VOC Newsletter 53
October 2014

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:	Berrie Zielman, Algemene rekenkamer, Directie Beleid en Communicatie, Afd. Statistiek, Lange Voorhout 8, 2414 ED Den Haag, Nederland (a.Zielman@rekenkamer.nl)
	Postbankrekening (IBAN) NL86 INGB 0000 161723 t.n.v. Vereniging voor Ordinatie en Classificatie, Doornenburg 84, 2261 XD Leidschendam.
Editor:	Tom Wilderjans, KU Leuven, Faculteit Psychologie en Pedagogische Wetenschappen, Andreas Vesaliusstraat 2 bus 3762, B-3000 Leuven, België (tom.wilderjans@ppw.kuleuven.be)
VOC-home page:	http://www.voc.ac

25 years VOC Jubilee Meeting

6-7 November 2014 Rolduc

November 6

11.30	Welcome and Daniel Oberski
12.20	Lunch
13.30	Lianne Ippel, Jeroen Jansen and Marieke Timmerman
16.15	Paul Eilers and Jelle Goeman
17.45	Drinks and dinner

November 7

9.00	Willem Heiser and Lieven de Lathauwer
11.00	Christian Hennig and Marco Riani
12.30	Lunch
13.45	Patrick Groenen and Denny Borsboom
15.15	Closing

In this issue:

Registration details for the 25th Anniversary VOC Jubilee Meeting

We like to encourage all VOC members to take part in the 25th Anniversary VOC Jubilee Meeting that will be held next month (Tuesday 6-Friday 7 November) at Rolduc (Kerkrade). The program of the meeting can be found further on in this newsletter. To register, please go to the VOC website (www.voc.ac) and go to '*meeting*'. The participation fee is 275 euros and includes, besides a set of interesting presentations, a 1-night stay, 2 lunches, 1 dinner (drinks excluded), 1 reception and coffee/tea breaks.

From the President

Dear VOC members,

On Tuesday and Friday November 6-7, we will celebrate the 25th anniversary of the VOC. For this occasion, the VOC board together with Paul Eilers organizes a two-day Jubilee Conference in Kerkrade (at Rolduc). The scientific part of the program, which can be found further on in this newsletter, is really excellent. It contains a mix of contributions of researchers from different generations (going from PhD student to just retired, and everything in between), researchers from the Netherlands/Flanders and abroad, and researchers from a variety of substantive disciplines in which supervised and unsupervised classification is of core interest. In other words, the scientific program represents what the VOC stands for. I am sure that attending this meeting will give you an excellent overview on what is going on in our field and will moreover provide you lots of inspiration for your own research.

Of course, at the occasion of an anniversary, not only the scientific part but also the social part is very important. You will be able to meet your old classification friends and be able to make new friends. With long coffee breaks, two lunches and a dinner, the program is created such that there is plenty of time to meet and talk with one another.

Thus, for the ones who didn't do so yet: go to the VOC website and register for the VOC anniversary meeting!

I hope to see you all November 6-7 at Rolduc.

Jeroen Vermunt

News from IFCS

The next IFCS conference will take place July 6-8 2015 in Bologna, Italy.

Program 25th Anniversary VOC Jubilee Meeting (Rolduc, Kerkrade, 6-7 November 2014)

<u>Thursday November 6, 2014</u>

11.00	*Arrival*	
11.30	*Welcome*	
11.35	Daniel Oberski	Model fit evaluation by sensitivity analysis
12.20	*Lunch*	
13.30	Lianne Ippel	Estimating multi-level models in data-streams
14.15	Jeroen Jansen	Revealing the information within Flow Cytometry data using advanced and dedicated Chemometrics
15.00	Marieke Timmerman	Unraveling multivariate effects resulting from an experimental design
15.45	*Coffee & Tea*	
16.15	Paul Eilers	Generalized exponential tilting
17.00	Jelle Goeman	Two folklores: ridge regression versus the lasso
17.45	*Drinks*	
18.30	*Dinner*	

<u>Friday November 7, 2014</u>

9.00	Willem Heiser	From preference mapping to preference learning, with an example of a prediction tree for rankings
9.45	Lieven de Lathauwer	Tensor decompositions: golden tools for data mining
10.30	*Coffee & Tea*	
11.00	Christian Hennig	Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters
11.45	Marco Riani	Robust modern multivariate data analysis and classification
12.30	*Lunch*	
13.45	Patrick Groenen	Some recent biplots approaches
14.30	Denny Borsboom	All quiet on the psychometric front? Goals and challenges for 21st century psychometrics
15.15	*Closing*	

Publications

Albers, C. J., & Gower, J. C. (2014). A contribution to the visualisation of three-way arrays. *Journal of Multivariate Analysis, 132,* 1-8. doi:10.1016/j.jmva.2014.07.013

Albers, C. J., & Gower, J. C. (2014). Canonical Analysis: Ranks, ratios and fits. *Journal of Classification, 31*(1), 2-27. doi:10.1007/s00357-014-9146-y

Almansa, J., Vermunt, J. K., Forero, C. G., & Alonso, J. (2014). A factor mixture model for multivariate survival data. An application to the analysis of lifetime mental disorders. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 63*, 85-102.

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating latent class assignments to external variables: standard errors for corrected inference. *Political Analysis*, *22*, 520-540.

Barendse, M. T., Albers, C. J., Oort, F. J., & Timmerman, M. E. (in press). Measurement bias detection through Bayesian factor analysis. *Frontiers in Psychology*. doi:10.3389/fpsyg.2014.01087

Bennink, M., Croon, M. A., Keuning, J., & Vermunt, J. K. (2014). Measuring student ability, classifying schools, and detecting item-bias at school-level based on student-level dichotomous attainment items. *Journal of Educational and Behavioral Statistics, 39*, 180-202.

