

VOC Newsletter 59 March 2019 Vereniging voor Ordinatie en Classificatie / Dutch-Flemish Classification

Society

Chairman:Mark de Rooij, Universiteit Leiden, Faculteit Sociale Wetenschappen, Departement Methoden en
Technieken, Postbus 9555, 2300 RB Leiden, Nederland (rooijm@fsw.leidenuniv.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:Pieter Schoonees, Erasmus Universiteit Rotterdam, Rotterdam School of Management, Departement of
Marketing Management, Postbus 1738, 3000 DR Rotterdam, Nederland (schoonees@rsm.nl)

VOC website: <u>http://www.voc.ac</u> Postbankrekening (IBAN) NL86 INGB 0000 161723 t.n.v. Vereniging voor Ordinatie en Classificatie.

8 th VOC Conference 5 April 2019 Groningen University		In this issue:	
Room B.0126, Bouman Building		Short program 8 th VOC Conference	1
Grote Rozenstraat 31, 9712 TG Groningen		Registration details for the 8 th VOC Conference	1
11:00-11:15	Welcome and registration	From the President	2
11:15-11:45	VOC annual members meeting	Publications	2
11.45-12.30	Contributed namer session 1	Other Meetings	4
11.45 12.50		Annual Report of the Secretary for the year 2018	5
12:30-13.30	Lunch	Minutes of the Members Meeting 2018 (Utrecht)	5
13:30-14:20	Keynote address: Casper Albers	Agenda Annual Members Meeting 2019 (Groningen)	6
	Intensive longitudinal data analysis in practice – Possibilities and limitations	Financial Report for 2018	7
		Programme 8th VOC Conference	8
14:20-15:00	Contributed paper session 2	Route description	9
15:00-15:20	Coffee break	Book of Abstracts 8th VOC Conference	11
15:20-16:20	Contributed paper session 3		
16:20-16:25	Short break	Registration details for the 8 th VOC Conference	
16:25-16:55	Contributed paper session 4		
16.55-17:00	Announcement of the VOC PhD Student Presentation Award	Those who would like to join the 8 th VOC Conference are welcome and are kindly requested to register through our website http://www.voc.ac. Details are provided through the website.	
17.00-	Closing and drinks		

From the President

Dear VOC members,

This is an exciting year for the VOC because we celebrate our 30th anniversary. For this reason, the VOC board, together with Paul Eilers, organizes a two day conference on November 20 - 21 in Wageningen. We will provide more detailed information on the program at a later moment but it promises to be a highly exciting meeting with distinguished invited international and national speakers. The large variety in topics of expertise of the speakers is characteristic for our society. Besides all scientific work there will be plenty of time to talk with classification friends in between the sessions and during lunches/dinner. So, already mark these dates in your calendar, you do not want to miss this event!

Before the anniversary meeting, we organize our yearly conference on April 5 at Groningen University. This year the organization is in the hands of our new board member, Matthijs Warrens. Casper Albers is our key-note speaker and we are very happy that he is willing to present his work to the VOC. Casper was just appointed to be a professor in Applied Statistics and Data Visualization (inaugural lecture is on March 19). A detailed program for this annual conference can be found in this newsletter, where you also find the material for our member meeting.

Furthermore, the International Federation of Classification Societies has its biennial conference this year in Thessaloniki, Greece (August 26 - 29). The VOC organizes an invited session where members of the VOC will present their work to the community. More information about this conference can be found on http://ifcs2019.gr/.

Thus, for the ones who didn't do so yet: go to the VOC website and register for our annual meeting! I hope to see you all in Groningen or later this year in Wageningen and/or Thessaloniki.

Mark de Rooij

Publications

Al-Thani, A. M., Voss, S. C., Al-Menhali, A. S., Barcaru, A., Horvatovich, P., Al Jaber, H., ... & Segura, J. (2018). Whole Blood Storage in CPDA1 Blood Bags Alters Erythrocyte Membrane Proteome. *Oxidative Medicine and Cellular Longevity*.

