



Vereniging voor Ordinatie en Classificatie / Dutch-Flemish Classification Society

Chairman:	Jeroen Vermunt, Universiteit van Tilburg, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken van Onderzoek, Postbus 90153, 5000 LE Tilburg, Nederland (j.k.vermunt@uvt.nl)
Secretary:	Katrijn Van Deun, Universiteit van Tilburg, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken van Onderzoek, Postbus 90153, 5000 LE Tilburg, Nederland (K.VanDeun@uvt.nl)
Treasurer:	Tom Wilderjans, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, Postbus 9555, 2300 RB Leiden, Nederland (t.f.wilderjans@fsw.leidenuniv.nl)
Editor:	Ralph Rippe, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Pedagogische Wetenschappen, Algemene en Gezinspedagogiek, Postbus 9555, 2300 RB Leiden, Nederland (rrippe@fsw.leidenuniv.nl)

VOC website: <http://www.voc.ac>

Postbankrekening (IBAN) NL86 INGB 0000 161723 t.n.v. Vereniging voor Ordinatie en Classificatie.

5th VOC Conference

May 27th, 2016 **Tilburg University**

Dante Building, Warandelaan 2, 5037 AB Tilburg

10.15-10:45	VOC Annual Members Meeting
10.45-11:15	Welcome and registration
11.15-12:45	Submitted paper session 1
12.45-13:30	Lunch
13.30-14:20	Keynote address: Eric Postma, <i>The return of Neural Networks</i>
14.20-15:20	Submitted paper session 2
15.20-15:40	Coffee break
15:40-17:00	Submitted paper session 3
17.00-17:10	Announcement of the PhD Presenter Award Winner
17:10-??:??	Drinks

In this issue:

Short program 5 th VOC Conference (Tilburg, May 27)	1
Registration details for the 5 th VOC Conference	1
From the President	2
Publications	2
Program 5 th VOC Conference (Tilburg, May 27)	3
Annual Report of the Secretary for the year 2015	4
Minutes of the Members Meeting 2015 (Nijmegen)	4
Agenda Annual Members Meeting 2016 (Tilburg)	5
Financial Report for 2015	6
Route description	7
Meetings	9
Book of Abstract 5th VOC Conference	11

Registration details for the 5th VOC Conference

Those who would like to participate in the 5th VOC Conference are welcome and are kindly requested to register through our website <http://www.voc.ac> (go to 'meeting'). Details are provided through the website.

From the President

Dear VOC members,

Our annual meeting will be held May 27 at Tilburg University. We again have a very interesting program with submitted paper presentations of both junior and senior VOC members. As should be the case at VOC, these presentations deal with a broad range of statistical methods and application types. In addition, we have been able to attract an excellent keynote speaker. Our keynote speaker is Eric Postma, a professor in Artificial Intelligence and one of leading persons in the Tilburg Data Science initiative. He will talk about exciting recent developments in the (old) field of neural networks.

The co-organization of the meeting in Tilburg will be my last task as a VOC board member. My second term as president ended, and it is time for another person to take it over. I enjoyed this job very much, especially because of the excellent collaboration with my fellow board members. It is also satisfying to see that with the new style annual meetings that we started a few years ago, VOC has been able to attract the new generation of classification scientists. Two other highlights I would like to mention are the hosting of the IFCS meeting in 2013 and the great 25 years VOC jubilee meeting in 2014.

I am happy to be able to announce that the VOC board has been able to find an excellent candidate for the vacant president position. At our annual meeting, the VOC board will nominate Mark de Rooij for this position. Mark is very motivated to make the VOC even more successful.

I would like to thank you all very much for the confidence you have placed in me as your president. And, of course, I hope to see many of you at our annual meeting in Tilburg.

Jeroen Vermunt

Publications

Will presented again in the VOC Fall 2016 Newsletter.

Program 5th VOC Conference (Tilburg, May 27th, 2016)

10.15		VOC Annual Members Meeting
10:45		Welcome and registration
11.15-12.45		<i>Submitted paper session 1</i>
11.25	Paul Eilers	Functional data analysis of juggling trajectories
11.45	Frank de Vos	Alzheimer's disease classification using resting state fMRI*
12.05	Andrei Barcaru	Towards Bayesian analysis of GCxGC and GCxGC-MS data*
12.25	Zhengguo Gu	A variable selection method for simultaneous component* based data integration
12.45		Lunch
13.30-14.20		<i>Keynote address</i>
13.30	Eric Postma	The return of Neural Networks
14.20-15.20		<i>Submitted paper session 2</i>
14.20	Norman Verhelst	Profile Analysis: A complementary way of reporting results in large assessments
14.40	Lieke Voncken	Continuous norming of psychological tests: A comparison of stepwise model selection procedures*
15.00	Yu Han	Understanding social-class differences in the transition to adulthood using Markov chain models*
15:20		Coffee and tea break
15.40-17.00		<i>Submitted paper session 3</i>
15.40	Jonas Haslbeck	Structure estimation for time-varying mixed graphical models in high-dimensional data*
16.00	Giulio Flore	A comparative analysis of psychometric unfolding models*
16.20	Rita Folcarelli	A novel multivariate method for selection of immunological cells with multicolour flow cytometry*
16.40	Katrijn Van Deun	Big Data in the Social Sciences: Statistical methods for multi-source high-dimensional data
17.00		Announcement of the PhD Presentation Award Winner
17:10		Closing and drinks