Bryan, S. R., Vermeer, K. A., Eilers, P. H. C., Lemij, H. G., & Lesaffre, E. M. E. H. (2013). Robust and censored modeling and prediction of progression in glaucomatous visual fields. *Investigative Ophthalmology & Visual Science, 54*(10), 6694-6700.

Bulteel, K., Ceulemans, E., Thompson, R. J., Waugh, C. E., Gotlib, I. H., Tuerlinckx, F., & Kuppens, P. (2014). DeCon: A tool to detect emotional concordance in multivariate time series data of emotional responding. *Biological Psychology, 98*, 29-42. doi:10.1016/j.biopsycho.2013.10.011

de Rooi, J. J., Ruckebusch, C., & Eilers, P. H. C. (2014). Sparse deconvolution in one and two dimensions: Applications in endocrinology and single-molecule fluorescence imaging. *Analytical Chemistry, 86*(13), 6291-6298.

de Rooi, J. J., van der Pers, N. M., Hendrikx, R. W. A., Delhez, R., Bottger, A. J., & Eilers, P. H. C. (2014). Smoothing of X-ray diffraction data and K alpha(2) elimination using penalized likelihood and the composite link model. *Journal of Applied Crystallography*, *47*(3), 852-860.

De Roover, K., Timmerman, M. E., De Leersnyder, J., Mesquita, B., & Ceulemans, E. (2014). What's hampering measurement invariance: Detecting non-invariant items using clusterwise simultaneous component analysis. *Frontiers in Psychology, 5,* 1-11. doi:10.3389/fpsyg.2014.00604

De Roover, K., Timmerman, M. E., Van Diest, I., Onghena, P., & Ceulemans, E. (2014). Switching principal component analysis for modeling means and covariance changes over time. *Psychological Methods, 19*, 113-132. doi:10.1037/a0034525

Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis, 81*(1), 10-23. doi:10.1016/j.csda.2014.07.008

Doove, L. L., Dusseldorp, E., Van Deun, K., & Van Mechelen, I. (in press). A comparison of five recursive partitioning methods to find person subgroups involved in meaningful treatment-subgroup interactions. *Advances in Data Analysis and Classification*. doi:10.1007/s11634-013-0159-x

Dusseldorp, E., Kamphuis, M., & Schuller, A. (in press). Impact of lifestyle factors on caries experience in three different age groups: 9, 15, and 21-year-olds. *Community dentistry and oral epidemiology*.

Eilers, P. H. C. (2013). Discussion: The beauty of expectiles. *Statistical Modelling*, *13*(4), 317-322.

Eilers, P. H. C., & Kroonenberg, P. M. (2014). Modeling and correction of Raman and Rayleigh scatter in fluorescence landscapes. *Chemometrics and Intelligent Laboratory Systems*, *130*, 1-5.

Erbas, Y. I., Ceulemans, E., Pe, M. L., Koval, P., & Kuppens, P. (in press). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition and Emotion*. doi:10.1080/02699931.2013.875890

Eusebi, P., Reitsma, J. B., & Vermunt, J. K. (2014). Latent class bivariate model for the meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology, 14*, 88.

Forero, C. G., Almansa, J., Adroher, N. D., Vermunt J. K., Vilagut, G., De Graaf, R., … Alonso Caballero, J. (2014). Partial likelihood estimation of IRT models with censored lifetime data: An application to men-tal disorders in the ESEMeD surveys. *Psychometrika, 79*, 470-488.

Fried, E. I., Tuerlinckx, F., & Borsboom, D. (2014). Mental health: More than neurobiology. *Nature, 508*, 458-458. doi:10.1038/508458c

Gijtenbeek, M., Bogers, H., Groenenberg, I. A. L., Exalto, N., Willemsen, S. P., Steegers, E. A. P., Eilers, P. H. C., & Steegers-Theunissen, R. P. M. (2014). First trimester size charts of embryonic brain structures. *Human Reproduction, 29*(2), 201-207.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., ... Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö & Evermann (2013). *Organizational Research Methods, 17*(2), 182-209. doi: 10.1177/1094428114526928

Henseler, J., Ringle, C. M., & Sarstedt, M. (in press). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*. doi:10.1007/s11747-014-0403-8

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (in press). On the added value of bootstrap analysis for K-means clustering. *Journal of Classification*.

Hofstetter, H., Dusseldorp, E., Van Empelen, P., & Paulussen, T. W. (2014). A primer on the use of cluster analysis or factor analysis to assess co-occurrence of risk behaviors. *Preventive medicine, 67*, 141-146.

Huijg, J. M., Dusseldorp, E., Gebhardt, W. A., Verheijden, M. W., van der Zouwe, N., Middelkoop, B. J., ... Crone, M. R. (2014). Factors associated with physical therapists' implementation of physical activity interventions in the Netherlands. *Physical Therapy*.

Kadengye, D. T., Ceulemans, E., & Van den Noortgate, W. (2014). A generalized longitudinal mixture IRT model for measuring differential growth in learning environments. *Behavior Research Methods, 46*, 823-840. doi:10.3758/s13428-013-0413-3

Kankaras, M., & Vermunt, J. K. (2014). Simultaneous latent class analysis across groups. In A. C. Michalos (ed.), *Encyclopedia of quality of life and well-being research* (pp. 5969-5974). Dordrecht, The Netherlands: Springer.

