Nieuwsbrief no. 43
Oktober 2009

Vereniging voor Ordinatie en Classificatie

Voorzitter: Ron Wehrens, Radboud University Nijmegen, Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen (r.wehrens@science.ru.nl)
Secretaris: Hugo Duivenvoorden, Erasmus Universiteit, Fac. Geneeskunde en Gezondheidswetenschappen, Postbus 1738, 3000 DR Rotterdam (h.duivenvoorden@erasmusmc.nl)
Penningmeester: Berrie Zielman, Algemene rekenkamer, Directie Beleid en Communicatie, Afd. Statistiek, Lange Voorhout 8, 2414 ED Den Haag (a.Zielman@rekenkamer.nl) Postbankrekening 161723 t.n.v. Vereniging voor Ordinatie en Classificatie, Louise Henriettestraat 163, 2595 TP Den Haag.
Redactie: Eva Ceulemans, Centrum voor Methodologie van het Pedagogisch Onderzoek, Katholieke Universiteit Leuven, Vesaliusstraat 2-bus 3762, B-3000 Leuven, België (Eva.Ceulemans@ped.kuleuven.be)
VOC-home page: http://www.voc.ac

From the President

Within a couple of weeks we will be celebrating the 25th anniversary of the VOC with a two-day symposium in Wageningen, at `De Wageningse Berg'. The program is an intriguing mix of classification and ordination in the fields of biology, psychology and sociology, providing a good representation of the width of our field. The opening lecture, providing a philosophy-of-science perspective on classification, will be given by Trudy Dehue, of the University of Groningen. Furthermore, there will be sessions on Biostatistics, Biological and Social Networks, and Psychometrics. The full program, including abstracts, can be found elsewhere in this newsletter. For those of you who have not yet registered, there still is room!

For next year, the board of the VOC is already busy planning the spring meeting - we hope to bring you more news on that during the Jubilee Meeting. Moreover, the VOC will organise a special session on the Statistics Day, April 1st 2010 (no joke...). You will hear more on these meetings in the next newsletter.

See you all in Wageningen!
Ron Wehrens

In this issue:

VOC jubilee meeting: November 12-13, 2009 Hotel de Wageningse Berg

Thursday November 12

11.00 Arrival

11.30 Trudy Dehue: *Other times, other suffering: on the changing classification of despair*

12.15 Lunch

13.30 Korbinian Strimmer: *Gene Sets, False Discovery Rates, and High-Dimensional Prediction*

14.15 Age Smilde: *From metabolomics data to biological networks and back*

14.50 Christian Steglich: *Modelling interdependent actors: Cross-sectional and longitudinal approaches in social network analysis*

15.25 Tea

15.45 Cajo ter Braak: *Spectral decomposition and fuzzy clustering of network data with an application in genetics*

16.20 Marcel Reinders: *Characterisation and inference of biological networks*.

17.05 Drinks

18.30 Dinner

Friday November 13

9.00 Iven Van Mechelen: *Classification models to retrieve the sequential process basis of person-in-context behavior*

9.45 Marian Hickendorff: *Latent variable modeling of strategy choice and strategy accuracy in primary school mathematics*

10.20 Ingmar Visser: *Classification through time: Markov models for time series data*

10.55 Coffee

11.15 Francis Tuerlinckx: *Some applications of stochastic differential equations in psychological research*

11.50 Han van der Maas: *Sudden change and types*

12.25 Lunch

13.45 Ritsert Jansen: *Gene and QTL networks*

14.20 Mark van der Laan: *Targeted Maximum Likelihood Machine Learning: Applications to Causal effect/Variable Importance Assessment and Prediction with Censored Data*

15.10 Closing

Abstracts for the Jubilee Meeting

Trudy Dehue (Theory and History of Psychology, University of Groningen): Other times, other suffering: on the changing classification of despair

This presentation discusses issues of classification from a philosophy of science point of view. It illustrates them at the vexed question of why since the 1990's and particularly in affluent countries the number of people combating depression has rapidly increased.

One standard answer is that not depression itself increased but only its detection and treatment. Wealthy countries are just paying more attention to lack of happiness. Surprisingly perhaps, this explanation is given by biological psychiatry but also by its critics who state that the discipline itself, together with pharmaceutical companies, has aggressively mongered the disease.