(*) applicant for the VOC Presenter Award

Annual Report of the Secretary for the year 2015

1. Number of members

The VOC started end 2014 with 120 members and counted 118 members at the end of 2015. Two memberships were terminated, nine members could not be contacted (inactive e-mail address), and there were nine new members registered. In 2015, 26 members paid contribution.

2. Board

The Board of the VOC was composed as follows in 2015:

Jeroen Vermunt - President
 Katrijn Van Deun - Secretary
 Tom Wilderjans - Treasurer
 Ralph Rippe - Newsletter Editor +Webmaster
 Jeroen Jansen - Member

The Board met once in 2015. The main topic was the organization of the 5th VOC conference.

3. Activities

The main activity of the VOC was the fourth VOC conference.

The fourth VOC Conference took place at Radboud University Nijmegen (The Netherlands) on the 29th of May 2015 with a full day program, including nine contributions by VOC members. On the 29th of May, 2015, the VOC had its *fourth conference* with a key note contribution from Morten A. Rasmussen (University of Copenhagen, Denmark) on 'Notes on some ideas from the Netherlands - seen from a Danish perspective' and many excellent contributions on a range of topics (including 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'. The conference had approximately 25 participants.

4. Publicity

The newsletter appeared once. The VOC conference was also announced to non-VOC members, using the IFCS newsletter, the VVS-site and mailing lists (IOPS, BBC, ICS).

Minutes of the VOC Annual Members Meeting (May 29th 2015, Nijmegen)

1. Opening of the meeting

The president, Jeroen Vermunt, opens the meeting.

2. Minutes of the previous meeting

The minutes of the previous meeting were approved.

3. Annual Report

- Contributions by members

There are 120 possible members (after updating the file with members). However, there were only 43 paying members. This is because we have many PhD-members which do not have to pay.

-Miscellaneous

It is possible to do VOC payments through an app for mobile phones. Also online payments can be done.

4. Financial Report

There is a decline in the financial reserve of the VOC, mainly because of the backlog in the contributions and the small number of participants to the Jubilee Meeting in Rolduc (not all costs were covered). This was the first Jubilee Meeting that implied a negative balance for the VOC. Ideally there is about 4000-5000 Euro on the VOC account.

The new treasurer and secretary will have to work together to update the information on the contribution by the members and to keep this information smoothly updated. An overview of which members paid their contribution and which ones did not is needed for the last three years so that the backlog in contributions can be eliminated.

At this moment, Berrie and Jeroen are the only ones that can do payments for the VOC (Tom, however, will replace Berrie soon).

5. Cash Committee

Elise Dusseldorp has controlled the accounts together with Lianne Ippel: The accounts are in order. The cash committee commented that it was very difficult to control the accounts because they had to search through many document. The committee proposes to use a

standardized form to declare costs (see attachment). The cash committee also advised that when a payment is made, always the name of the person and the reason for the payment are included as information. Because this was not always present, the cash committee had a hard time to control the accounts.

The audit committee remains on duty for the next year.

6. Composition of the board

Tom Wilderjans takes note of the fact that Ralph Rippe replaces Tom Wilderjans as Newsletter Editor of the VOC for a term of three years; the President (Jeroen Vermunt) takes note of the fact that Tom Wilderjans becomes the new treasurer of the VOC for the next three years. The board is still looking for new board members. Jeroen Jansen has been elected as new board member.

7. Miscellaneous

Elise Dusseldorp will organize a VOC-session on the next IFCS conference (Bologna 2015) on the topic "new developments in classification and regression trees".

8. Closure of the meeting

Jeroen Vermunt closes the meeting.

Agenda for the VOC Annual Members Meeting (May 29, 2015, Nijmegen)

1. Opening of the Members Meeting

2. Minutes of the Members Meeting 29th May 2015

The Minutes of this Meeting are included in this Newsletter (see p. 4-5).

3. Annual Report of the Secretary on the year 2015

The Annual Report is included in this Newsletter (see p. 4).

4. Financial report of the treasurer on the year 2014

The Financial Report is included in this Newsletter (p. 6).

Report of the cash committee (Elise Dusseldorp and Lianne Ippel).

Appointment of a new cash committee.

5. Composition of the Board

The board is composed of the following members (with their remaining term, in years, between brackets):

- Jeroen Vermunt, President (0)
- Katrijn Van Deun, Secretary (1)
- Tom Wilderjans, Treasurer (2)
- Ralph Rippe, Newsletter Editor (2)
- Jeroen Jansen, board member (2)

The term of Jeroen Vermunt ends. Candidates for this position can register up to 24 hours before the meeting with the secretary.

