summary (two pages or less) of the work or investigation on which the candidacy is based, which explains its relevance to a) the field of BRCA, b) the data analytic method proposed, and c) its potential benefit for the early diagnosis and/or medical prevention of BRCA-related cancer,
 2. up to three published papers, an accepted thesis, or an accepted manuscript,
 3. a curriculum vitae, and
 4. the names and addresses of two referees,
- to BRCA-Award@gfkl.de, until September 30, 2016. Candidates will be notified by the end of 2016. For further information see the website www.isw.rwth-aachen.de/BRCA-Award.php.

Hans-Hermann Bock

Call for Nominations: The Classification Society Distinguished Dissertation Award

Each year, the Classification Society offers a Distinguished Dissertation Award (the Award, hereafter) for an outstanding Ph.D. dissertation on the theme of clustering, classification, or related areas of data analysis, encompassing associated theory and/or applications. The 2016 Award totals US\$500 in book vouchers from Chapman and Hall/CRC and the winner will be invited to make a presentation in a special session at the 2016 Classification Society Annual Meeting, which will be held at the University of Missouri from June 1-4. The Classification Society will cover travel expenses to the 2016 Annual

Meeting of at most US\$1,000 for a North American-based Award winner or at most US\$1,500 for an Award winner based elsewhere. Criteria for nomination for the Award include: (i) the dissertation contains innovative or impressive work in theory/methodology, or innovative or well-developed application(s); and (ii) the literature review is thorough. To be eligible for nomination, a Ph.D. dissertation must have been successfully completed – including successful defense – within the calendar year 2015. Nominations may be made by the dissertation author, advisor/supervisor, or related person. A nomination must include contact information for the nominator and nominee (if different), a URL where an online copy of the dissertation can be accessed by the members of the Award Committee, and an electronic (e.g., scanned) cover letter, written and signed by the nominator, stating the date of completion of the Ph.D. dissertation and outlining why the dissertation merits the award. The nominator must also arrange for two external referee reports to be sent directly to the chair of the Award committee. If possible, the referees should be arms length from the dissertation author and their supervisor.

Cover letters should include full bibliographic details for any published papers, proceedings, book chapters, or similar scholarly material that has arisen from the dissertation. For non-English language dissertations, nominators are requested to provide an extended abstract in English and, where relevant, to highlight any papers, proceedings, book chapters, or similar scholarly material associated with the thesis that has been published in English.

The 2016 Award Committee comprises:

- Kathleen Gates (The University of North Carolina at Chapel Hill, United States),
- Herbie Lee (University of California, Santa Cruz, United States), and
- Paul McNicholas (McMaster University, Canada).

Nominations and external referee reports must be submitted to the chair of the Award Committee, Paul McNicholas, at paulmc@at.mcmaster.ca, on or before **February 5, 2016**. Results will be announced approximately two months after this submission deadline.

Paul McNicholas

News from CLAD

JOCLAD 2015

The XXII Annual Meeting of the Associação Portuguesa de Classificação e Análise de Dados (CLAD), **JOCLAD 2015**, was held April 10-11 2015 at the *Escola Superior de Tecnologia do Barreiro – Instituto Politécnico de Setúbal*. There were about 65 participants and 50 papers were presented. The conference program included 3 plenary invited lectures, 4 thematic sessions, 8 parallel sessions and a poster session. Two short courses took place April 10. Further information is available on the conference website: <http://joclاد2015.estbarreiro.ips.pt/>. The authors of the best applied research papers presented at the conference were invited to submit full papers for publication in the next issue of CLAD Books Series. The publication will contain a cluster of refereed articles which were presented in latest JOCLAD meetings. Editors are H-Bacelar-Nicolau, Fernanda Sousa, Carlos Marcelo and Otilia Dias. Publisher is the *Instituto Nacional de Estatística*, INE.

CLAD New Cooperation Protocol

A Cooperation Protocol of associate membership between CLAD and APD, the *Associação Portuguesa de Demografia*, was signed during the XXII Annual Meeting by the APD president, Maria Filomena Mendes, and the CLAD president José Gonçalves Dias.