Both parties maintain that psychiatry is increasingly treating minor depression too, be it to our benefit according to the first and to our detriment according to the second. I will argue, however, that since its 19th century introduction the classification 'depression' has been in constant flux. In contrast to those who maintain that the disease itself did not increase, I put forward that the disease 'itself' does not exist.

Without denying that people can deeply suffer from feelings of despair and without implying that such feelings did actually increase, I demonstrate that in the course of time the word depression has come to refer to quite different phenomena. In addition, since the 1990's it has acquired the extra meaning of 'lacking independence, entrepreneurship and success'. The presupposition that depression can only vary in severity has actually hindered the search for an explanation of the 'depression-epidemic'. Thus my analysis also exemplifies the crucial role classifications play in both social science and society at large.

Korbinian Strimmer (Institute of Medical Informatics, Statistics, and Epidemology, University of Leipzig): Gene Sets, False Discovery Rates, and High-Dimensional Prediction

My talk will revisit the problem of feature selection in high-dimensional classification with dependent variables, a common task in genomics and proteomics. Exploiting the connection between feature selection and gene ranking in linear discriminant analysis (LDA) I describe an effective framework [1] for high-dimensional prediction that is based on three key elements: James-Stein shrinkage for learning prediction rules, feature ranking by (grouped) correlation-adjusted *t*-scores (cat scores) [2], and feature selection by thresholding and controlling false non-discovery rates (FNDR).

Relative to competing LDA approaches (such as SCRDA) our algorithm is computationally inexpensive and makes practical high-dimensional LDA analysis. Furthermore, we show on experimental data sets and by comparing with the "higher criticism" approach that feature selection by FNDR control is very effective not only for LDA but also for diagonal discriminant analysis. The proposed shrinkage discriminant and variable selection procedure is implemented in the R package "sda" available from the R repository CRAN.

[1] Ahdesmäki, M., and Strimmer, K. (2009) Feature selection in `omics" prediction problems using cat scores and false non-discovery rate control, arXiv:0903.2003.
[2] Zuber, V., and Strimmer, K. (2009). Gene ranking and biomarker discovery under correlation, arXiv:0902.0751.

Age K. Smilde (Biosystems Data Analysis, University of Amsterdam): From metabolomics data to biological networks and back

Systems biology is the study of biology as an integrated system of genetic-, protein-, metabolite-, cellular- and pathway events that are in flux and interdependent. Due to the availability of advanced instrumentation it is possible to generate very complex data sets and a systems biology approach becomes a possibility. A part of systems biology is metabolomics. The amount of data generated in metabolomics studies is huge and the type of data can be very complicated and information rich, especially if it has been collected in a time-resolved way.

A key notion in (systems-) biology is the idea of a biological network. These occur in several forms, i.e. gene regulatory networks and protein-protein interaction networks. One of the challenging areas of systems biology is network inference: to what extend can the biological network be reconstructed from functional genomics data.

A short overview will be given of different networks. More detailed examples will be given of reconstructing metabolic networks and building association networks in the field of endocrinology.

Christian Steglich (ICS/Department of Sociology, University of Groningen): Modelling interdependent actors: Cross-sectional and longitudinal approaches in social network analysis

Social network analysis focuses on relational patterns between social actors, such as friendship among schoolchildren, contracts between firms, or trade between nations. Typical for this research is the non-independence of actors and of dyads. On the one hand, individual properties of the actors can crucially depend on relational partners' properties. As examples one can think of similar substance use patterns among befriended schoolchildren, or of positively correlated profit margins of partnering firms. On the other hand, a relationship between two actors can depend on presence or absence of relationships to third partners. An example would be transitive closure, as expressed in the saying, "friends of my friends are my friends." When analysing observed patterns of relations in a fixed group of actors (a so-called *complete network design*), statistical procedures need to be able to take these interdependencies into account. In the presentation, it will be sketched how this is achieved in two prominent families of statistical network models: exponential random graph models (for single observations of a network structure), and actor-based models for network evolution (for repeated observations). The methods will be illustrated by studying the structure of the gossip network in an organisation, and the co-evolution of substance use patterns with the friendship network in a school cohort.

Cajo J.F. ter Braak (Biometris, Wageningen University and Research Centre, Wageningen): Spectral decomposition and fuzzy clustering of network data with an application in genetics

Spectral decomposition and agglomerative clustering are proven tools in the search for modular structure in large networks. After reviewing some of this work, we provide an attractive interpretation for an existing, but rarely used fuzzy clustering model and apply it to undirected network data in genetics.