The board is looking for additional members. Candidates may contact the President before the meeting. The members may give their opinion during the meeting.

6. Miscellaneous

-Raise member contribution to 25 Euros?

-Timing of collecting the contributions: ask a first time in January, a general reminder in March, and if needed a reminder in May

-Communication channel between VOC members?

7. Questions before closure of the meeting

8. Closure of the Members Meeting

Financial Report for 2015

Revenue		Expenditure	
membership fees (26 paying members)	520	transaction costs ING	101,8
overdue membership fees	25	hosting website	121
Revenue spring meeting Nijmegen	250	IFCS due 2013-2014	155
subsidy spring meeting	1000	IFCS due 2015-2016	185
interest savings account	23,53	Rent location and catering spring meeting	809,74
		Costs guest speaker spring meeting 2015	340,97
		meeting VOC board	36,6
		Costs guest speaker spring meeting 2014	89
		ADAC subscription 2015	44,34
		ADAC subscription 2014	41,34
Total	1818,53	Total	1924,79

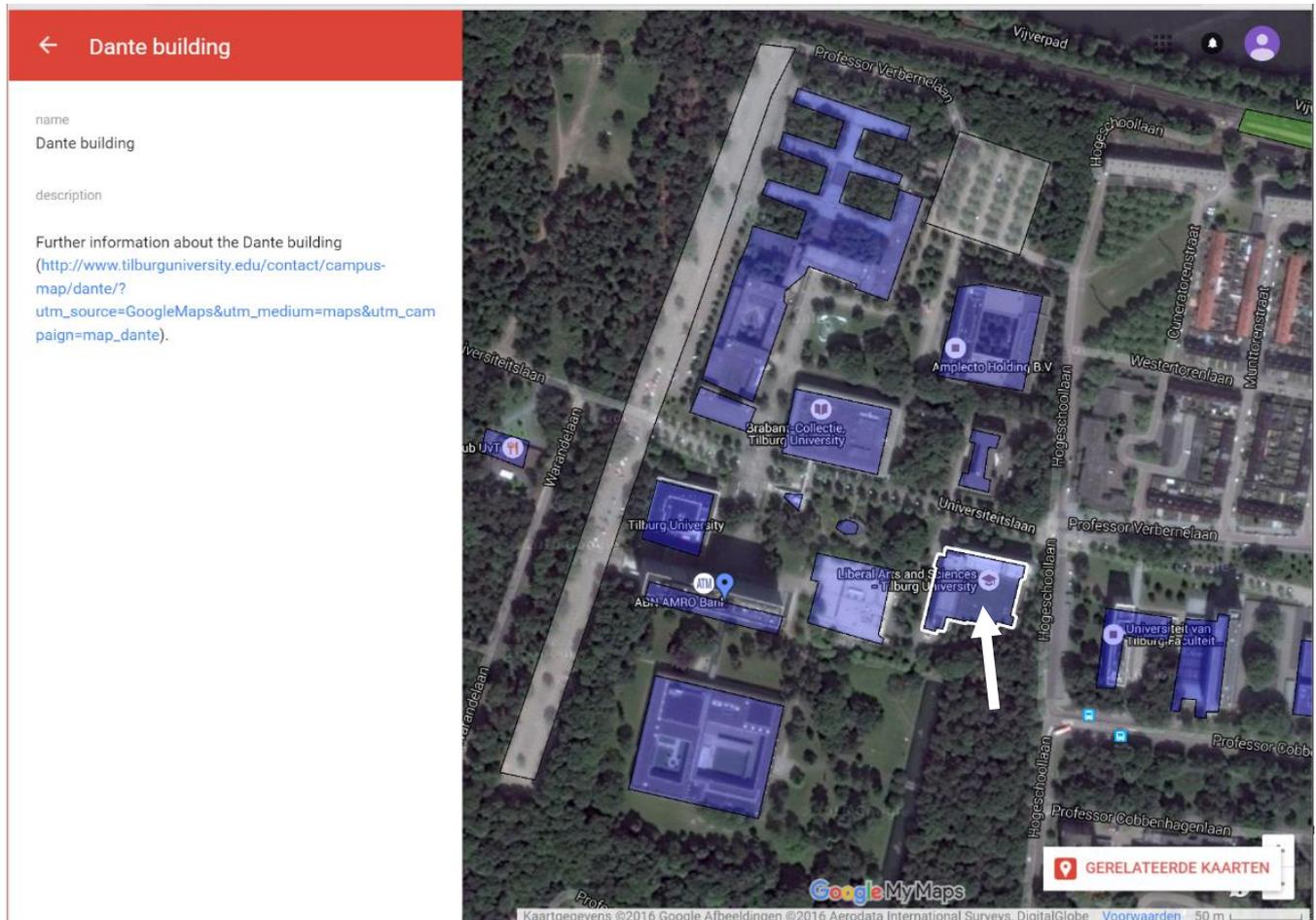
Debet		Credit	
Balance ING account	1072,76	Accounts payable	0
Balance savings account	2840,9	Equity	3913,66
Total	3913,66	Total	3913,66