Kerver, A. L. A., Carati, L., Eilers, P. H. C., Langezaal, A. C., Kleinrensink, G. J., & Walbeehm, E. T. (2013). An anatomical study of the ECRL and ECRB: Feasibility of developing a preoperative test for evaluating the strength of the individual wrist extensors. *Journal of Plastic Reconstructive and Aesthetic Surgery, 66*(4), 543-550.

Koppenol-Gonzalez, G. V., Bouwmeester, S., & Vermunt, J. K. (2014). Short term memory development: Differences in serial position curves between age groups and latent classes. *Journal of Experimental Child Psychology, 126*, 138–151.

Kusche, D., Kuhnt, K., Ruebesam, K., Rohrer, C., Nierop, A., Jahreis, G., & Baars, T. (in press). Fatty acid profiles and antioxidants of organic and conventional milk from low- and high-input systems during outdoor period. *Journal of the Science of Food and Agriculture*.

Magis, D. (in press). A note on the equivalence between observed and expected information functions with polytomous IRT models. *Journal of Educational and Behavioral Statistics*.

Magis, D., & Facon, B. (2014). deltaPlotR: An R package for differential item functioning analysis with Angoff's delta plot. *Journal of Statistical Software, 59*, 1-19.

Moors, G., Kieruj, N., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology, 44*, 369–399.

Müller, A., Claes, L., Wilderjans, T. F., & de Zwaan, M. (2014). Temperament subtypes in treatment seeking obese individuals: A latent profile analysis. *European Eating Disorders Review, 22*, 260-266. doi:10.1002/erv.2294

Pavlopoulos, D., Fouarge, D., Muffels, R., & Vermunt, J. K. (2014). Who benefits from a job change: The dwarfs or the giants? *European Societies, 16*, 299-319.

Rousseau, S., Grietens, H., Vanderfaeillie, J., Ceulemans, E., Hoppenbrouwers, K., Desoete, A., & Van

Leeuwen, K. (2014). The distinction of 'psychosomatogenic family types' based on parents' self reported questionnaire information: A cluster analysis. *Families, Systems, and Health, 32*, 207-218. doi:10.1037/fsh0000031

Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the emancipation of PLS-SEM: a comment on Rigdon (2012). *Long Range Planning, 47*(3), 154-160. doi:10.1016/j.lrp.2014.02.007

Schouteden, M., Van Deun, K., Wilderjans, T. F., & Van Mechelen, I. (2014). Performing DISCO-SCA to search for distinctive and common information in linked data. *Behavior Research Methods, 46*, 576-587. doi:10.3758/s13428-013-0374-6

Schuppert, H. M., Albers, C. J., Minderaa, R. B., Emmelkamp, P. M., & Nauta, M. H. (in press). Severity of borderline personality symptoms in adolescence: Relationship with maternal parenting stress, maternal psychopathology and rearing styles. *Journal of Personality Disorders*. doi:10.1521/pedi_2104_28_155

Sikorska, K., Lesaffre, E., Groenen, P., & Eilers, P. (2013). Fast mixed models for GWAS with longitudinal data. *Human Heredity, 76*(2), 95.

Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organization Research Methods, 17*, 77-106.

Turner, B., Claes, L., Wilderjans, T. F., Pauwels, E., Dierckx, E., Chapman, A. L., & Schoevaerts, K. (2014). Personality profiles in eating disorders: Further evidence of the clinical utility of examining subtypes based on temperament. *Psychiatry Research, 219*, 157-165. doi:10.1016/j.psychres.2014.04.036

Vermunt, J. K. (2014). Latent class model. In A. C. Michalos (ed.), *Encyclopedia of quality of life and well-being research* (pp. 3509-3515). Dordrecht, The Netherlands: Springer.

Vucic, S., de Vries, E., Eilers, P. H. C., Willemsen, S. P., Kuijpers, M. A. R., Prahl-Andersen, B., ... Ongkosuwito, E. M. (2014). Secular trend of dental development in Dutch children. *American Journal of Physical Anthropology, 155*(1), 91-98.

Wiech, K., Vandekerckhove, J., Zaman, J., Tuerlinckx, F., Vlaeyen, J. W. S., & Tracey, I. (in press). Influence of prior information on pain involves biased perceptual decision-making. *Current Biology*.

Willemsen, S. P., de Ridder, M., Eilers, P. H. C., Hokken-Koelega, A., & Lesaffre, E. (2014). Modeling height for children born small for gestational age treated with growth hormone. *Statistical Methods in Medical Research, 23*(4), 333-345.

Personalia

Mark de Rooij has been appointed Full Professor (‘hoogleraar’) Methodology and Statistics of Psychological Research at the Faculty of Social Sciences of Leiden University.’

From May 2014 on, Jörg Henseler holds the Chair of Product-Market Relations in the Faculty of Engineering Technology, University of Twente. For more information see www.henseler.com.

Meetings

You are invited to the Comprehensive PLS Seminar using SmartPLS 3 that will take place 5-8 November 2014 in the Hamburg University of Technology (TUHH), Hamburg, Germany. From the foundations to the latest advances, this four-day seminar introduces participants to the state-of-the-art of PLS path modeling using the new SmartPLS 3 software. This seminar will also cover recent developments in PLS and variance-based structural equation modeling, including the new PLSc algorithm, the SRMR overall goodness of fit criterion, the novel HTMT criterion for discriminant validity assessment, the new PLS-POS segmentation tool, and a procedure to assess measurement invariance. For more information and registration, see http://november2014.pls-school.com.

Casper Albers will be giving a workshop on "Modelmatige analyse voor beleid op het Bindend StudieAdvies (BSA)" during the "Proof: Truth, hard truth, and statistics" conference of the Dutch Association for Institutional Research, 5-6 November 2014 at Doorn, The Netherlands. For more information, go to www.dair.nl.