Bogers, Hein, Evelyne M. van Uitert, Sharon van Ginkel, Elisabeth DH van der Mooren, Irene AL Groenenberg, Paul HC Eilers, Niek Exalto, Eric AP Steegers, and Régine PM Steegers-Theunissen. (2018). Human embryonic curvature studied with 3D ultrasound in ongoing pregnancies and miscarriages. *Reproductive Biomedicine*, 36(5), 576-583.

Cariou, V., & Wilderjans, T. F. (2018). Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. *Food Quality and Preference*, 67, 18-26. https://doi.org/10.1016/j.foodqual.2017.01.006.

D'Ambrosio, A. & Heiser, W.J. (2018). A distributionfree soft-clustering method for preference rankings. *Behaviormetrika*, online

(https://doi.org/10.1007/s41237-018-0069-5).

De Raadt, A., Warrens, M. J., Bosker, R. J., & Kiers, H. A. (2019). Kappa Coefficients for Missing Data. *Educational and Psychological Measurement*, 0013164418823249.

Eilers, P. H. (2018). The truth about the effective dimension. *Statistica Neerlandica*, 72(3), 201-209.

Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruyen, P. M. (2018). An empirical analysis of alleged misunderstandings of coefficient alpha. International Journal of Social Research Methodology, 1-14.

Holmgren, A., Aimon Niklasson, Andreas F.M. Nierop, Lars Gelander, A. Stefan Aronson, Agneta Sjöberg, Lauren Lissner and Kerstin Albertsson-Wikland. (2018). Estimating secular changes in longitudinal growth patterns underlying adult height with the QEPS model: the Grow Up Gothenburg cohorts. *Pediatric Research*, 84(41–49), doi: 10.1038/s41390-018-0014-z.

Joachim G.C. Deru, Jaap Bloem, Ron de Goede, Nyncke Hoekstra, Harm Keidel, Henk Kloen, Andreas Nierop, Michiel Rutgers, Ton Schouten, Jan van den Akker, Lijbert Brussaard, Nick van Eekeren. (2019). Predicting soil N supply and yield parameters in peat grasslands. *Applied Soil Ecology*, 134, 77-84, doi: 10.1016/j.apsoil.2018.10.018

Kroonenberg, P.M., & Verbeek, A. (2018). The tale of Cochran's Rule. My contingency table has so many expected values smaller than five. What am I to do? *The American Statistician*, 72, 175-183. DOI: 10.1080/00031305.2017.1286260.

Kroonenberg, P.M. (2018). Bibliographic analyses of handbooks: A proposal. Handbooks as a source of data for mapping scientific disciplines. In CWTS (Ed.), *STI 2018 Conference Proceedings*. (pp. 1609–1618). Leiden, The Netherlands. On-line only:

https://openaccess.leidenuniv.nl/handle/1887/652 25.

Kwiatkowski, M., Krösser, D., Wurlitzer, M., Steffen, P., Barcaru, A., Krisp, C., ... & Schlüter, H. (2018). Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability for the Analysis of Complex Proteomes. *Analytical Chemistry*, 90(16), 9951-9958.

Martin, M., Legat, B., Leenders, J., Vanwinsberghe, J., Rousseau, R., Boulanger, B., Eilers, P.H., De Tullio, P. and Govaerts, B., (2018). PepsNMR for 1H NMR metabolomic data pre-processing. *Analytica Chimica Acta*, 1019, pp.1-13.

van der Hoef, H., & Warrens, M. J. (2019). Understanding information theoretic measures for comparing clusterings. *Behaviormetrika*, 1-18. Van Schie, C.C. Chiu, C.-D., Rombouts, S. A. R. B., Heiser, W.J., & Elzinga, B.M. (2018). When compliments do not hit but critiques do: An fMRI study into self-esteem and self-knowledge in processing social feedback. *Social Cognitive and Affective Neuroscience*, 13, 404-417 (https://doi.org/10.1093/scan/nsy014).