An overview of the evolution of the equity

2015	€ 3913,66
2014	€ 4019,92
2013	€ 5444,46
2012	€ 5524,70
2011	€ 6194
2010	€ 7621
2009	€ 8189
2008	€ 6248
2007	€ 5914
2006	€ 6869
2005	€ 6057
2004	€ 5019
2003	€ 6795
2002	€ 6408
2001	€ 5898
2000	€ 5731
1999	€ 4871
1998	€ 5100

Route description

The 5th VOC Conference will take place at the **Dante Building**, Room DZ-4, (Warandelaan 2, 5037 AB Tilburg) Tilburg University, The Netherlands



By public transport

Visitors travelling by stop-train (from Eindhoven, Breda and 's-Hertogenbosch) can take the 'Tilburg Universiteit' NS train station which is just a few minutes walking from Tilburg University.

Travellers on fast intercity trains need to get off the train at 'Tilburg Centraal Station' (downtown) and can reach Tilburg University from there by cab approximately € 15, bus (bus line 4), or a connecting stop-train to 'Tilburg Universiteit'.

By Car

From Schiphol Airport, Amsterdam (A4.A2)

Take the A4, direction Utrecht. The A4 will ultimately connect to the A2, which will take you all the way to 's-Hertogenbosch. Just beyond 's-Hertogenbosch take the A-58/E-312 to Breda which will take you to Tilburg. Take exit 11 (Goirle/Tilburg/Turnhout). At the traffic lights, take a left turn to Tilburg. At the next traffic lights, take a left turn to "Universiteit". Follow the "Universiteit" signs all the way to the University. There is car park at the edge of the campus.

From Breda, Eindhoven, and 's-Hertogenbosch (A58/E312)

Take exit number 11 (Centrum West). At the traffic lights, follow direction Tilburg. Keep left. At the next traffic lights, take a left turn to "Universiteit". Follow the "Universiteit" sign all the way to the University. There's a parking lot right at the entrance of the campus.

From Waalwijk, Kaatsheuvel and Loon op Zand

As you approach Tilburg from the North, follow the signs to "Centrum". At the roundabout, go straight ahead and follow the signs "Universiteit" from there on.

From Brussels Airport

Take the E17 motorway to Antwerp. In Antwerp, take the motorway to Breda. From Breda, take the A58 to Tilburg, exit 11 (Centrum West). At the traffic lights, take a right turn to Tilburg. At the next traffic lights, take a left turn to "Universiteit". Follow the "Universiteit" sign all the way to the University. There's a parking lot right at the entrance of the campus.

From Eindhoven Airport

Take the A2 motorway Eindhoven to Tilburg, and take the exit 11 (Centrum West). At the traffic lights, take a left turn to Tilburg. At the next traffic lights, take a left turn to "Universiteit". Follow the "Universiteit" sign all the way to the University. There's a parking lot right at the entrance of the campus.

Meetings

ICSSM 2016: The *9th International Conference on Social Science Methodology* is organized by RC33 (Research Committee for Logic and Methodology in Sociology), Leicester (UK), 11-16 September 2016.

Close of call for papers /abstracts submission 21 January 2016.

<http://www2.le.ac.uk/departments/sociology/research/rc33-conference>

IBS 2016: XXVIIIth *International Biometric Conference* 10-15 July, 2016, Victoria Convention Centre, Victoria, Canada. <http://biometricconference.org>

ISCB 2016: The 37th *Annual Conference of the International Society for Clinical Biostatistics* will be held in Birmingham, UK from 21-25 August, 2016.

<http://www.birmingham.ac.uk/facilities/mds-cpd/conferences/iscb-2016/index.aspx>

IFCS 2017: The meeting of the *International Federation of Classification Societies 2017* will take place in Tokyo, Japan, from August 8-10, 2017.

IMPS 2016: The *International Meeting of the Psychometric Society 2016* will be held at the Renaissance Hotel in Asheville, NC, USA.

Pre-conference workshops will be held on July 11, and the main conference will run from July 12 through July 15.

<https://www.psychometricsociety.org/content/imps-2016>

IWSM 2016: The 31th edition of the *International Workshop on Statistical Modeling* will be held in Rennes (France) from 4 to 8 July 2016, hosted by the Institut National des Sciences Appliquées.