The IOPS winter conference will take place December 11-12 at UvA, Amsterdam. For more information, see www.iops.nl.

The 2nd International Symposium on Partial Least Squares Path Modeling (The Conference for PLS Users) will take place 16-19 June 2015, University of Seville, Spain. The International Symposium on Partial Least Squares Path Modeling is the user conference for researchers in business and social sciences who apply and improve partial least squares (PLS) path modeling. The main symposium takes place on 17 and 18 June 2015. There is a pre-conference workshop on 16 June 2015 providing an introduction to PLS path modeling (in English and Spanish), and on 19 June 2015, there will be a post-conference workshop on new developments in the context of PLS path modeling. Several special issues of journals from the business and social sciences are connected to the conference. The submission portal is already open. For more information and registration see www.pls2015.org.

The 7th CARME (Correspondence Analysis and Related Methods) Conference is scheduled to take place 20-23 September, 2015, in Naples, Italy. The objective of this conference is to spotlight the very latest research in correspondence analysis and related methods (CARME) of multidimensional visualization, as well as to discuss future developments. We aim to bring together theoretical and applied researchers in all the areas where correspondence analysis and related methods are currently being used, notably sociology, psychology, education, ecology, archaeology, geology, linguistics, philosophy, genetics, biomedical research, health economics, marketing and management. Interdisciplinary contributions will be particularly welcome. More information is given on the CARME network website (www.carme-n.org) at the conference web page: www.carme-n.org/carme2015/.

ADANCO 1.0 (ADvanced ANalysis of COmposites) is a new software for variance-based structural equation modeling with a graphical user interface. It implements several limited-information estimators, such as partial least squares path modeling (including consistent PLS) or ordinary least squares regression based on sum scores. For more information and download see www.compositemodeling.com.

The European Conference on Data Analysis (ECDA2015) will take place September 2-4 in Colchester UK.

During the next months different workshops will be held:

- Analysis of measurement instruments (IOPS-course, Cees Glas, 16-19 December, Twente university)
- Generalized latent variable modeling (IOPS course, Jeroen Vermunt, 14-15 january, Tilburg University)

Route description

The 25th Anniversary VOC Jubilee Meeting 2014 will take place at abbey Rolduc, Heyendallaan 82, 6464 EP Kerkrade, Netherlands.

By car

From Eindhoven: Follow the A2 motorway in the direction of 'Maastricht/Heerlen'. After the 'Kerensheide' junction, take direction 'Heerlen/Aken (A76)'. After the 'Ten Esschen' junction, take direction 'Kerkrade' (A76) and take the road N281 after the junction 'Bocholtz'. Take the exit 'Kerkrade' and follow 'Kerkrade'. Follow the brown signs indicating 'Rolduc'.

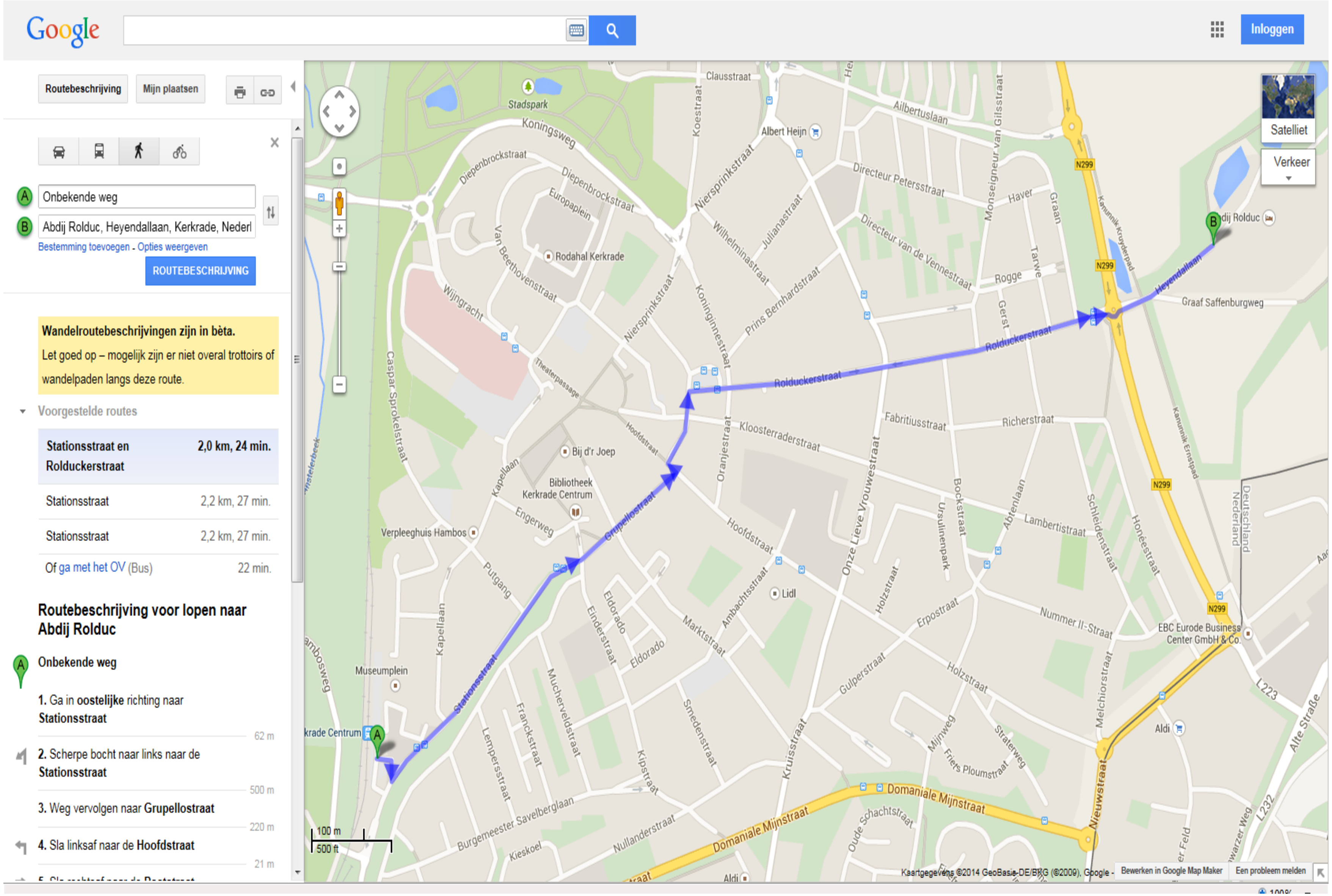
From Aken (Germany): After crossing the border 'Aken/Heerlen (A4/A76)', go direction Kerkrade (N281) at junction 'Bocholtz'. Take the exit 'Kerkrade' and follow 'Kerkrade'. Follow the brown signs indicating 'Rolduc'.