CLAD Sponsored Courses

Following the series of CLAD sponsored courses, four short courses took place on April, May and on October:

- Two half-day courses were held April 10 in the scope of JOCLAD2015, hosted by the *Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal*: “*Model-based Clustering - presentation of the Mixmod software*” (Gilles Celeux, INRIA Saclay-Île-de-France, France) and “*Análise de Dados nas Ciências da Saúde*” (Cristina Rocha, Faculdade de Ciências da Universidade de Lisboa, Portugal)

- A one day course was held May 30 under the topic “*Modelos com Equações Estruturais - aplicações com LISREL e AMOS*” (Maria de Fátima Salgueiro, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal). The course was hosted by the *Centro de Sondagens e Estudos Estatísticos, Instituto Politécnico de Tomar*.

- A one day course on “*Como criar um package na linguagem R*” (Pedro Duarte Silva, Universidade Católica, Porto, Portugal) was held October 24, at the *INE - Lisboa*. The course organization was associated with the second World Statistics Day, WSD2015, which was observed on 20 October 2015, under the theme 'Better Data, Better Lives'.

JOCLAD 2016

The XXIII Annual Meeting will be held from March 31 to April 2, 2016, at the *Universidade de Évora*, the second oldest university in the country. Évora is the capital of the Alentejo region, a UNESCO World Heritage Site and a member of the Most Ancient European Towns Network. More information on JOCLAD 2016 may be found at <http://www.clad.pt/>.

Helena Bacelar-Nicolau

News from Gesellschaft für Klassifikation (GfKI) Data Science Society

The *European Conference on Data Analysis 2015 (ECDA2015)*, co-hosted by the *British Classification Society (BCS)*, *Gesellschaft für Klassifikation e.V. (GfKI)* and *Sekcja Klasyfikacji i Analizy Danych (SKAD)* and held at the University of Essex, Colchester, UK from 1st to 4th September 2015 (conference webpage www.ecda2015.eu). After the successful ECDA conferences in Luxembourg (2013) and Bremen (2014), participants came from 19 different countries: Austria, Belgium, Canada, Cyprus, Germany, Finland, Greece, Italy, Luxembourg, Morocco, Mexico, New Zealand, Poland, Portugal, South Africa, Tunisia, United Kingdom, United States of America and Switzerland. The conference included the *39th Annual Meeting of the German Classification Society* and the *Librarians and Information Science (LIS) Workshop*. Keynotes of the conference were

- Hendrik Blockeel, Leuven: *Declarative data analysis*
- Ulf Brefeld, Darmstadt: *Capturing user behaviour*
- Andrzej Dudek, Wrocław: *Cluster analysis in the XXI century, new algorithms and tendencies*

- Arthur Gretton, London: *Kernel nonparametric tests of homogeneity, independence and multi-variable interaction*
- Janette McQuillan, London: *Crowdsourcing classifications to accelerate cancer research*
- Christine Müller, Dortmund: *Data depth*
- Iris Pigeot, Bremen: *Challenges in the statistical analysis of longitudinal data*
- Claus Weihs, Dortmund: *Efficient global optimization: motivation, variation, and application*
- Stefan Wrobel, Bonn: *Data analytics in a networked world*

The conference had invited and contributed sessions under the headings *Data Analysis, Data Science, Clustering, Classification, Machine Learning and Knowledge Discovery, Library and Information Science, Finance and Economics, Digital Humanities and Social Sciences, Geosciences and Archaeology, Marketing, Musicology, Machine Learning and Knowledge Discovery, Outliers in Classification, Life Sciences, Mathematical Foundations of Data Science, Education, Big Data, Engineering, Logistics and Optimisation.*

The GfKl Best Paper Award 2014 was awarded for the paper at ECDA2014 *Incremental Generalized Canonical Correlation Analysis* by Angelos Markos (Democritus University of Thrace) and Alfonso Iodice D'Enza (University of Cassino and Southern Lazio). Professor Hans Kestler were the laudator and Professor Angelo Markos gave a short presentation on the awarded paper in opening session of the conference. Sponsors of the ECDA2015 were the data science company *Profusion Ltd, London* and the *Institute for Analytics and Data Science, University of Essex*. The conference dinner was at Colchester Castle, the biggest Keep ever built in the United Kingdom and the largest that remains in existence throughout Europe.