Baltz, A., & Kliemann, L. (2005). Spectral analysis. In *Network Analysis* (eds. U. Brandes and T. Erlebach), New York, Springer, pp. 373–416.
Newman, M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69*, e066133.
Newman, M.E.J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 74*, e036104.
ter Braak, C.J.F., Kourmpetis, Y.A.I., Kiers, H.A.L., & Bink, M.C.A.M. (2009). Approximating a similarity matrix by a latent class model: a reappraisal of additive fuzzy clustering. *Computational Statistics and Data Analysis, 53*, 3183-3193.
ter Braak, C.J.F,. *et al.* (2009). *Identity-by-Descent Matrix Decomposition using Latent Ancestral Allele Models.* Manuscript submitted for publication.

Marcel Reinders (Information and Communication Theory Group, Delft University of Technology): Characterisation and inference of biological networks

The cell is a complex system of molecular events that can be represented as biological networks. Today's measurement technologies allow us to capture these

events or their effects at a large scale. Associated with a proper modelling approach, this eventually will give us insights on how such a complex system behaves under different conditions and how it evolves over time. The modelling, however, is still hampered since we only get a partial view on all possible events due to unknown or missing interactions since the measurement abilities are still limited or noisy. Consequently, in most cases a detailed mechanistic modelling of the cell is not possible and we have to rely on coarse modelling approaches, simplifying the molecular events, or, assume uncertainties in situations a more detailed model is possible. This gives room for statistical approaches when modelling biological networks.

In this talk I will walk along two avenues; the characterization and inference of biological networks. Analyzing general properties of biological networks, such as the connectivity, is important to gain insight in the evolutionary principles but also generates characteristics that can be exploited when inferring them from data. Within this context, I will report on our findings with respect to the overrepresentation of robust network structures.

The inference of biological networks is mostly based on association networks. I will report on a number of different modelling techniques of increasing complexity that we have explored over the years. These range from simple linear models to models in which Boolean interactions are incorporated. A recurring theme in these modelling approaches is the balance between model complexity and data fit.

Both topics show that statistical techniques are currently indispensible to advance the field of molecular biology due to the complexity of the cell and the available amount of data.

Iven Van Mechelen (Research Group of Quantitative Psychology and Individual Differences, University of Leuven): Classification models to retrieve the sequential process basis of person-in-context behavior

In many research domains, responses from entities or organisms to stimuli are being studied, often with the conjecture that a number of variables act as mediators between stimuli and responses. As a guiding example, contextualized personality psychology focuses on behaviors as displayed by a person across a broad range of situational contexts, with the assumption that various cognitive-affective process variables (such as interpretations, expectancies, and emotions) act as mediators between situation and behavior. The study of such sequential processes constitutes a major challenge, especially if large numbers of stimuli, potential mediators, and responses are involved. Key research questions at this point include: Which stimuli are functionally equivalent and which stimulus features play a critical role in the response elicitation process? Which responses co-occur/what are the response systems involved? What are the relevant mediators and how does the structure of the mediation sequences look like? Furthermore, when sequential processes are studied for multiple organisms or entities (e.g., several individuals in the contextualized personality case), one may wish to know the structure of between-organism differences at several places in the sequential processes.

In this talk, I will explain how some custom-made unsupervised classification models can be most helpful in addressing the questions as outlined above. The use of these models will be illustrated with psychological data on contextualized anger and self-injurious behavior.

Marian Hickendorff (Leiden University): Latent variable modeling of strategy choice and strategy accuracy in primary school mathematics

Something that is certainly in flux is Dutch mathematics education in primary school. Over the last 3 to 4 decades, not only didactics and educational goals have undergone a large reform, but students' performance on specific areas have changed too – some for the better, but some for the worse. In my research project, I focus on the role that solution strategies play in these changes in achievement.

From a psychometric perspective, analyzing strategy use – choice and accuracy of the different strategies – is a challenging task. Strategies are mostly measured in nominal categories, and are often repeatedly observed within a subject. For example, a student takes a mathematics test of several items, and for each item the strategy used is coded into one of several distinct categories. Due to the repeated observations, various sources of variability play a role in the data analysis: persons and items can affect strategy choice as well as strategy accuracy. In this presentation, I will show how latent variable models (latent class models and explanatory IRT models) can be used to analyze research questions into changes of strategy use, illustrated by two real-data examples on primary school mathematics.