<http://www.lebesgue.fr/content/sem2016-iwsm2016>

CoDaWork 2017: IASC promotes the 7th *International Workshop on Compositional Data analysis (CoDaWork 2017)* that will be held in Val d'Orcia (Tuscany, Italy) from June 5-9. CoDaWork 2017 offers a forum of discussion for people concerned with the statistical treatment and modelling of compositional data or other constrained data sets, and the interpretation of models or applications involving them. The primary goal of the workshop is to identify important potential lines of future research and gain insight as to how they might be tackled. All necessary details concerning registration and abstract submission are available at the conference website

<http://www.compositionaldata.com//material/others/Antonellacodawork2017.pdf>

CoDaCourse 2016: The CoDaCourse 2016 will be held in Girona, Italy from 4-8 July, 2016. It provides an introduction to the theoretical and practical aspects of statistical analysis of compositional data, as well as an informal discussion forum on more advanced modeling topics. This course will present the current state of the art in this field of active research and will cover the following topics:

- Hypothesis underlying statistical data analysis (sample space, scale).
- The Aitchison geometry of the simplex.
- Coordinate representation; distributions on the simplex.
- Exploratory analysis (centering, variation array, biplot, balances-dendrogram).
- Introduction to available software: CoDaPack and R-packages.

http://www.compositionaldata.com//pages/codacourse_s.php

Chemometrics: Introduction to Advanced Data Analysis
August 8th-12th, 2016

The Analytical Chemistry/Chemometrics group and the Radboud Summer School (RSS) have the pleasure to present the first edition of the course *Chemometrics: Introduction to Advanced Data Analysis (Chemometrics-RSS16)*.

The research field of chemometrics focuses on the development and improvement of advanced data analysis methods for the analysis of chemical data, including e.g. removal of instrumental artefacts by data pre-processing, visualization by explorative data analysis, prediction of sample properties and comprehensive assessment of significance of the models by validation.

A week-long programme will cover the major aspects of Chemometrics with an introduction to the most commonly used data analysis method. These include methods for Data preprocessing, exploratory data analysis methods such as Principal Component Analysis (PCA) and regression/classification based on Partial Least Squares (PLS) method. Each subject will be introduced with a lecture, after which you will gain hands-on experience with the methods during (computer) exercises.

The course will be led by the internationally highly recognized chemometricians from the Analytical Chemistry department at Radboud University. The course will have guest lectures by Prof. Beata Walczak (University of Silesia) on strategies for data pre-processing and by Dr. Federico Marini (University Sapienza of Rome) on validation for the classification methods.

-
- Chemometrics-RSS16 will be held on 8th to 12th of August, 2016
 - Chemometrics-RSS16 is addressed to advanced Bachelor, Master and PhD students, as well as Postdoc researchers and industrial professionals.
 - The course does not require programming experience or affinity, all practical work will be done in interactive click-and-play software.
 - Chemometrics-RSS16 is worth 2 ECTS

- Application deadline: 15 June 2016
- How to enroll in the Chemometrics-RSS16? Please, fill in the application form at <http://www.ru.nl/radboudsummerschool/application/how-apply/> sending your curriculum vitae and a motivation letter. For more information, please visit <http://www.ru.nl/radboudsummerschool/> or contact r.folcarelli@science.ru.nl



5th VOC Conference
May 27, 2016
Tilburg University, The Netherlands
Dante Building, Room DZ-4

Book of Abstracts

Scope

The Dutch/Flemish Classification Society, VOC, aims at communicating scientific principles, methods, and applications of ordination and classification. The VOC is a member of the International Federation of Classification Societies (IFCS).

Functional data analysis of juggling trajectories

Paul Eilers

*Department of Methodology & Statistics,
Erasmus MC, Rotterdam, The Netherlands*

Seeing someone juggling balls is a fascinating sight. It becomes even more interesting if the movements of the juggler have been measured and are available for statistical modeling. The data I will analyze have been collected by Jim Ramsay and colleagues (Ramsay et al., 2013). They consist of ten trials of approximately ten seconds. The movements of a light emitting diode on the index finger of the juggler were recorded with high precision, in three directions: left-right, forward-backward and up-down.

A special section on functional data analysis in the Electronic Journal of Statistics (Volume 8, part 2, 2014) contains nine papers on the analysis of these juggling data. They all first split the data in cycles, based on chosen landmarks, and then apply different alignment or warping procedures within the cycles. In contrast, I model the series with extensions of the smooth complex logarithm model (Eilers, 2010). This model was developed for the analysis of chirp-like signals, like the sounds of crickets or bats. It fits a sine wave with variable frequency and amplitude to a time series. The juggling data, especially the left-right and forward-backward components have a more complex shape, but by adding harmonics with multiples of the fundamental frequency we get a very good fit to the data.

References

- Eilers, P.H.C. (2010) The Smooth Complex Logarithm and Quasi-Periodic Models. In: Statistical Modelling and Regression Structures. Kneib, T. and Tutz, G. (eds.) Springer.
- Ramsay, J.O., Gribble, P. and Kurtek, S. (2014) Description and processing of functional data arising from juggling trajectories. Electronic Journal of Statistics 8, 1811-1816.