Voncken, L., Albers, C. J., & Timmerman, M. E. (2018). Improving confidence intervals for normed test scores: Include uncertainty due to sampling variability. *Behavior Research Methods*. doi: 10.3758/s13428-018-1122-8

Waaijenborg, S., Korobko, O., Willems van Dijk, K., Lips, M., Hankemeier, T., Wilderjans, T. F., Smilde, A. K., & Westerhuis, J. A. (2018). Fusing metabolomics data sets with heterogeneous measurement errors. *PLoS ONE*, 13, e0195939, 1-19. https://doi.org/10.1371/journal.pone.0195939

Warrens, M. J. (2019). Similarity measures for 2× 2 tables. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-14.

Warrens, M. J., & de Raadt, A. (2018). Properties of Bangdiwala's B. *Advances in Data Analysis and Classification*, 1-13.

Warrens, M. J. (2018). On the Negative Bias of the Gini Coefficient due to Grouping. *Journal of Classification*, 35(3), 580-586.

Wijsen, L.D., Borsboom, D., Cabaco, T., & Heiser, W.J. (2019). An Academic Genealogy of Psychometric Society Presidents. *Psychometrika*, on line (https://doi.org/10.1007/s11336-018-09651-4).

Book announcement

Greenacre, M. (2018). Compositional Data Analysis in Practice. Chapman and Hall/CRC.

Compositional Data Analysis in Practice is a user-oriented practical guide to the analysis of data with the property of a constant sum, for example percentages adding up to 100%. Compositional data can give misleading results if regular statistical methods are applied, and are best analysed by first transforming them to logarithms of ratios. This book explains how this transformation affects the analysis, results and interpretation of this very special type of data. All aspects of compositional data analysis are considered: visualization, modelling, dimension-reduction, clustering and variable selection, with many examples in the fields of food science, archaeology, sociology and biochemistry, and a final chapter containing a complete case study using fatty acid compositions in ecology. The applicability of these methods extends to other fields such as linguistics, geochemistry, marketing, economics and finance.



Other Meetings

16th Conference of the International Federation of Classification Societies

26 – 29 August, Thessaloniki, Greece

The IFCS 2019 conference theme is 'Data Analysis and Rationality in a Complex World'. The conference opening will take place on August 26 late afternoon and pre-conference workshops will be held. The conference sessions will start on August 27 in the morning, and will close on August 29 with a full day conference program and a conference dinner. The conference will include a President's invited session and a Presidential address, invited presentations, invited and contributed sessions. See https://ifcs.gr/ for more information.

The programme includes several keynotes. Abstract submission closes 5 May 2019. Early bird registration is available until 26 May.



26-29 Aug 2019 Thessaloniki Concert Hall Thessaloniki, Greece #IFCS2019

Data Science, Statistics & Visualisation (DSSV 2019)

13 – 15 August 2019, Doshisha University, Kyoto, Japan

The conference of Data Science, Statistics & Visualisation (DSSV 2019) will take place 13 – 15 August in Kyoto, Japan; see https://iasc-isi.org/dssv2019/. This meeting shall create a forum to discuss recent progress and emerging ideas in these different disciplines and encourage informal contacts and discussions among all the participants.

The conference welcomes contributions to practical aspects of data science, statistics and visualisation, and in particular those which are linking and integrating these subject areas. Presentations should thus be oriented towards a very wide scientific audience, and can cover topics such as machine learning and statistical learning, the visualisation and verbalization of data, big data infrastructures and analytics, interactive learning, advanced computing, and other important themes.



Keynotes will be given by Di Cook, Kwan-Liu Ma and Richard Samworth. The deadline for abstract submission is due to 7 May 2019. Early bird registration ends on 30 April 2019.

Annual Report of the Secretary for the year 2018

1. Number of members

The VOC started 2018 with 113 members and counted 93 members at the end of 2018. Twenty two memberships were terminated and there were two new members registered. In 2018, 40 members paid contribution.