Ingmar Visser (Department of psychology, University of Amsterdam): Classification through time: Markov models for time series data

We present the depmixS4 package which is developed for the R statistical programming language to fit a broad class of (hidden) Markov mixture models. The basic model fit by depmixS4 is the hidden Markov model for longitudinal and time series data, and it also includes latent class and mixture models. The data can be multi-variate, and can be modeled with a range of distributions including all the distributions available in the generalized linear model (glm) function from R, the multivariate normal distribution, and the multinomial logistic distribution. Through the S4 interface, adding other measurement

models and distributions is straightforward. DepmixS4 includes the possibility of incorporating covariates on the initial or prior probabilities of hidden Markov models and mixture type models, as well as on the transition probabilities of the hidden Markov model. We present the program using a number of illustrative applications, ranging from a latent class model with covariates to the analysis of single case multivariate time series data.

Francis Tuerlinckx (Research Group of Quantitative Psychology and Individual Differences, University of Leuven): Some applications of stochastic differential equations in psychological research

The most natural way of modeling change in psychological and physical phenomena is by considering a continuous flow of time. However, in psychological research this is not a common approach, not in the last place because of the methodological difficulties that come with continuous-time modeling. Stochastic differential equations provide a natural framework for thinking about change in continuous time. In this presentation I will give an intuitive account of stochastic differential equations and then show how they can be applied in areas such as response time modeling and longitudinal data analysis.

Han van der Maas (Department of psychology, University of Amsterdam): Sudden change and types

Techniques such as latent class analysis, finite mixture modeling and hidden Markov model, apply the idea of a categorical latent variable. Categorical latent variable techniques have become more advanced and more popular in recent years. There are different theoretical justifications for using categorical latent variables. In this talk we focus on the nonlinear dynamical system perspective on quantitative and qualitative different states in systems. Qualitative different states are often demarcated by phase transitions in the dynamics of complex systems. We will argue that phase transitions imply categorical latent variables and visa versa. We will discuss examples where phase transition research and application of categorical latent structure models go hand in hand.

Ritsert Jansen (Groningen Bioinformatics Centre, University of Groningen): Gene and QTL networks

Genetically different individuals can exhibit large quantitative trait variation. Such variation stems, at least partly, from variations in the DNA. Modern sequencing technologies can reveal the variations in the genome, and genome-wide linkage (GWL) analysis and genome-wide association (GWA) analysis can then link or associate them to variations in the trait of interest. These strategies are increasingly applied to a growing number of organisms, including human, mouse, rat, cattle, pigs, *A. thaliana*, tomato, corn, yeast, *C. elegans*, and *D. melanogaster*, and have pinpointed many quantitative trait loci (QTL) on the genome. To lift the veil that covers the genome-to-phenotype relation we may need to monitor the whole trajectory of intermediate biomolecular phenotypes. Today's molecular technologies, particular microarray and deep sequencing for epigenome and transcriptome, and high resolution mass spectrometry and nuclear magnetic resonance for proteomics and metabolomics, have reached a cost-efficiency level allowing for comprehensive molecular profiling of many samples at multiple biomolecular levels. We here discuss the promises, statistical methods, results and pitfalls for QTL analysis, network reconstruction and causal inference using system-wide data on chromatin modification (epiQTL), gene expression (eQTL), proteins (pQTL), metabolites (mQTL) and classical phenotypes (phQTL) from studies on human, plant and shorebird.

Mark van der Laan (UC Berkeley): Targeted Maximum Likelihood Machine Learning: Applications to Causal effect/Variable Importance Assessment and Prediction with Censored Data

Current statistical practice to assess an effect of an intervention or exposure on an outcome of interest often involves either maximum likelihood estimation for a priori specified regression model, or, manual and/or data adaptive interventions to fine tune a choice of model. In both cases, bias in the point estimates and the estimate of the signal to noise ratio are rampant, causing an epidemic of false claims based on data analyses.

In this talk we present our efforts to construct machine learning algorithms for estimating a causal or adjusted effect that take away the need for specifying regression models, while still providing maximum likelihood based estimators and inference. Two fundamental concepts underlying this methodology are super learning, i.e., the very aggressive use of cross-validation to select optimal combinations of many model fits, and subsequent targeted maximum likelihood estimation to target the fit towards the causal effect of interest. Our maximally unbiased and efficient estimates are accompanied with statistical inference. In addition, multiple testing methods are employed in case one pursues effect estimation across a large set of variables. Our approach naturally integrates machine learning into a formal statistical framework for statistical inference.