Alzheimer's disease classification using resting state fMRI

Frank de Vos

Leiden University, Leiden, The Netherlands

Introduction

Alzheimer's disease (AD) affects millions of people worldwide and its prevalence increases rapidly. In order to develop a cure, or to slow down the disease progression, it is essential to diagnose AD in an early stage (Prince et al., 2011). Resting state functional magnetic resonance imaging (RSfMRI) scans reveal different patterns of functional connectivity (FC) in AD patients (Binnewijzend et al., 2012). FC is defined as the statistical dependency (e.g., correlation) between the time courses of brain regions. Various approaches exist to analyse FC, and it is unclear which of these approaches are most informative for the individual classification of AD patients. Moreover, the combination of different approaches might improve AD classification (de Vos et al., 2016).

Methods

We investigated various RSfMRI measures in a sample of 77 mild to moderate AD patients (MMSE = 20.4 ± 4.5) and 173 controls (MMSE = 27.5 ± 1.8). We calculated i) FC matrices between brain regions using full and sparse partial correlation, ii) the dynamics of these FC matrices over time, iii) the graph properties of the FC matrices, and iv) the time spent within different FC states. Furthermore, for each voxel we calculated, v) FC with the hippocampus, vi) FC with 10 resting state networks and, vii) the eigenvector centrality. Additionally, we measured the amplitude of the low frequency fluctuations within the resting state signal. We used the eight sets of features separately and combined as predictors in an elastic net logistic regression. We used a nested cross validation approach to tune the penalty parameters and to fit the logistic regression model independently. We calculated the area under the receiver operating characteristic curves (AUC) to determine classification performance.

Results

The FC matrices (AUC = 0.80) and FC dynamics (AUC = 0.81) provided the best classification, followed by the FC states (AUC = 0.74), the graph metrics (AUC = 0.73), and the amplitude of low frequency fluctuations (AUC = 0.68). FC with the resting state networks (AUC = 0.61) and eigenvector centrality (AUC = 0.61) performed poorly and FC with the hippocampus did not perform better than chance (AUC = 0.49). The combination of all measures (AUC = 0.81) did not improve over the best single measure.

Discussion

Our results demonstrate that it is possible to obtain moderate to good AD classification using FC matrices between ICA components and the dynamics of those FC matrices and that the combination of several resting state measures does not improve classification accuracy.

References

Binnewijzend M a a, Schoonheim MM, Sanz-Arigita E, Wink AM, van der Flier WM, Tolboom N, Adriaanse SM, Damoiseaux JS, Scheltens P, van Berckel BNM, Barkhof F (2012): Resting-state fMRI changes in

- Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 33:2018–28.
<http://www.ncbi.nlm.nih.gov/pubmed/21862179>.
- Prince, M., Bryce, R., & Ferri, C. (2011). *World Alzheimer Report 2011*. London. Retrieved from
<http://www.alz.co.uk/research/world-report-2011>
- Sheline YI, Raichle ME (2013): Resting state functional connectivity in preclinical Alzheimer's disease. *Biol Psychiatry* 74:340–347. <http://dx.doi.org/10.1016/j.biopsych.2012.11.028>.
- De Vos, F., Schouten, T., Hafkemeijer, A., Dopper, E., van Swieten, J., de Rooij, M., ... Rombouts, S. (2016). Combining multiple anatomical MRI measures improves Alzheimer's classification. *Human Brain Mapping*, 00 (October 2015). <http://doi.org/10.1002/hbm.23147>

Towards Bayesian analysis of GCxGC and GCxGC-MS data

Andrei Barcaru, Gabriel Vivó-Truyols

University of Amsterdam, Amsterdam, The Netherlands

The applications and the complexity of the multidimensional gas chromatographic data sometimes require solutions that are beyond those used in the “traditional chemometrics”. PCA, PLS, MCR and frequentist statistics are a few of the currently widely used in our field. However, modern Big Data analysis makes use of a wider spectra of algorithms. We intend to through a glimpse of light on Bayesian statistic using practical applications such as peak tracking and pairwise comparison of two dimensional gas chromatographic datasets. We present two examples in which the Bayesian framework is specially suitable. In the first example, a Bayesian method for “Peak tracking” is presented. Peak matching consists of a peak assignment of two chromatographic sets of the same sample taken in different conditions (i.e. changing temperature gradient and/or pressure). When the detector is not multi-channel (like the case of FID), the peak assignment carries a high degree of uncertainty. These uncertainties originates in the lack of mass spectral dimension and in the change of the elution order due to changes in experimental conditions. In this situation, a Bayesian approach of ranking probabilistically all possible candidates of the best match, seems more appropriate than completely excluding all but one outcome. In the second example, the comparison of GCxGC-MS chromatograms (to find differences) is presented. This comparison is also a challenging task even with mass spectral dimension present. Here, the uncertainty lies in the data shift from one experiment to another, and therefore a simple overlay of the data sets can generate many false positives. Aligning two samples of this size and complexity is not the most optimal solution. For this purpose we applied Information technology metric, namely Jensen-Shannon divergence, combined with Bayesian hypothesis testing to retrieve a probabilistic map of the location of meaningful differences.