2. Board

The Board of the VOC was composed as follows in 2018:

Mark de Rooij Katrijn Van Deun Tom Wilderjans Pieter Schoonees Matthijs Warrens Jeroen Jansen Hilde Tobi President Secretary Treasurer Newsletter Editor Webmaster Member Member

The Board met once in 2018. The main topic was the organization of the 7th VOC conference.

3. Activities

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

The seventh VOC Conference took place at Utrecht University (the Netherlands) on the 25th of May 2017 with a full day program, including eight contributions by VOC members on a range of topics. Keynote contributions were given by Ellen Hamaker (Utrecht University, the Netherlands) on 'Intensive longitudinal data in the social and behavioral sciences: Challenges and solutions' and by Johannes Textor (Radboud University Medical Center, the Netherlands) on 'How the Immune System Learns to Tell Apart Self from Foreign'. 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.

Minutes from the VOC Annual Members Meeting (25th of May 2018, Utrecht)

1. Opening of the Members Meeting

2. Minutes of the Members Meeting 27th May 2016

The minutes of this meeting were approved.

3. Annual Report of the Secretary on the year 2016

We discussed this report and the difficulty with the annual conferences. There does not seem to be a lot of enthusiasm among the members to visit conferences. Maybe the timing of the conference is not ideal because the month May is for many quite busy. Earlier meetings, like in February/March, would be better.

4. Financial report of the treasurer on the year 2016

Mark reads the rapport of the cash committee. The members decharge the treasurer.

5. Composition of the Board

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

- Mark de Rooij, President (1)
- Katrijn Van Deun, Secretary (2)
- Tom Wilderjans, Treasurer (0)
- Pieter Schoonees, Newsletter Editor (2)
- Jeroen Jansen, board member (0)
- Hilde Tobi, board member (2)

Tom and Jeroen terms finish; both are willing to serve another three years. The board suggests to add Matthijs Warrens to the board. The members agree.

6. Miscellaneous

7. Questions before closure of the meeting

Paul Eilers asks when there will be a jubilee meeting again. This will be next year, in 2019. Paul is willing to help in organizing this event.

8. Closure of the Members Meeting

Meeting closed.

Agenda for the VOC Annual Members Meeting (5th April 2019, Groningen University)

- 1. Opening of the Members Meeting
- 2. Minutes of the Members Meeting 25th May 2018

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

3. Annual Report of the Secretary on the year 2018

The Annual Report is included in this Newsletter (see p. 5).

4. Financial report of the treasurer on the year 2018

The Financial Report is included in this Newsletter (see p. 7).

5. Composition of the Board

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

- Mark de Rooij, President (0)
- Katrijn van Deun, Secretary (1)
- Tom Wilderjans, Treasurer (2)
- Pieter Schoonees, Newsletter Editor (1)
- Matthijs Warrens, Webmaster (2)
- Jeroen Jansen, board member (2)
- Hilde Tobi, board member (1)

The terms of Mark de Rooij ends. Candidates for this position can register up to 24 hours before the meeting with the Secretary.

- 6. Miscellaneous
- 7. Questions before closure of the meeting
- 8. Closure of the Members Meeting

Financial Report for 2018

Revenue		Expenditure	
membership fees (40 paying members)	1000	IFCS dues 2017-2018	199,53
overdue membership fees	280	hosting website	121
interest savings account	4,7	present speakers VOC meeting Utrecht	75
		room and catering VOC meeting Utrecht	717,05
		domain registration	44,25
		drinks meeting organization jubilee	27,05
		transaction costs ING	110,64
Total	1284,7	Total	1294,52
Debit		Credit	
Balance ING account	1388,39	Accounts payable	0
Balance savings account	5866,13	Equity	7254,52
Total	7254,52	Total	7254,52

Notes to the balance sheet

(1) Contributions from 40 members have been collected in 2018
(2) In 2018 we collected 280 euro's of overdue membership fee's
(3) the equity stayed the same as last year (revenue and expenditure cancel each other out)
(4) An overview of the evolution of the equity