We illustrate this method in observational studies for assessing the effect of mutations in the HIV virus that cause resistance to a particular drug regimen. We also illustrate the performance for assessing the effect on the outcome or response to treatment of single nucleotide polymorphisms and gene-expressions in genomic studies, including randomized trials. In particular, we demonstrate the performance of the super learning in prediction of time till event based on censored data.

Book review

Marginal Models: For dependent, clustered, and longitudinal data. W. Bergsma, M. Croon, and J.A. Hagenaars. Springer: New York.

A lot of research within the field of statistics nowadays is devoted to the analysis of clustered categorical data. Examples include longitudinal data, where the interest is in comparing the distributions of the response at different time points, or data gathered in families (groups, schools) where the interest is in comparing the distributions of the parents' response with that of the children. Research questions for such data focus on the marginal distributions (i.e. the distributions at time point 1, time point 2, etc.) instead of the joint distribution. A statistical analysis should take into account the dependencies among these marginal distributions.

Mainstream methodologies for such data are generalized linear mixed models (GLMMs) and generalized estimating equations (GEE). The first (GLMMs) belong to the family of conditional models that explicitly model the dependencies by including subject specific parameters into the model. The second approach (GEE) treats the dependencies as nuisances and replaces them by a simple structure (independence, uniform association) such that estimation becomes `simple' and consistent estimates are obtained for the parameter estimates. Efficiency depends on the adequacy of the approximation of the dependency structure by the simple form. When the dependency structure is misspecified GEE may lead to non-optimal solutions, especially for the standard errors.

This book deals with marginal models that compete with the GEE methodology. Contrary to GEE it makes no assumptions about the dependency structure and uses full Maximum Likelihood estimation with constraints to find parameter estimates. The only assumption made in the book is that of a multinomial sampling scheme of the joint distribution of the responses. This reduction of assumptions is a main step forward compared to the GEE methodology. A drawback, however, is that the presented methodology is for categorical variables only, whereas GEE can use continuous explanatory variables.

The book has seven chapters, starting with a very nice introduction to marginal models, where it is explained what marginal models are and for what type of research questions marginal models are needed. Chapter 2 and 3 provide the main methodology. Chapter 4 gives applications to longitudinal data. Chapter 5 discusses marginal models in relation to causal modeling, i.e. path models for categorical data. Chapter 6 discusses marginal modeling with latent variables. If needed additional methodology is explained in these chapters (4, 5, and 6). The last chapter gives a conclusion, other potential application areas of marginal modeling, and some topics for future research.

The authors write on the back cover that the book is intended for applied researchers. Although the text is well written and the material is introduced using many examples, in my opinion the text is quite technical at some points and probably far too difficult for the intended group. These technical sections are marked such that the reader can skip them, and I think the applied researcher will do so. This skipping poses no problem in following the main arguments of the book. The more technical sections seem to be targeted towards researchers in more statistically oriented areas of science, like psychometrics and sociometrics.

A main advantage of the book is its accompanying website where computer programs are presented for the examples in the book. The authors write that both R and Mathematica code is given on that website, but at the time of writing only R code is available. I don't think that this harms anyone, since R is much more popular than Mathematica. The R code can be used without much difficulty, and without too much trouble can be adapted to `own data situations'.

In conclusion, this book is a very well written book about an important topic in categorical data analysis. The methodology is presented with clear examples from applied research. The book is intended for applied researchers but I think that psychometricians, biometricians, and alike can learn a lot from the material in this book. Therefore, the book certainly deserves a spot on the categorical data analysis bookshelf for anyone that regularly encounters categorical data!

Mark de Rooij

Personalia

From january 1st onwards, **Ron Wehrens** will lead the biostatistics and chemometrics group at the Fondazione Edmund Mach (www.iasma.it).

Pieter Kroonenberg (kroonenb@fsw.leidenuniv.nl) is finalising a Windows-interface for his three-way programs. He is looking for persons who are willing to test the new versions of his analysis programs and the interface itself. The models which at present are attached to the interface are the Parafac, Tucker2, Tucker3 models as well as Basford's model for three-mode mixture method of clustering plus a couple of service programs for three-way data manipulation, and programs to construct joint biplots, calculate residuals and Kiers's programs to rotate components and core arrays.