People using GPS: it is safest to enter 'Roderlandbaan' as your final destination ('Heyendallaan' is just a side street of 'Roderlandbaan'). Do never follow the directions for 'Kloosterlindenweg' (because this road is not suitable for cars).

Plan your road: use ANWB-routeplanner at www.anwb.nl/verkeer/routeplanner.

By public transport

The closest railway station is 'Kerkrade Centraal' (another option is to take the 'Herzogenrath' railway station which is located just across the German border). From both railway stations, there are busses to the Rolduc abbey (take line 30 or 41 till busstop 'Rolduc'; to plan your trip to Rolduc by public transport, go to http://9292.nl/). On the map below it is indicated how you can walk from the Kerkrade Centraal railway station to Rolduc (about 2km). For more information, go to www.rolduc.com/NL/routebeschrijving.


Routebeschrijving
Mijn plaatsen
Onbekende weg
Abdij Rolduc, Heyendallaan, Kerkrade, Nederl
Bestemming toevoegen - Opties weergeven
ROUTEBESCHRIJVING
Wandelroutebeschrijvingen zijn in bèta.
Let goed op – mogelijk zijn er niet overal trottoirs of wandelpaden langs deze route.
Voorgestelde routes
Stationsstraat en Rolduckerstraat 2,0 km, 24 min.
Stationsstraat 2,2 km, 27 min.
Stationsstraat 2,2 km, 27 min.
Of ga met het OV (Bus) 22 min.
Routebeschrijving voor lopen naar Abdij Rolduc
Onbekende weg
1. Ga in oostelijke richting naar Stationsstraat
62 m
2. Scherpe bocht naar links naar de Stationsstraat
500 m
3. Weg vervolgen naar Grupellostraat
220 m
4. Sla linksaf naar de Hoofdstraat
21 m
Satelliet
Verkeer
Rolduckerstraat
Heyendallaan
Graaf Saffenburgweg
Stationsstraat
Grupellostraat
Hoofdstraat
Domaniale Mijnstraat
Inloggen

Vereniging voor Ordinatie en Classificatie
Dutch/Flemish Classification Society

25th Anniversary VOC Meeting 2014

November 6-7, 2014

Abbey Rolduc, Kerkrade, The Netherlands

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).

Program

Thursday November 6, 2014

11.00	Arrival	
11.30	Welcome	
11.35	**Daniel Oberski**	Model fit evaluation by sensitivity analysis
12.20	Lunch	
13.30	**Lianne Ippel**	Estimating multi-level models in data-streams
14.15	**Jeroen Jansen**	Revealing the information within Flow Cytometry data using advanced and dedicated Chemometrics
15.00	**Marieke Timmerman**	Unraveling multivariate effects resulting from an experimental design
15.45	Coffee & Tea	
16.15	**Paul Eilers**	Generalized exponential tilting
17.00	**Jelle Goeman**	Two folklores: ridge regression versus the lasso
17.45	Drinks	
18.30	Dinner	

Friday November 7, 2014

9.00	**Willem Heiser**	From preference mapping to preference learning, with an example of a prediction tree for rankings
9.45	**Lieven de Lathauwer**	Tensor decompositions: golden tools for data mining
10.30	Coffee & Tea	
11.00	**Christian Hennig**	Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters
11.45	**Marco Riani**	Robust modern multivariate data analysis and classification
12.30	Lunch	
13.45	**Patrick Groenen**	Some recent biplots approaches
14.30	**Denny Borsboom**	All quiet on the psychometric front? Goals and challenges for 21st century psychometrics
15.15	Closing	

Model fit evaluation by sensitivity analysis

Daniel Oberski
Department of Methodology & Statistics,
Tilburg University, Tilburg, The Netherlands

Latent variable models involve potential "misspecifications": restrictions with a model-based meaning. Examples include zero cross-loadings in factor analysis, zero local dependencies in latent class modeling, and “measurement invariance” or “differential item functioning” in IRT. Such misspecifications can potentially disturb the main purpose of latent variable modeling. This possible disturbance makes model fit evaluation essential, because conclusions are unlikely to be affected when the model fits the data.

In practice, however, the model rarely fits the data. Which should we then stop doing: the modeling or the evaluation of the modeling? Both choices are bad. Abandoning modeling will needlessly throw away information when the misspecifications are irrelevant to the conclusions at hand. Abandoning evaluation, meanwhile, is disastrous when the misspecifications *are* relevant to the conclusions.