2018	€ 7 254,52
2017	€7264,34
2016	€ 5 432,34
2015	€ 3 913,66
2014	€ 4 019,92
2013	€ 5 444,46
2012	€ 5 524,70
2011	€ 6 194
2010	€7621
2009	€ 8 189
2008	€6248
2007	€ 5 914
2006	€6869
2005	€6057
2004	€ 5 019
2003	€ 6 795
2002	€ 6 408
2001	€ 5 898
2000	€ 5 731
1999	€4871
1998	€ 5 100

Programme: 8th VOC Meeting Groningen, 5 April 2019

Room B.0126 (Gadourekzaal), Bouman Building, Grote Rozenstraat 31, 9712 TG Groningen

- 11:00-11:15 Welcome and registration
- 11:15-11:45 VOC Annual Members Meeting
- 11:45-12:30 Submitted paper session 1

Wouter van Loon* – Selecting views in multi-view learning

Cajo ter Braak – Model-based or simpler? The example of trait-environment association in ecology

12:30-13.30 Lunch

13:30-14:20 Keynote address

Casper Albers – Intensive longitudinal data analysis in practice – Possibilities and limitations

14:20-15:00 Submitted paper session 2

Jorien Vugteveen* – Using the multi-informant Strengths and Difficulties Questionnaire for predicting adolescent psychiatric disorders

Tim Offermans* – One-class modelling of PAT data for detection of byproduct formation

15:00-15:20 Coffee break

15:20-16:20 Submitted paper session 3

Andrei Barcaru – Chasing the interesting in the data with the Supervised Projection Pursuit

Denise Blom – Identifying school motivation profiles of students in secondary education using cluster analysis

Jeffrey Durieux* – Detecting disease subtypes by means of Clusterwise Independent Component Analysis (C-ICA) for multi-subject resting state fMRI data

16:20-16:25 Short break

16:25-16:55 Submitted paper session 4

Patrick Groenen – A Fast MM Algorithm for Clusterpath

16.55-17:00 Announcement of the PhD student Presentation Award

17.00- Closing and drinks

* = applicant for the VOC PhD student Presentation Award

Route description to meeting location

Room B.0126 (Gadourekzaal), Bouman Building Grote Rozenstraat 31, 9712 TG Groningen

Please note that the room (Gadourekzaal, B.0126) is in the Bouman building. This building hosts the Sociology department. Do NOT confuse the room name with that of the Gadourek building, which is close by.

A map of the Faculty of Behavioural and Social Sciences is available at: <u>https://www.rug.nl/staff/location/plattegrond-gmw.pdf</u>

Car park

There are limited (paid) parking spaces in the area. Visitors are therefore advised to use the general parking facility Ossenmarkt (<u>https://www.rug.nl/staff/location/0001</u>). Starting at the Ossenmarkt, walk into the first street (Nieuwe Boteringestraat) on your right-hand side without crossing the bridge. Pass by the 'Nieuwe Kerk' church to the right. Take the first turn left into the Kleine Rozenstraat and walk straight ahead over the crossing. From here, you should be able to see the Bouman building (number 31) located on the right.

Disabled people

The GMW-complex has a number of parking spaces available for disabled people. These are also intended for GMW employees who have temporary disabilities. Please contact the reception upon arrival if you wish to make use of these parking spaces. Reservations in advance cannot be made. The Bouman building has an adapted elevator for the disabled and special toilet facilities.

On foot or by bicycle

The conference location is at a 20 – 25 minute walk / 10 minute cycle from Groningen train station.

Public transportation

The conference location is 15 - 20 minutes by bus from Groningen train station. Several buses could be used: please see <u>https://9292.nl/</u> for more information. Note that all bus options require a 5 - 10 minute walk to reach the building.





8th VOC Conference 5 April 2019

Groningen University, the Netherlands Room B.0126, Bouman Building

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).

KEYNOTE

Intensive longitudinal data analysis in practice – Possibilities and Limitations

Casper Albers

Psychometrics, Groningen University

The past years have seen a steep rise in the use of intensive longitudinal data methods, such as experience sampling and momentary assessment, in psychological science. In my talk, I will discuss two recent studies involving longitudinal data.