At the general annual meeting the members of GfKl decided to add *Data Science Society* to the name of the Society. The members awarded Professor Gunter Ritter, Passau, the honorary membership of GfKl. As honorary member Gunter Ritter is a member of the GfKl advisory board together with the honorary president Hans Bock and the honorary members Wolfgang Gaul, Otto Opitz and Claus Weihs. The other members of the advisory board were elected - as usual - by

the board of the society for a term of three years (2013-16): Christian Hennig (London), Lars Schmidt-Thieme (Hildesheim), Frank Scholze (Karlsruhe), Myra Spiliopoulou (Magdeburg) and Alfred Ultsch (Marburg).

GfKl Data Science Society will celebrate its *40th Annual Meeting* at the Fourth Joint Statistical Meeting of the Deutsche Arbeitsgemeinschaft Statistik "Statistics under one Umbrella" 14 - 18 March 2016 at the Georg-August University Göttingen (DAGStat2016) in Göttingen, DAGStat 2016 includes the 40th Annual Meeting of GfKl, 62. Kolloquium of the German Region of the International Biometric Society and the Pfingsttagung of the German Statistical Society.

In September 2016, we plan to have the next joint workshop with the Japanese Classification Society at Schloss Reisenburg in Germany and in August 2017 before or after the IFCS2017 conference in Japan. The next German-Polish meeting is planned in September 2017 before the *ECDA2017 in Wroclaw*. In spring 2017 we will celebrate 40 years *Gesellschaft für Klassifikation Data Science Society* (founded February 1977 in Frankfurt am Main).

Berthold Lausen

News from JCS

(I) New Executive Members of JCS

The following list includes the new staff of the Japanese Classification Society elected according to the articles of the JCS governing the election of Executive Members. The tenure for these members is April 1, 2015 to March 31, 2017.

PRESIDENT:

Tadashi Imaizumi, Tama University,
imaizumi@tama.ac.jp

SECRETARIAT MEMBERS:

Secretary: Atsuhiko Nakayama, Tokyo Metropolitan University. **General affairs:** Takafumi Kubota, Tama University. **Public relations:** Satoru Yokoyama, Teikyo University. **Treasurer:** Yasunari Kono, Rikkyo University. Koji Kurihara, Okayama University; Sadaaki Miyamoto, University of Tsukuba; Masahiro Mizuta, Hokkaido University; Akinori Okada, Tama University; Fumitake Sakaori, Chuo

University; Yuki Toyoda, Hosei University; Hiroyuki Tsurumi, Yokohama National University; Hiroshi Yadohisa, Doshisha University; Yoshiro Yamamoto, Tokai University.

AUDITORS:

Akinobu Takeuchi, Jissen Women's University; Atsuhiko Hayashi, Kyushu University.

EXECUTIVE COMMITTEE MEMBERS:

Noboru Ohsumi, Tatsuo Otsu, Yutaka Kano, Koji Kurihara, Fumitake Sakaori, Mika Sato-Ilic, Kazuo Shigemasu, Yuki Toyoda, Atsuhiko Nakayama, Yasumasa Baba, Fumi Hayashi, Masahiro Mizuta, Takashi Murakami, Hiroshi Yadohisa, Kazue Yamaoka, Kazunori Yamaguchi, Yoshiro Yamamoto, Satoru Yokoyama, Ryozo Yoshino, Michiko Watanabe.

(II) The Japanese Classification Society Symposium on “Correspondence Analysis, Dual Scaling, and Quantification Theory, the present and the future”.

The symposium was held on September 9th 2015 at Okayama University, Tsushima Campus in Okayama Pref. under the 2015 Japanese Joint Statistical Meeting. The organizer, discussant, speakers were as follows:

- Organizer: Tadashi Imaizumi, Tama University.
- Discussant: Kohei Adachi, Osaka University.

Speaker and co-speaker(s): Tadashi Imaizumi, Tama University; Takashi Murakami, Chukyo University; Yoshiro Yamamoto and Sanetoshi Yamada, Tokai University; Fumi Hayashi, Toyo Eiwa University; and Kazue Yamaoka, Teikyo University. The official Web site of the 2015 Japanese Joint Statistical Meeting is <http://www.jfssa.jp/taikai/2015/>.