I therefore propose a third option. When the model does not fit the data according to a null hypothesis test, I suggest evaluating whether the conclusions could be substantively affected by the misspecification.

To do this, I define a measure based on the likelihood of the restricted model that approximates the change in the parameters of interest if the misspecification were freed: the “EPC-interest”. Examining EPC-interest allows the researcher to free those misspecifications that are “important” while ignoring those that are not. The measure is implemented in the lavaan software for structural equation modeling and the Latent Gold software for latent class analysis.

References

Preprints of the papers can be found at http://daob.nl/publications

Oberski, DL. (2013). “Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models”. Political Analysis. (http://dx.doi.org/10.1093/pan/mpt014)

Oberski, DL & Vermunt, JK (2013). "A Model-Based Approach to Goodness-of-Fit Evaluation in Item Response Theory", Measurement: Interdisciplinary Research & Perspectives, vol. 11, pp. 117-122. (http://dx.doi.org/10.1080/15366367.2013.835195)

Oberski, DL, Vermunt, JK & Moors, G. (submitted) “Ensuring the cross-group comparability of latent variable models for ranking data with the EPC-interest”.

Vermunt, JK & Magidson, J (2013). “Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax”. Belmont, MA: Statistical Innovations Inc. (http://statisticalinnovations.com/technicalsupport/LGtechnical.pdf)

Estimating multi-level models in data-streams

Lianne Ippel
Department of Methodology & Statistics,
Tilburg University, Tilburg, The Netherlands

Recent technological advances in measurement techniques have led to an increase in data-streams in the social sciences: more and more social phenomena can be measured continuously. This leads to datasets which are continuously augmented with new data. Examples of such data-streams include measurements of the web browsing behavior of individuals, or the continuous performance of students in (a series of) examinations. In this presentation we focus on data-streams that have a nested structure. Examples include data-streams containing multiple observations nested within individuals, or measurements of pupils nested within school classes.

Currently, researchers often decide when the process of data gathering is “finished” and start analyzing the (at that point fixed) data set. However more recent analysis methods, known as streaming analyses or online learning, are capable of analyzing the data while these enter the data set. For analysis methods which can be denoted in summation form (e.g., the computation of averages, variances, or even linear regression), the transformation of “offline” methods to “online” methods is rather straightforward. For an (unbalanced) multi-level model, however, this transformation is more complex. This is due to the fact that for a multi-level model no closed form expression exists to fit the model. Therefore, the model is fitted using algorithms such as “Expectation Maximization” algorithm. By iterating through a dataset, this algorithm recursively maximizes the likelihood of the model. However in the case of data-streams such an iterative approach is computationally infeasible: As the stream grows large one has neither the time nor the memory capacity to store all data and iterate through it multiple times.

In this presentation, I introduce a new approach to fit a multi-level model that is applicable to data-streams, using an approximation of the EM algorithm based on Stochastic Gradient Descent. I will show that the performance of this “one-pass” algorithm is competitive to iterative methods of fitting the model.

Revealing the information within Flow Cytometry data using advanced and dedicated Chemometrics

Jeroen Jansen
Analytical Chemistry
Radboud University, Nijmegen, The Netherlands

Flow cytometry is a very well-established platform to analyse comprehensive cell suspensions, such as blood and the algal communities in water. Recent developments in this technology allow the simultaneous analysis of an increasing number of cellular characteristics, either indicated by fluorescent markers or through innate fluorescence in *e.g.* chlorophyll. Monitoring such changes may give unprecedented insight in the system underlying immunology and ecotoxicology, which opens up great possibilies for specific diagnosis of perturbations in the cellular populations. However, the currently used methods for the analysis of Flow Cytometry data cannot handle the increased multivariability, such that dedicated methods are required.

We present several such methods, incorporating ideas from fields including Process Analytical Technology, Image Analysis and Multiset Analysis, to optimally use all information present within Flow Cytometry data. These novel methods allow simultaneously (1) dedicated diagnosis of diseases and ecology, (2) revealing the system of in- and decreasing cellular populations and (3) selective isolation of cells that are specific to the perturbation.

Unraveling multivariate effects resulting from an experimental design

Marieke E. Timmerman[1], Eva Ceulemans[2], Kim De Roover[2], Huub C.J. Hoefsloot[3] & Age K. Smilde[3]
[1]Psychometrics and Statistics, Heymans Institute for Psychological Research, University of Groningen, Groningen, The Netherlands
[2]KU Leuven, Leuven, Belgium;
[3]Biosystems Data Analysis, Faculty of Sciences, University of Amsterdam, Amsterdam, The Netherlands

In many experiments, data are collected on a large number of variables. Typically, the manipulations involved yield differential effects on subsets of variables, and possibly on individuals. The key challenge is to unravel the nature of those differential effects and the associated subsets of variables. An effective method to achieve this goal is to first decompose the observed data matrix into a series of additive effect matrices, according to the experimental design, and second to perform some kind of component modeling on each additive effect matrix of interest. This general method encompasses many different component models, like ASCA (ANOVA-simultaneous component analysis) and clusterwise SCA. In this paper, we provide an overview of the general method, and the specific component models that are of use for modeling an effect matrix. Further, we devote attention to model selection and the issue of scaling. To illustrate the power of the approach, we present analysis results from real-life data, both from metabolomics and psychometrics. We will show that insight can be obtained into multivariate experimental effects, in terms of similarities and differences across individuals. The latter is highly relevant for subtyping.