The first study is of technical nature. Standard discrete time series models assume stationarity, which implies that the estimated parameters are constant throughout the time series. This is a mathematically convenient, but unrealistic assumption. In psychopathology research, the emotional dynamics are known to be varying both abrubtly as gradually. Models for abrubt change, such as regime switching models, already exist and so do models for gradual change. However, models that incorporate both gradual and abrupt changes were lacking. In this presentation, I will explain our time-varying change point autoregression model and show its merits for emotion research.

The second study is of applied nature. The number of models for psychological time series grows rapidly and each model claims to be the best model. Driven by curiosity about how different the inferences from these models would be, we asked twelve teams consisting of both clinical and methodological experts, to analyse the same data set from an individual diagnosed with major depressive disorder. From the results we can see that, with the current models, statistical prediction does not win from clinical prediction. Research and guidelines on how to best analyse psychological time series is necessary before it can be used in practice.

References:

C. J. Albers & L. F. Bringmann (2019). The time-varying change point autoregressive model. *In preparation*.

J. A. Bastiaansen, Y. K. Kunkels, ..., C. J. Albers & L. F. Bringmann (2019). Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology. *Submitted for publication*.

L. F. Bringmann & C. J. Albers (2019). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *Submitted for publication*.

Chasing the interesting in the data with the Supervised Projection Pursuit

Andrei Barcaru

UMCG Groningen

An important step in multivariate analysis is the dimensionality reduction, which allows for an easier visualization of the class structures in the data and sometimes a better classification. The techniques like PCA, PLS-DA and LDA dominated this side of data analysis for many decades. Yet the data does not always reveal properly the structures wen these techniques are applied. To this end, a supervised projection pursuit (SuPP) is proposed, based on Jensen-Shannon divergence . The combination of this metric with powerful Monte Carlo based optimization algorithm, yields a versatile dimensionality reduction technique capable of working with highly dimensional data and missing observations. In some cases, SuPP is capable to separate the classes in a latent space better than PLS-DA and PCA. Combined with Naïve Bayes classifier, SuPP is a powerful preprocessing tool for classification.

Identifying School Motivation Profiles of Students in Secondary Education using Cluster Analysis

Denise Blom

GMW Groningen

The importance of motivation of students is acknowledged in order to explain why some students learn and thrive in school contexts, while others struggle academically. As higher levels of motivation are associated with higher achievement in children, research on motivation in students is required as it will allow to design interventions to improve their future academic achievement. School motivation of 13,933 9th grade students in the Netherlands was investigated using cluster analysis methods. The aim consisted of identifying school motivation profiles in a four-dimensional motivation space, including mastery, performance, social, and extrinsic motivation.

Currently, there is no consensus on which clustering method is best suited for identifying school motivation profiles, and which validity index is best in determining the number of profiles (clusters). Therefore, we compared and will discuss the performance of two different clustering methods, a partitioning method (k-medoids) and a model based method (latent cluster analysis), and various validity indices (Dunn index, CH-criterion, BIC, and AIC). Furthermore, in many applications of cluster analysis, researchers aim at a relatively small number of clusters, approximately 10 or less. However, with small number of clusters objects in the same clusters may have large within cluster distances, which is not necessarily desirable. Therefore, we compared the performance of the clustering methods and validity indices using solutions with up to 40 clusters.

Selecting Views in Multi-View Learning

Wouter van Loon

Leiden University

Integrating information from different feature sets describing the same set of objects is known as multi-view learning. In biomedical research, such feature sets (views) may correspond to different data sources such as medical imaging modalities, questionnaires, and omics data. Views can also be defined within data obtained from the same source, for example as different feature sets derived from the same image, as different brain regions, or as gene sets. Integrating the information from different views can increase the accuracy of medical classification models. However, collection of biomedical data can be expensive and/or burdening for patients. Identifying the views that are most important for prediction can improve the understanding of disease and can contribute to reducing the amount of required data collection. This leads to a group-wise feature selection or 'view selection' problem.