Agenda

Applying PLS Path Modeling - Introduction, Extensions, Advances
Berlin, Germany, 5-7 November 2009
Instructors: Jörg Henseler & Christian Ringle

Partial Least Squares Path Modeling (PLS) is a powerful method for estimating structural equation models with latent variables and observed indicators. Market researchers and academics from social and business science appreciate the advantages of PLS in cases of small samples, complex models, and formative measurement models.

This seminar is designed for practicing business professionals, full-time faculty, and students who are interested in structural equation modeling using PLS. The seminar covers:

BASICS of PLS
- Essential characteristics of PLS path modeling
- PLS algorithm essentials
- Creating valid PLS path models
- Formative vs. reflective measurement models
- Assessment of PLS path modeling estimates
- Bootstrap

EXTENSIONS of PLS
- Goodness of fit index
- Single vs. multiple item measurement
- Blindfolding
- Second-order constructs
- Common Method Bias
- Mediating & moderating effects
- PLS multigroup analysis

ADVANCES in PLS
- LISREL vs. PLS
- Convergence of the PLS algorithm
- Testing measurement models with TETRAD-PLS
- Cohen's inner model path directionality analysis
- Finite mixture PLS (FIMIX-PLS) segmentation
- PLS-IPMA: Importance-performance matrix analysis
- Investigating non-linear effects

The seminar includes a comprehensive software tutorial and "hands on" applications using the SmartPLS software applications for PLS path modeling (http://www.smartpls.de).

Further information and registration under http://www.pls-school.com/.

Publications

Blasius, J., Eilers, P.H.C., & Gower, J. (2009). Better biplots. *Computational Statistics and Data Analysis, 53*, 3145-3158.

Blasius, J., Greenarcre, M., Groenen, P.J.F. & Velden, M. van de (2009). Special issue on correspondence analysis and related methods. *Computational Statistics and Data Analysis, 53*, 3103-3106.

Capelle, L.G., de Vries, A.C., Haringsma, J., Looman, C.W.N., Nagtzaam, N.M.A., van Dekken, H., ter Borg, F.R., de Vries, A. & Kuipers, E.J. (2009). Serum level of leptin: a potential marker for patients at high risk of gastric cancer? *European Journal of Gastroenterology & Hepatology, 21:3*, A85-A86.

Ceulemans, E., & Kiers, H.A.L. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical & Statistical Psychology, 62*, 601-620.

Eilers, P.H.C., Li, B., & Marx, B.D. (2009). Multivariate calibration with single-index signal regression. *Chemometrics and Intelligent Laboratory Systems, 96*, 196-202.

Frederickx, S., Kuppens, P., Tuerlinckx, F., & Van Mechelen, I. (2009). Desequens: an R-package for the variance decomposition of sequential processes. *Behavior Research Methods, 41*, 524-530.

Geerdink, L.M., Prince, F.H.M., Looman, C.W.N., & van Suijlekom-Smit, L.W.A. (2009). Development of a digital childhood health assessment questionnaire for systematic monitoring of disease activity in daily practice. *Rheumatology, 48*, 958-963.

González, J., Tuerlinckx, F., & De Boeck, P. (2009). Analyzing structural relations in multivariate dyadic binary data. *Applied Multivariate Research, 13*, 77-92.

Graaf, E.S. van de, Felius, J., van Kempen-du Saar, H., Looman, C.W.N., Passchier, J., Kelderman, H., & Simonsz, H.J. (2009). Construct validation of the Amblyopia and Strabismus Questionnaire (A&SQ) by factor analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 247*, 1263.

Jansen, P.W., Tiemeier, H., Looman, C.W.N., Jaddoe, V.W.V. , Hofman, A., Moll, H.A., Steegers, E.A.P., Verhulst, F.C., Mackenbach, J.P., & Raat, H. (2009). Explaining educational inequalities in birthweight: the Generation R Study. *Paediatric and Perinatal Epidemiology, 23*, 216-228.

Kiers, H.A.L., & Harshman, R.A. (2009). An efficient algorithm for PARAFAC with uncorrelated Mode-A components applied to large IxJxK data sets with I>>JK. *Journal of Chemometrics, 23*, 442-447.

Kroonenberg, P.M., Harshman, R.A., & Murakami, T. (2009). Analysing three-way profile data using the Parafac and Tucker3 models illustrated with views on parenting. *Applied Multivariate Research, 13*, 5-41.

Kuchuk, N.O., Pluijm, S.M.F., van Schoor, N.M., Looman, C.W.N., Smit, J.H., & Lips, P. (2009). What is the optimal level of serum 25-hydroxyvitamin D for bone health in older people? Different thresholds for different outcomes. *Bone, 44*, S363-S363.