Generalized exponential tilting

Paul Eilers
Erasmus University Medical Centre,
Rotterdam, the Netherlands

Suppose we are given a density and we ask for the density that is closest to it, but has a different expectation. If we express closeness using the Kullback-Leibler distance, we find that we have to add a linear function to the logarithm of the given density. This is equivalent to multiplication with an exponential function, leading to the name exponential tilting.
We can go further and specify expected values of a set of functions. Then we find that we have to add a linear combination of these functions to the logarithm of the given density. The proper values of the coefficients can be computed efficiently with Newton-Raphson iterations.
Suppose now that we are given a set of densities. We can ask the reverse question: can we find a “mother density” such that the observed densities are (approximately) the results of exponential tilting? The answer is affirmative, but it is of little value if we work with observed data. However, starting from histograms with narrow bins and applying penalized Poisson regression (to get a smooth estimate) we can obtain excellent results. We call this exploratory exponential tilting (EET).
In EET it is assumed that the tilting functions are known. A more ambitious goal is to estimate them from the data in a semi-parametric way. As will be shown, this goal is attainable. It can lead to a parsimonious but accurate description of large sets of densities. Also the patterns in the tilting functions can teach us something about the underlying processes. We call this generalized exponential tilting (GET).
Applications to real data show the usefulness of GET.

This is joint work with Giancarlo Camarda (Institut Nationale d’Études Démographiques, Paris) and Jutta Gampe (Max Plank Institute for Demographic Research, Rostock).

Two folklores: ridge regression versus the lasso

Jelle Goeman
Biostatistics, department for Health Evidence,
Radboud University Medical Center, Nijmegen, The Netherlands

Lasso and ridge regression are two forms of penalized regression that shrink the parameters of the fitted regression model to zero. Both can be used in high-dimensional prediction models, allowing regression models to be fitted even when there are more parameters than observations. The difference between the two is that lasso also returns a sparse model, setting many regression parameters to exactly zero, whereas ridge regression always leaves all covariates in the model. Research in mathematical statistics had uncovered many interesting properties of variants of the lasso, showing in particular some oracle properties. These oracle properties say that asymptotically these variants of the lasso have the same mean squared error for estimating the regression coefficients as an oracle that already knows which regression coefficients are truly non-zero. Because of such properties, which ridge regression does not have, mathematical statisticians tend to claim superiority of lasso over ridge regression. Research in biostatistics, however, tends to show that for many high-dimensional data sets ridge regression has a better predictive potential than the lasso. Researchers also find that the lasso tends to be unstable, selecting very different covariates upon slight perturbations of the data. Because of this, biostatisticians often claim superiority of ridge regression over the lasso, at least where prediction is concerned, and warn against overinterpretation of the results of lasso models. In this talk I will review the arguments on both sides, discussing the usefulness of oracle properties. I will end with practical recommendations.

From preference mapping to preference learning, with an example of a prediction tree for rankings

Willem Heiser
Mathematical Institute and Institute of Psychology,
Leiden University, Leiden, The Netherlands

In the early days of the VOC there used to be an emphasis on methodology for the exploration of multivariate data by clustering, ordering, and mapping the units of analysis so that their structural characteristics could be discovered. Classification in the sense of trying to predict class membership on the basis of multivariate data was regarded as a well-understood task, for which standard methods sufficed and were readily available. In the case of preferences or compositional data, most of us usually first mapped the preference rankings as points or vectors in some Euclidean space by principal components analysis, correspondence analysis or unfolding, and possibly in a next step related them to a relevant set of covariates. An important exception was one of the founders of the VOC, Cajo ter Braak, who already in 1986 had proposed a technique for predicting a multivariate outcome vector (species composition) directly as a function of environmental variables, and developed a whole new methodology around it.
Meanwhile, the machine learning revolution has brought ordination and classification in a mainstream where statistics and computer science meet, and where prediction and predictability became major objectives. Preference learning emerged as a fast growing subfield of machine learning, with fresh perspectives on data collection (internet!), model building, and algorithms. As an example, I discuss a tree-based supervised classification method dealing with preference rankings as the outcome variable.

Tensor decompositions: golden tools for data mining

Lieven de Lathauwer
Group Science, Engineering and Technology
KU Leuven – Kulak
Kortrijk, Belgium
Stadius Center for Dynamical Systems, Signal Processing and Data Analytics
Department of Electrical Engineering (ESAT)
KU Leuven, Leuven, Belgium

Counting from L.R. Tucker's seminal paper on the extension of factor analysis to three-dimensional matrices, tensor-based data analysis celebrates its golden anniversary in 2014. Initially appearing in dedicated fields such as psychometrics, chemometrics and higher-order statistics, decompositions of higher-order tensors are nowadays intensively used in many disciplines. Especially the last 25 years have shown a tremendous increase in tensor-related research. Tensor methods open up remarkable new possibilities in signal processing, array processing, data mining, machine learning, system modelling, scientific computing, statistics, wireless communication, audio and image processing, biomedical applications, bio-informatics, and so forth. On the other hand, these methods have firm roots in multilinear algebra, algebraic geometry, numerical mathematics and optimization.

We give a short general introduction and discuss new trends and perspectives. We pay special attention to new developments in factor analysis, multi-set analysis and big data mining. We also pay attention to current progress in computational issues.

Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters

Christian Hennig
Department of Statistical Science,
University College London, Londen, United Kingdom

Many cluster analysis methods deliver a clustering regardless of whether the dataset is indeed clustered or homogeneous, and need the number of clusters to be fixed in advance. Validation indexes such as the Average Silhouette Width are popular tools to measure the quality of a clustering and to estimate the number of clusters, usually by choosing the number of clusters that optimizes their value. Such indexes can be used for testing the homogeneity hypothesis against a clustering alternative by exploring their distribution, for a given number of clusters fitted by a given clustering method, under a null model formalising homogeneous data. The same approach can be used for assessing the number of clusters by comparing what is expected under the null model with what is observed under different numbers of clusters. Many datasets include some structure such as temporal or spatial autocorrelation that distinguishes them from a plain Gaussian or uniform model, but cannot be interpreted as clustering.

The idea is to specify a null model for data that can be interpreted as homogeneous in the given application, which captures the non-clustering structure in the dataset by some parameters, which are estimated from the data, and then bootstrapping a cluster validity index can be used for testing homogeneity against a clustering alternative and for assessing the number of clusters. Applications will be presented.

Robust modern multivariate data analysis and classification

Marco Riani
Department of Economics, Division of Statistics and Computing,
University of Parma, Italy

Robust methods are little applied although much studied by statisticians. In this paper we sketch what we see as some of the reasons for this failure and suggest a system of interrogating robust analyses, which we call “monitoring”, whereby we consider fits from very robust to highly efficient and follow what happens to aspects of the fitted model. The resulting procedure provides insight into the structure of the data including outliers and the presence of more than one population. Monitoring overcomes the hindrances to the routine adoption of robust methods, being informative about the choice between the various robust procedures.
We also propose some computational improvements of the robust routines and provide a recursive implementation of the so called concentration steps.The output is a set of efficient routines for fast updating of the model parameter estimates, which do not require any data sorting, and fast computation of likelihood contributions, which do not require any inverse matrix or qr decomposition.
Finally, we describe the new routines inside the FSDA (Flexible Statistics Data Analysis) toolbox for MATLAB, which go from the possibility of simulating regression or multivariate mixtures with a prespecified degree of overlap among groups, to the implementation of robust clustering routines based on trimming and eigenvalue constraints, from the possibility of brushing and linking different objects which come out from the application of robust methods, to the implementation of new routines for robust heteroskedastic regression.

Some recent biplots approaches

Patrick Groenen
Econometric Institute
Erasmus University, Rotterdam, The Netherlands

Biplots provide a fantastic tool for visualizing the relations between two entities often with principle components analysis or (multiple) correspondence analysis. Here I discuss several recent developments in this context in which I was involved. The first one is the use of the so-called *area biplot* that can be used as an alternative to every standard projection biplot. Its main difference is that the estimate of the data is given by the area formed by the origin and two points that (Gower, Groenen, & Van de Velden, 2010). The second variety is the *nonlinear biplot* with a distance interpretation: the reconstructed value on a variable of each sample point is obtained by finding the nearest marker point on a nonlinear curve representing the variable (Groenen, Le Roux, Gardner-Lubbe, 2014). The third type of biplot stems from an application of tooth emergence data for school children. The difficulty here lies in the fact that the tooth emergence is not being directly observed, but only intervals are available in which the emergence must have taken place. This *interval-censored* biplot handles this case by providing a biplot and estimating the exact emergence times simultaneously. A fourth method is a two-step approach suitable for linear correspondence analysis biplots that use non-Euclidean distance measures. This approach is developed in collaboration with Michael Greenacre. Then, the first step consists of a constrained multidimensional scaling that only allows for column weights to be estimated. Then, the original data matrix multiplied by their column weights serves as input for a correspondence analysis. Each type of biplot will be explained briefly and an example will be presented.

References

Cecere, S., Leroy, R., Groenen, P.J.F., Lesaffre, E., Declerck, D. (2012). Estimating Emergence Sequences of Permanent Teeth in Flemish Schoolchildren using Interval-Censored Biplots: a graphical display of tooth emergence sequences. *Community Dentistry and Oral Epidemiology*, 40 (suppl 1), 50-56. DOI: 10.1111/j.1600-0528.2011.00666.x

Cecere, S, Groenen, P.J.F. & Lesaffre, E.M.E.H. (2013). The Interval-Censored Biplot. *Journal of Computational and Graphical Statistics*, 22(1), 123-134.

Gower, J.C. and Hand, D.J. (1995). *Biplots*. Monographs on statistics and applied probability, 54. London, U.K.: Chapman & Hall.

Gower, J.C., Groenen, P.J.F. & Van de Velden, M. (2010). Area biplots. *Journal of Computational and Graphical Statistics*, *19*, 46-61.

Gower, J.C., Gardner-Lubbe, S., Le Roux , N. (2011). *Understanding Biplots*. Chichester, Wiley.

Groenen, P.J.F., Le Roux, N.J., and Gardner-Lubbe, S. (2014). Spline-based nonlinear biplots. Advances in Data Analysis and Classification. DOI 10.1007/s11634-014-0179-1

All quiet on the psychometric front?
Goals and challenges for 21st century psychometrics

Denny Borsboom
Institute of Psychology
University of Amsterdam, Amsterdam, The Netherlands

Psychometric tests, and the statistical models used to analyze them, are arguably among the most important fruits harvested in the development of scientific psychology so far. Notwithstanding the importance of these contributions, however, psychometrics has not succeeded in penetrating the intellectual core of psychology. Instead, it has tended to operate as an ancillary discipline, tasked with the development of measurement models that were subsequently used to test psychological theories. These theories are typically developed without the assistance of formalized models. As a result, the current modus operandi in psychology involves a verbal, informal stage of theory formation, and a largely disconnected data-analytic exercise based purely on statistical considerations. In this presentation, I will argue that this division of labor does not actually work very well, and will propose that psychometrics should orient itself towards the development of formalized theories.