A popular method for these kinds of problems is the group lasso, which places a penalty on the sum of L2-norms of predefined groups of features. However, this method is known to have a high false positive rate and can be very slow to compute for high-dimensional problems. We propose an alternative called Stacked Penalized Logistic Regression (StaPLR). Compared to the group lasso, StaPLR can make use of faster optimization algorithms and is easily parallelized. We compare StaPLR with the group lasso on both simulated and real data. We observe that StaPLR has a consistently lower false positive rate than the group lasso without incurring a major penalty in terms of predictive accuracy.

Detecting disease subtypes by means of Clusterwise Independent Component Analysis (C-ICA) for multi-subject resting state fMRI data

Jeffrey Durieux

Leiden University

An emerging challenge in the study of brain diseases and mental disorders, like dementia and depression, consists of revealing systematic differences and similarities between subgroups of patients in functional connectivity networks (FCN), that is, coordinated activity across brain regions. As such, existing subtypes of the disease may be characterized in terms of FCN and disease subtypes may get detected which transcend the current diagnostic boundaries.

In order to obtain FCN, researchers often collect resting state functional Magnetic Resonance Imaging (rs-fMRI) data and analyze this data with Independent Component Analysis (ICA). ICA is a data reduction technique that decomposes an rs-fMRI dataset into a set of FCN patterns and a mixing matrix that contains time courses of activation. The extracted FCN patterns can then be visualized as a three-dimensional brain image, indicating regions that show synchronized activity.

Analyzing the brain data of each patient separately with ICA has a major drawback that each patient will be characterized by different FCN, which makes it difficult to detect the systematic differences and similarities in FCN between (groups of) patients. Therefore, we propose Clusterwise Independent Component Analysis (C-ICA). The goal of this method is to cluster the patients into homogenous groups based on the similarities and differences in their FCN. As such, patients allocated to the same cluster are assumed to have similar FCN, whereas patients belonging to different clusters will be described by different FCN. This allows a data-driven detection of disease/disorder subtypes based on different FCN.

In this presentation, the C-ICA model is proposed, along with an alternating least squares type of algorithm to estimate its parameters. Further, the results of a simulation study to evaluate the performance of the novel C-ICA method are presented. In this simulation study, data are generated based on well-known FCN and the associated time courses of activation are simulated with fluctuation frequencies that are commonly encountered in rs-fMRI data. Lastly, the use of C-ICA is illustrated on a large empirical dataset concerning Alzheimer patients and healthy (matched) control subjects.

One-class modelling of PAT data for detection of byproduct formation

Tim Offermans

Radboud University Nijmegen

The goal of the presented study is to develop a method for automatic determination of a transition point of an industry-scaled biochemical batch reaction in real-time. Continuing the reaction beyond this transition point causes an increased fraction of byproducts, which leads to a decrease in yield and to a waste of energy and raw materials.

The measured data are spectroscopic signals that are collected on-line to follow the reaction in real-time, and which are available for different production batches. For each production batch, the transition point is indicated by an expert operator. A principal component analysis-based approach to one class-classification was applied to distinguish spectroscopic signals that were collected before the transition point from spectroscopic signals that were collected after the transition point.

An initial analysis showed that it is possible to, retrospectively, detect the reaction transition point in the spectroscopic signals for individual production batches with high accuracy. However, because of a high batch-to-batch variation of the spectroscopic signals, a model trained for one production batch performs unsatisfactory for another production batch. Current efforts therefore focus on extending the classification approach for production batch-invariant prediction of the transition point. Such an approach will allow for automated control of the ending of the reaction.

Model-based or simpler? The example of trait-environment association in ecology

Cajo ter Braak

Wageningen University

Statistical analysis of trait-environment association is challenging owing to the lack of a common observation unit: Community weighted mean regression (CWM) uses site points, multilevel models use species points, and the fourth corner correlation uses all species-site combinations. This situation invites the development of new methods capable of using all observation-levels. To this end, new multilevel and weighted averaging-based regression methods are proposed.