Kuchuk, N.O., Pluijm, S.M.F., van Schoor, N.M., Looman, C.W.N., Smit, J.H., & Lips, P. (2009). Relationships of serum 25-hydroxyvitamin D to bone mineral density and serum parathyroid hormone and markers of bone turnover in older persons. *Journal of Clinical Endocrinology & Metabolism, 94*, 1244-1250.

Kruijshaar, M.E., Essink-Bot, M.L., Donkers, B., Looman, C.W.N., Siersema, P.D., & Steyerberg, E.W. (2009). A labelled discrete choice experiment adds realism to the choices presented: an example relating to preferences for surveillance tests for Barrett esophagus. *BMC Medical Research Methodology, 9:31*.

Lorenzo-Seva, U., van de Velden, M. & Kiers, H.A.L. (2009). Oblique rotation in correspondence analysis: a step forward in the search of the simplest interpretation. *British Journal of Mathematical and Statistical Psychology, 62*, 583-600.

Nijssen, E.J., & Van Herk, H. (2009). Conjoining international marketing and relationship marketing: Exploring consumers' cross-border service relationships. *Journal of International Marketing, 17*, 91-115.

Redden, R.J., Basford, K.E., Kroonenberg, P.M., Islam, F.M.A., Ellis, R., Wang, S. , Yongsheng, C., & Wang, X. (2009). Variation in adzuki bean (*Vigna angularis)* germplasm grown in China. *Crop Science, 49*, 771-782.

Rosmalen, J.M. van, Groenen, P.J.F., Trejos, J., & Castillo, W. (2009). Optimization strategies for two-mode partitioning. *Journal of Classification, 26*, 155-181.

Rosmalen, J.M. van, Koning, A.J., & Groenen, P.J.F. (2009). Optimal scaling of interaction effects in generalized linear modelling. *Multivariate Behavioral Research, 44*, 59-81.

Roswall, J., Bergman, S., Almqvist-Tangen, G., Alm, B., Niklasson, A., Nierop, A.F.M., & Dahlgren, J. (2009). Population-based waist circumference and waist-to-height ratio reference values in preschool children. *Acta Pædiatrica, 98*, 1632–1636.

Schepers, J., & Hofmans, J. (2009). TwoMP: A MATLAB graphical user interface for two-mode partitioning. *Behavior Research Methods, 41*, 507-514.

Schnabel, S.K., & Eilers, P.H.C. (2009). An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research, 21*, 109-134.

Smilde, A.K. Kiers, H.A.L., Bijlsma, S., Rubingh, C.M., & van Erk, M.J. (2009). Matrix correlations for high-dimensional data: the modified RV-coefficient. *Bioinformatics, 25*, 401–405.

Tendeiro, J.N., ten Berge, J.M.F., & Kiers, H.A.L. (2009). Simplicity transformations for three-way arrays with symmetric slices, and applications to Tucker-3 models with sparse core arrays. *Linear Algebra and its Applications, 430*, 924-940.

Ter Braak, C.J.F., Kourmpetisa, Y., Kiers, H.A.L., & Bink, M.C.A.M. (2009) Approximating a similarity matrix by a latent class model: a reappraisal of additive fuzzy clustering. *Computational Statistics and Data Analysis, 53*, 3183-3193.

Timmerman, M.E., Kiers, H.A.L., Smilde, A.K., Ceulemans, E., & Stouten, J. (2009). Bootstrap confidence intervals in multilevel simultaneous component analysis. *British Journal of Mathematical & Statistical Psychology, 62*, 299-318.

Valdivieso, L., Schoutens, W., & Tuerlinckx, F. (2009). Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type. *Statistical Inference for Stochastic Processes, 12*, 1-19.

van den Berg, R.A., Rubingh, C.M., Westerhuis, J.A., van der Werf, M.J., & Smilde, A.K. (2009). Metabolomics data exploration guided by prior knowledge. *Analytica Chimica Acta, 651*, 173-181.

van den Brink, P.J., den Besten, P.J., bij de Vaate, A., & ter Braak, C.J.F. (2009). Principal response curves technique for the analysis of multivariate biomonitoring time series. *Environmental Monitoring and Assessment, 152*, 271-281.