A variable selection method for simultaneous component based data integration

Zhengguo Gu, Katrijn Van Deun

Tilburg University, Tilburg, The Netherlands

The integration of multiblock high throughput data from multiple sources is one of the major challenges in several disciplines including metabolomics, computational biology, genomics, and clinical psychology. A main challenge in this line of research is to obtain interpretable results 1) that give an insight into the common and distinctive sources of variations associated to the multiple and heterogeneous data blocks and 2) that facilitate the identification of relevant variables. We present a novel variable selection method for performing data integration, providing easily interpretable results, and recovering underlying data structure such as common and distinctive components. The flexibility and applicability of this method are showcased via numerical simulations and an application to metabolomics data.

The return of Neural Networks

Eric Postma

Cognitive Science & AI, TiCC (Tilburg University, The Netherlands) & Jheronimus Academy of Data Science (JADS)

Neural networks are again in the spotlight . This time under the name of " Deep Learning " and "Convolutional Neural Networks." The use of neural networks results in huge performance improvements in visual and auditory classification tasks amongst others. Technology companies like Facebook, Google and Baidu effectively apply neural networks to their extensive databases and start-ups make good use of pattern recognition capabilities of neural networks. Neuroscientists and psychologists use neural networks as an effective model of the visual system. Renowned scientists and entrepreneurs warn about the dangers of AI techniques that are based on neural networks. In short, there is a technological revolution.

During this presentation, these exciting developments will be put into a scientific perspective in a down-to-earth way.

Profile Analysis: a complementary way of reporting results in large assessments

Norman Verhelst

Eurometrics, The Netherlands

In large national or international assessments commonly a unidimensional IRT model is used to analyse the data, and groups are compared to each other assuming the validity of the measurement model. The comparisons are always related to the level of performance, e.g., is country A performing better than country B. But other interesting comparisons are possible, which are in a sense orthogonal to the previous ones: given the level of performance, can we show that a country performs better than expected in some subdomain, while another country performs more poorly. The expectation is expressed relative to the measurement model used. The subscores for several domains considered jointly constitute the (observed) profile, and the analysis finds out whether the deviations of the observed profile to the expected profile are significant or not. The TIMSS-2011 mathematics data will be used to demonstrate the power of such an analysis.

Continuous norming of psychological tests: A comparison of stepwise model selection procedures

Lieke Voncken, Casper Albers, Marieke Timmerman

University of Groningen, Groningen, The Netherlands

The availability of proper norms is essential to sensibly interpret test scores. For psychological tests scores that change with age, the state-of-the-art methodology for estimating norms is the so-called continuous norming, rather than norming per subgroup. The core idea is to build a statistical model that relates age to the distribution of test scores. This distribution is then subsequently used to derive the norms. In this way, the available information from the whole norm group is used in estimating the norms. A promising continuous norming approach is the use of the generalized additive models for location, scale, and shape (GAMLSS, originating from Rigby & Stasinopoulos, 2005). GAMLSS allows for modelling differences in center, spread, skewness, and kurtosis as a function of age. This can be implemented as a kind of polynomial regression. However, applying GAMLSS for norming to empirical norming data involves the selection of a specific model. For instance, one needs to select the degree of the polynomial. In our study, we will compare different stepwise model selection procedures and evaluate their precision in estimating norms using a simulation study. Specifically, we compare two stepwise model selection procedures (the default in the GAMLSS R-package vs. a newly developed, more flexible procedure) using four model selection criteria (AIC, BIC, GAIC(3) and cross-validation) and two different sample sizes. The data complexity is varied by manipulating the complexity of the relationship between age and the different model parameters. The performance is evaluated in terms of the difference between the true and estimated distributions of test scores against age. We provide an overview of the comparative performance of the procedures and selection criteria in the various conditions of data complexity. We will discuss implications for model selection in continuous norming with the GAMLSS approach.

References

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, 54(3), 507-554.

Understanding social-class differences in the transition to adulthood using Markov chain models

Yu Han¹, Cees. H. Elzinga², Aart. C. Liefbroer³

1NiDi/KNAW & RUG, Den Haag, The Netherlands

2NiDi/KNAW & VU

3NiDi/KNAW, VU & UMCG

Recent theories about social and demographic change, such as individualization theory and Second Demographic Transition theory, suggest the emergence of a type of late, protracted and complex pathway to adulthood. In recent years, studies offer qualified support for the emergence of this new pattern of transition to adulthood in most European countries. However, the transition to adulthood is a complex process of a series of events that are often interlinked. Even though life courses are greatly varying sequences of roughly the same life course events, the complexity is caused by the fact that these sequences consist of correlated events and spells and these correlations depend on gender, social class, cohort and cohort-related macro events. Our previous work demonstrated that the application of stochastic models like the Latent-Class model helps to describe the variation in life courses and its correlation with gender and social class. But the Latent-Class model cannot account for correlated events within life courses nor can it account for switches between latent types during the life course. We argue that (Hidden) Markov models, as a simple generalization of the Latent-Class model, has the ability to account for correlations between events and spells and also allows for switches between latent types or model life courses. Therefore, this study will use (Hidden) Markov models to produce a typology of trajectories of the transition to adulthood. Furthermore, we will test hypotheses on social class- and gender differences in observed life courses and latent types or model-life courses, using data from the Gender and Generation Programme (GGP), which provides full monthly life course sequence data between age 15 to 40.