The new multi-level model contains not only a variance component for species-specific response to environment but also one for site-specific trait-response. The new weighted averaging approach is a weighted version of the existing approach, in which each site and species is weighted by Hill's effective number of values that are averaged. The sophisticated model-based method and the simpler weighted averaging method had similar power in simulations.

We note also that the natural extension of the weighted averaging method to the multi-trait, multi-environmental variable case is double constrained correspondence analysis, leading to novel insights in this method.

References

ter Braak, C. J. F., P. Šmilauer, and S. Dray. 2018. Algorithms and biplots for double constrained correspondence analysis. *Environmental and Ecological Statistics* 25: 171-197. https://doi.org/10.1007/s10651-017-0395-x

ter Braak, C.J.F. (2018) New robust weighted averaging- and model-based methods for assessing trait-environment relationships. *PeerJ Preprints*, 6, e27439v27431. https://doi.org/10.7287/peerj.preprints.27439v1

Using the multi-informant Strengths and Difficulties Questionnaire for predicting adolescent psychiatric disorders

Jorien Vugteveen

Groningen University

Knowledge on the validity of the Strengths and Difficulties Questionnaire (SDQ) among adolescents is limited but essential for the interpretation of SDQ scores in the screening or diagnostic process. I will present the results of a study that assessed the predictive value of the adolescent- and the parent-rated SDQ versions for psychiatric disorders in a clinical sample of 2,753 Dutch adolescents aged 12-17. Four types of DSM-IV diagnoses, established by professionals in outpatient community clinics, were taken into account: Attention-Deficit/Hyperactivity Disorder (ADHD), Conduct/Oppositional Defiant Disorder (CD/ODD),

Anxiety/Mood disorder, Autism Spectrum Disorder (ASD). Per type of diagnose the predictive accuracy of the SDQ scale that content-wise is most closely related to that particular disorder was assessed with logistic regression analysis. That is, 12 logistic regression analyses were performed: one for each combination of DSM-IV diagnosis (4, as mentioned above), informant (3: adolescent, parent, both). Additional logistic regression analyses were performed to assess the discriminative strength of the SDQ scales. The results show both fair predictive and fair discriminative strength for the adolescent- and parent-reported hyperactivity scales, the parent-reported conduct scale, and the parent-reported social and prosocial scales, indicating that these scales provide useful information about the presence of ADHD, CD/ODD, and ASD, respectively. The SDQ emotional scale showed to be insufficiently predictive for Anxiety/Mood disorder. Our findings suggest that parent-rated SDQ scores can be used to provide clinicians with a preliminary impression of the type of problems for ADHD, CD/ODD, and ASD, and adolescent-rated scores for ADHD.

A Fast MM Algorithm for Clusterpath

Patrick Groenen

Erasmus University Rotterdam

Convex clustering is an elegant clustering approach proposed by Hocking et al. (2011) that rephrases k-means clustering into a convex optimization problem with a grouped-Lasso like penalty to induce equal cluster means. By changing the penalty strength parameter, a path is obtained that joins increasingly more clusters means. In this way, generally a hierarchy of clusters is obtained. The advantage of Clusterpath is that for each penalty strength the resulting minimization problem is convex in the means and hence a global minimum can be obtained. A disadvantage is that it cannot be guaranteed that it results in a hierarchical tree of clusters. A second disadvantage is that the algorithms proposed so far are quite slow. In this paper, we propose three innovations that each can be used independently. First, we propose to make use of MM algorithms that can decrease running time of the algorithm by a factor 3 to more than 10, depending on the options chosen. Second, to avoid bias of the cluster means towards the overall mean for larger penalty strengths, we suggest to use a concave shrinkage penalty. Third, a hierarchy of partitions can be explicitly imposed that comes with the additional bonus of seriously reducing the computational complexity. We present an illustrative example and some numerical results.