Van Deun, K., Hoijtink, H., Lieven, T., Van Lommel, L., Schuit, F., & Van Mechelen, I. (2009). Testing the hypothesis of tissue-selectivity: The Intersection-Union Test and a Bayesian approach. *Bioinformatics, 25*, 2588-2594.

Van Deun, K., Smilde, A.K., van der Werf, M.J., Kiers, H.A.L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, *10*, 246.

Van de Wiel, M.A., Brosens, R., Eilers, P., Kumps, C., Meijer, G., Menten, B., Sistermans, E., Speleman, F., Timmerman, M.E., & Ylstra, B. (2009). Smoothing waves in array CGH tumor profiles. *Bioinformatics, 25*, 1099-1104.

Velden, M. van de, Groenen, P.J.F., & Poblome, J. (2009). Seriation by constrained correspondence analysis: A simulation study. *Computational Statistics and Data Analysis, 53*, 3129-3138.

Vries, A.C. de, Capelle, L.G., Looman, C.W.N., van Blankenstein, M., van Grieken, N.C.T., Casparie, M.K., Meijer, G.A., & Kuipers, E.J. (2009). Increased risk of esophageal squamous cell carcinoma in patients with gastric atrophy: Independent of the severity of atrophic changes. *International Journal of Cancer, 124*, 2135-2138.

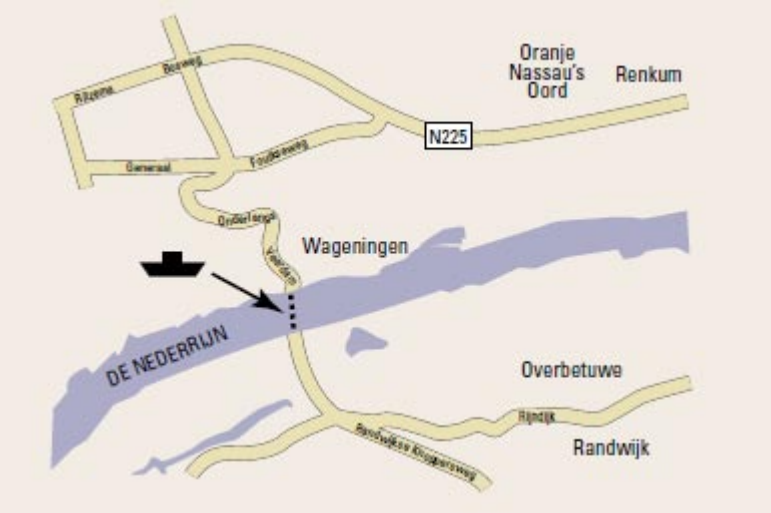
Vrugt J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B. A., Hyman, J.M., & Higdon, D.M. (2009). Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation, 10*, 273-290.

Wilderjans, T.F., Ceulemans, E., & Van Mechelen, I. (2009). Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics and Data Analysis, 53*, 1086-1098.

Wullfaert, J., Van Berkelaer-Onnes, I., Kroonenberg, P.M., Scholte, E., Bhuiyan, Z., & Hennekam, R. (2009). Simultaneous analysis of the behavioural phenotype, physical factors, and parenting stress in people with Cornelia de Lange syndrome. *Journal of Intellectual Disability Research, 53*, 604-619.

Route description

Hotel de Wageningse Berg
Generaal Foulkesweg 96, 6703 DS Wageningen, The Netherlands
Telephone +31 317 49 59 11



By car:
From Utrecht (A12)

- Take exit 24 "Wageningen"
- Take a left turn towards Wageningen (Dr. W. Dreeslaan, N781)
- Follow the road that continues as the Mansholtlaan
- At the roundabouts continue straight ahead
- You are driving on the Diedenweg
- At the traffic lights turn left towards Renkum (N225)
- Then turn right at the roundabout
- After 50 metres on your left you will find the exit towards Hotel de Wageningse Berg

From Arnhem / Apeldoorn (A50) or Nijmegen / Den Bosch (A50)

- Take exit 19 "Renkum"
- You are heading towards Wageningen (N225)
- Turn left at the roundabout
- After 50 metres you will find the exit towards Hotel de Wageningse Berg on your left handside

By public transportation:
From the NS station Arnhem: Bus 50, 251 or 86
From the NS station Ede-Wageningen: Bus 86 (towards Arnhem).
For all buses get off at the stop "De Wageningse Berg".
A taxi will take approximately 15 minutes from NS station Ede-Wageningen.
For more information contact the travel information line at 0900-9292 or go to www.9292ov.nl.