Structure estimation for time-varying mixed graphical models in high-dimensional data

Jonas Haslbeck, Lourens Waldorp

University of Amsterdam, Amsterdam, The Netherlands

Dependencies in multivariate systems (graphical models) have become a popular way to abstract complex systems and gain insights into relational patterns among observed variables. For temporally evolving systems, time-varying graphical models offer additional insights as they provide information about organizational processes, information diffusion, vulnerabilities and the potential impact of interventions. In many of these situations the variables of interest do not follow the same type of distribution, for instance, one might be interested in the relations between physiological and psychological measures (continuous) and the type of prescribed drug (categorical) in a medical context. We present a novel method based on generalized covariance matrices and kernel smoothed neighborhood regression to estimate time-varying mixed graphical models in a high-dimensional setting. In addition to our theory, we present a freely available software implementation, performance benchmarks in realistic situations and an illustration of our method using a dataset from psychopathology.

A comparative analysis of psychometric unfolding models

Giulio Flore, Willem Heiser, Mark de Rooij

Leiden University, Leiden, The Netherlands

We present the results of a comparative analysis of simulation data of four unfolding Item Response Theory (IRT) models: Multiple Stochastic Unidimensional Unfolding (MUDFOLD); Generalized Hyperbolic Cosine Model (GHCM); PARALLELogram Analysis (PARELLA); Generalized Graded Unfolding Model (GGUM).

We assess the sensitivity of each model to Differential Item Functioning (DIF), Local Dependence (LD), non-standard distributional features of model parameters, and different items' starting values information. The simulations are based on a new unfolding IRT model, called Generalized Squared Distance Logistic Model (GSDLM). The analysis is framed as a set of ANOVA experimental designs.

The experiments show that bias and precision of items estimates are most affected by local dependencies and by the distribution of person locations on the latent trait. The bias and precision of individual latent trait values estimates are most affected by skewed distributions of personal latent traits and by the type of estimation methods. MUDFOLD is shown to be highly sensitive to departures in item and person latent traits assumptions.

These results identify three critical areas of research for a more general application of unfolding IRT models: 1) how to yield precise estimates for item and person latent traits scores for both symmetrical or asymmetrical distributions of the person latent trait; 2) more effective ways to determine initial item estimates compared to currently used or suggested methods; 3) development of more efficient estimation algorithms that can assess large datasets and filter out inconsistent item orderings.

A novel multivariate method for selection of immunological cells with multicolour flow cytometry

R. Folcarelli¹, L. Koenderman², L. M. C. Buydens¹, J. Jansen¹

1Analytical Chemistry, Radboud University, Nijmegen, The Netherlands

2 Respiratory Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.

Flow Cytometry (FC)-based gating allows the selection of single cells based on their expression of surface markers. Currently gating mostly is done bivariately, even though many more markers may be measured on the same cell using current FC technology. Only multivariate approaches may extract all aspects of cell variability from the data, including those associated with co-expression of multiple surface markers. A quantitative multivariate approach dedicated to gating would lead to objective analysis of cells that are activated during the immune response. In this context we have developed a method called Elimination of Cells Laying in Patterns Similar to Endogeneity (ECLIPSE) that provides a multivariate filter for cells of which the surface markers expression is also observed in healthy individuals. With those cells that pass the filter we build a Response model that focuses only on variability present in activated cells. The objectivity of all models is warranted by validation procedures in widespread use, specifically adapted to the multiset structure of MFC data.

The resulting ECLIPSE method provides a focused view on the variability in surface marker expression observed in cells that were quantitatively determined to be response-related.

Big Data in the Social Sciences: Statistical methods for multi-source high-dimensional data

Katrijn Van Deun

Tilburg University, Tilburg, The Netherlands

Social science research has entered the era of big data: Many detailed measurements are taken and multiple sources of information are used to unravel complex multivariate relations. For example, in studying obesity as the outcome of environmental and genetic influences, researchers increasingly collect survey, dietary, biomarker and genetic data from the same individuals. Although linked more-variables-than-samples (called high-dimensional) multi-source data form an extremely rich resource for research, extracting meaningful and integrated information is challenging and not appropriately addressed by current statistical methods. A first problem is that relevant information is hidden in a bulk of irrelevant variables with a high risk of finding incidental associations. Second, the sources are often very heterogeneous, which may obscure apparent links between the shared mechanisms. Hence, a statistical framework is needed to (i) select the relevant groups of variables within each source and (ii) link them throughout data sources. To address these issues, we will present an extension of sparse simultaneous component analysis to a component method that 1) finds the common components and 2) that selects relevant clusters of variables.