(III) Information about the 34th Annual Meeting of the Japanese Classification Society

The 34th Annual Meeting will be held on February 29th and March 1st, 2016 at Tokai University, Takanawa Campus in Tokyo. The chairperson of the organizing committee is Toshiro YAMAMOTO (Tokai University). For additional information, please visit our website at <http://www.bunrui.jp/>.

Satoru Yokoyama

News from MDAG

The annual meeting of the Multivariate Data Analysis Group (MDAG) will take place during the South African Statistical Association (SASA) Conference from 29 November - 2 December 2015. The MDAG members will also be addressed by Professor Maurizio Vichi during a special interest session on Multivariate Data Analysis. Since the founding of the group in 2013, the number of active members has increased to a network of 54 multivariate data enthusiasts.

MDAG is also looking forward to hosting the upcoming Correspondence Analysis and Related Methods (CARME) conference in 2019 surrounded by the picturesque mountains of Cape Town, South Africa. This will be the first CARME conference not held on European soil.

Johané Nienkemper-Swanepoel

News from the VOC

On the 29th of May, 2015, the VOC had its fourth conference in Nijmegen (The Netherlands) in collaboration with the Analytical Chemistry-Chemometrics Research Group of the Radboud University Nijmegen. A key note lecture was given by Morten A. Rasmussen (University of Copenhagen, Denmark) on ‘Notes on some ideas from the Netherlands – seen from a Danish perspective’. Furthermore, there were many excellent contributions on a range of topics (including chemometrics, optimal treatment regimes, clustering, factor analysis, and latent class analysis). At this occasion, a best student presentation award was given to Mattis van den Bergh (Tilburg University) who presented a paper on ‘Divisive latent class analysis applied to social capital’.

Katrijn Van Deun

News from The Classification Society

CS 2015

The 2015 meeting of The Classification Society took place from June 3-6 at McMaster University, Hamilton, Ontario, Canada. This was the 44th meeting of The Classification Society, and just

the 6th time the meeting was held in Canada. Around 50 delegates attended the meeting, coming from eight countries: The Netherlands, Italy, France, UK, Belgium, Germany, US (six states), and Canada (three provinces). In all, there were five keynote lectures:

- President's Invited Address: Jeroen Vermunt, Tilburg University, The Netherlands.
- President's Address: Douglas Steinley, University of Missouri, United States.
- Invited Lecture on Big Data: Fionn Murtagh, Royal Holloway, University of London, UK.
- Invited Lecture on Classification Methodology: Vincent Vandewalle, University Lille 2, France.
- Invited Lecture on Psychometrics: Katie Gates, University of North Carolina at Chapel Hill, United States.

There were also three invited sessions, three contributed sessions, and a poster session. One of the invited sessions was on 50 Years of the Classification Society, and it sparked some very interesting historical discussion! The Classification Society Distinguished Dissertation Awards for 2013 and 2014 were presented at the banquet, which was held at the Royal Botanical Gardens. The winners are:

- Distinguished Dissertation Award 2014: Irene Vrbik.
- Distinguished Dissertation Award 2013: Kim De Roover.

CS 2016

The 2016 meeting of The Classification Society will take place from June 1-4 at the University of Missouri, Columbia, MO. Details will soon be available at www.classification-society.org.

Paul McNicholas

News from the Greek Society of Data Analysis

The 8th Panhellenic Conference of Data Analysis, Alexandroupolis, Greece, from September 10th to 12th, 2015 in Democritus University of Thrace. The conference was organized by Achilles Dramalidis and Angelos Markos (Democritus University of Thrace). The meeting was a very successful and fruitful event, where experts from academia and industry shared interdisciplinary research results and experiences.

This edition attracted more than 60 participants from different countries such as Greece, France, Italy, Germany, Spain and Nigeria.

Angelos Markos

News from the SFC

1. SFC Meeting (2015)

The 22th meeting of the SFC was held in Nantes (France) on 9-11 September 2015 (<http://sfc2015.sciencesconf.org/>). It was organized by the University of Nantes. About 70 researchers participated to the meeting and 31 papers were presented. Invited sessions were organized with the following speakers: Gilbert Saporta, Michel Grossetti, Jean Loup Guillaume. The Price Simon Regnier 2015 was awarded this year to François Queyroi for his work titled: Clustering large graphs: Measures, Algorithms and Visualization (Partitionnement de grands graphes : Mesures, Algorithmes et Visualisation).

2. AAFD & SFC Meeting (2016)

The next (23st) meeting of the French speaking Classification Society will held in Marrakach (Morocco) on 22-26 May, 2016. It will be colocated with 7th edition of the biennial symposium "Artificial Learning and Data Mining" (AAFD). It will be organized by the Moroccan Society of Classification (SMC), Moroccan Society of Applied Mathematics (SM2A), AAFD (LIPN, University of Paris 13) and SFC.

Mustapha Lebbah

News from the SKAD

The 24th Scientific Conference of Classification and Data Analysis Section (SKAD) of Polish Statistical Society (the 29th Taxonomic Conference) on "Classification and Data Analysis – Theory and Applications" was held on 14th-16th September in 2015 in Novotel Gdańsk Marina Hotel in Gdańsk. The conference was organized by SKAD and Department of Statistics of University of Gdańsk. The conference topics included:

a) theory (taxonomy, discriminant analysis, linear ordering methods, multivariate statistical analysis methods, continuous variables analysis methods,

discrete variables analysis methods, symbolic data analysis methods, graphical methods), b) applications (data analysis in finance, marketing, spatial data analysis, others applications of data analysis – medicine, psychology, archaeology, etc., computer applications of statistical methods).

There were 81 conference participants. During 2 plenary sessions and 13 parallel sessions the participants have presented 58 papers. Additionally, a poster session was organized, during which 14 posters were shown. The conference participants have prepared articles which will be reviewed and then published in *Taxonomy Series* no. 26 and 27 (as part of *Research Papers of Wrocław University of Economics*) and *Statistics in Transition New Series* (an international journal published jointly by the Polish Statistical Association and the Central Statistical Office of Poland).

Krzysztof Najman

IFCS Cluster Benchmark Data Repository: Call for Contributions

The IFCS Task Force on Benchmarking invites all researchers interested in clustering and classification to contribute data sets to the IFCS Cluster Benchmark Data Repository (<http://ifcs.boku.ac.at/repository/>). With this repository the Task Force aims to create a unique resource of data sets (along with comprehensive documentation on their background and on goals for subjecting them to cluster analyses) that can be used to compare the performance of clustering methods and associated data-analytic procedures.

Background

At present, very often new clustering methods, algorithms, and data-analytic procedures (e.g., procedures to determine the number of clusters in a data set) are being proposed without adequate comparisons with current best practices, and without sufficient evidence of good performance on benchmarking datasets. In particular, new methods and procedures are often proposed along with results from a few simulations and analyses of a few real datasets chosen in an unsystematic manner. Sometimes such simulations and analyses may involve comparisons with one or two alternative methods, but without discussion

of the specific conditions under which the new method or procedure is supposed to work well, what kind of clusters it attempts to find, and what would therefore be the most appropriate competitors. Otherwise, comparative work often implicitly assumes that there should be a unique best method for general clustering, and then suggests, based on a handful of experiments, that the newly proposed method is a promising candidate for this position, without giving any idea about potential weaknesses and limitations.

In general, it seems clear that benchmarking methods for cluster analysis (unsupervised classification) and related data-analytic methods is nontrivial. Particularly it is essentially more difficult than benchmarking in supervised classification. The reason is that, typically in supervised classification, there is a single true grouping, and performance can be measured by looking at (potentially weighted) misclassification percentages. In cluster analysis, however, different aims of clustering may lead to different clusterings on the same dataset that could be optimal according to different criteria (e.g., overall low within-cluster distances, or optimal representation of every object by the centroid object of the cluster to which it is assigned, or optimal fit by a mixture probability model). Otherwise, this also implies that it is particularly important for benchmarking clustering to define properly the clustering problem that a method aims to solve, by specifying as precisely as possible what kinds of clusters are of interest. With this specification in place, appropriate simulation setups, datasets for benchmarking, and competing existing clustering methods can be more systematically chosen.

That being said, datasets with known classes are often used for benchmarking clustering in a similar manner as in supervised classification. Sometimes these are simulated from mixture probability models with clusters defined by homogeneous mixture components, and sometimes these are real datasets in which the problem of interest was originally supervised classification. But in all of these cases, it would be conceivable that clusterings other than those assumed to be “true” could be valid as well. In fact, clusterings different from the “known” ones could sometimes be of even greater scientific interest, because cluster analysis is used in reality to find clusterings that are not yet known. For this

reason, it is also of interest to use datasets for benchmarking that represent genuine clustering problems, i.e., in which the “true” clustering is not yet known.

Aim and Setup

The aim of the IFCS Cluster Benchmark Data Repository is to stimulate better practice in benchmarking (performance comparison of methods) for cluster analysis by providing a variety of well-documented, high-quality datasets and simulation routines¹ for use in practical benchmarking. The overarching principle is that cluster benchmarking does not simply mean that methods are ranked in a one-dimensional manner, but rather that their advantages and shortcomings are characterized thoroughly against a more precise definition of what kind of clustering could be of interest, and the characteristics of the datasets to which the methods would be applied.

The IFCS Cluster Benchmark Data Repository (<http://ifcs.boku.ac.at/repository/>) offers a platform for all kinds of datasets on which clustering may be of interest, including standard multivariate data, data with mixed types of variables, similarity and dissimilarity data, and various kinds of structured data such as time series data, regression data, spatial data, symbolic data, graph data. The datasets may be offered with or without given “true” clusterings. A particular feature of the repository is that every dataset comes with comprehensive documentation, including information on the specific nature of the clustering problem in the dataset, and the characteristics that useful clusters should fulfill, with scientific justification. In cases where “true” clusters are provided, such information is still required in order to understand properly what it means that a certain clustering method performs well or badly, taking into account the possibility that clusterings other than the given one can also be good in some sense. In cases where “true” clusters are not provided, the given information can be used to evaluate the performance of clustering methods by other means than looking at misclassifications, for example based on quality criteria that can be computed from the clustered data, or based on other external variables or information if provided.

¹ The section of the repository that will contain simulation routines is still under construction.

Call for Contributions

The IFCS Task Force on Benchmarking (Anne-Laure Boulesteix, Rainer Dangl, Nema Dean, Isabelle Guyon, Christian Hennig, Fritz Leisch, Doug Steinley, and Iven Van Mechelen) are most happy to invite now all researchers interested in clustering and classification to contribute data sets to the IFCS Cluster Benchmark Data Repository (at <http://ifcs.boku.ac.at/repository/>). If you click on the link “Contribute a Data Set” on the website, you will immediately be guided to instructions and a questionnaire to contribute data along with (required and optional) metadata information. On the website you may also inspect a few examples of data sets that are already available. As soon as the repository includes a sufficient number of data sets, the Task Force will undertake several initiatives to make it highly visible to the international scientific community as a point of reference for benchmarking in cluster analysis and as a catalyst for better practices in this area.

Iven Van Mechelen

Upcoming Conferences and Meetings

February 2016

JCS 34th Annual Meeting, February 29 - March 1, Tokai University, Takanawa Campus. <http://www.bunrui.jp/>

March 2016

GfKI Data Science Society 40th Annual Meeting, March 14-18, at the Fourth Joint Statistical Meeting of the Deutsche Arbeitsgemeinschaft Statistik “Statistics under one Umbrella”, Georg-August University Göttingen (DAGStat2016). <http://www.uni-goettingen.de/de/485701.html>

JOCLAD 2016, March 31 - April 2, *Universidade de Évora*. <http://www.clad.pt/>

June 2016

The Classification Society meeting, June 1-4 at the University of Missouri, Columbia, MO. www.classification-society.org

Looking Ahead to IFCS Newsletter Number 53

Contributions for the forthcoming issue of the IFCS newsletter should be sent to Paul McNicholas, IFCS Publication Officer, at mcnicholas@math.mcmaster.ca by March 24.

Plain text or Word files are preferred.

Thanks!

Paul McNicholas

IFCS Publication Officer