



# Vereniging voor Ordinatie en Classificatie / Dutch - Flemish Classification Society

<b>Chairman:</b>	Mark de Rooij, Universiteit Leiden, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken, Postbus 9555, 2300 RB Leiden, Nederland ( <a href="mailto:rooijm@fsw.leidenuniv.nl">rooijm@fsw.leidenuniv.nl</a> )
<b>Secretary:</b>	Katrijn Van Deun, Universiteit van Tilburg, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken van Onderzoek, Postbus 90153, 5000 LE Tilburg, Nederland ( <a href="mailto:K.VanDeun@uvt.nl">K.VanDeun@uvt.nl</a> )
<b>Treasurer:</b>	Tom Wilderjans, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, Postbus 9555, 2300 RB Leiden, Nederland ( <a href="mailto:t.f.wilderjans@fsw.leidenuniv.nl">t.f.wilderjans@fsw.leidenuniv.nl</a> )
<b>Editor:</b>	Pieter Schoonees, Erasmus Universiteit Rotterdam, Rotterdam School of Management, Department of Marketing Management, Postbus 1738, 3000 DR Rotterdam, Nederland ( <a href="mailto:schoonees@rsm.nl">schoonees@rsm.nl</a> )

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

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

## 7<sup>th</sup> VOC Conference 25 May 2018 Utrecht University

Room Ruppert A, Marinus Ruppertgebouw  
Leuvenlaan 21, 3584 CE Utrecht, Nederland

10:15-11:00	Welcome and registration
10:30-11:00	VOC Annual Members Meeting
11:00-12:00	Keynote address: Ellen Hamaker <i>Intensive longitudinal data in the social and behavioral sciences: Challenges and solutions</i>
12:00-12:45	Contributed paper session 1
12:45-13:45	Lunch
13:45-15:15	Contributed paper session 2
15:15-15:45	Coffee break
15:45-16:45	Keynote address: Johannes Textor <i>How the Immune System Learns to Tell Apart Self from Foreign</i>
16:45-16:55	Announcement of the PhD Presenter Award Winner
16:55-	Closing and Drinks

## In this issue:

Short program 7 <sup>th</sup> VOC Conference	1
Registration details for the 7 <sup>th</sup> VOC Conference	1
From the President	2
Publications	2
Other Meetings	4
Annual Report of the Secretary for the year 2017	5
Minutes of the Members Meeting 2017 (Leiden)	5
Agenda Annual Members Meeting 2018 (Utrecht)	6
Financial Report for 2017	7
Programme 7th VOC Conference	8
Route description	9
Book of Abstracts 7th VOC Conference	11

## Registration details for the 7<sup>th</sup> VOC Conference

Those who would like to join the 7<sup>th</sup> VOC Conference are welcome and are kindly requested to register through our website <http://www.voc.ac> (go to 'meeting'). Details are provided through the website.

## From the President

On May 25 we will have our annual conference in Utrecht. It promises to be an interesting program with two invited speakers and several contributed talks. During the day we will have a very nice cocktail of topics ranging from visualization of compositional data, to clustering of multi-source data, to prediction rule ensembles. The first invited speaker, Ellen Hamaker, will give an overview of recent developments in intensive longitudinal data and discusses some challenges and opportunities. The second invited speaker is Johannes Textor, from Radboud University Medical Center.

All together it promises to be a very nice day. The board is thinking about developing other activities, besides our annual meeting, to further bring to life our society. We might for example start organizing expert meetings on topics of interest to both scientists from different academic fields as well as to business. If you have any thoughts about such or other activities that the VOC could organize please let me know (email address on the front page of this newsletter).

I am looking forward to see you all in Utrecht!

Mark de Rooij

## Publications

Albers, C, Gower, J (2017). Visualising interactions in bi- and triadditive models for three-way tables. *Chemometrics and Intelligent Laboratory Systems*, 167: 238-247.

Albers, C, Gower, J, Kiers, H (2018). Rank properties for centred three-way arrays. F. Mola, C. Conversano, M. Vichi (eds), In: *Classification, (Big) Data Analysis and Statistical Learning, Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, pp. 69-76.

Albers, C, Lakens, D (joint first author) (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74, 187-195.

Andrinopoulou, E. R., Eilers, P. H., Takkenberg, J. J., & Rizopoulos, D. (2017). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*.

Camarda, CG, Eilers, PHC, Gampe, J (2017). Modelling trends in digit preference patterns. *Journal of the Royal Statistical Society: Series C Applied Statistics*, 66 (5).

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.

De Vlaming, R., Okbay, A., Rietveld, C.A., Johannesson, M., Magnusson, P.K., Uitterlinden, A.G., van Rooij, F.J., Hofman, A., Groenen, P.J.F., Thurik, A.R. and Koellinger, P.D. (2017). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS Genetics*, 13(1).

Doove, L. L., Wilderjans, T. F., Calcagni, A., & Van Mechelen, I. (2017). Deriving optimal data-analytic regimes from benchmarking studies. *Computational Statistics and Data Analysis*, 107, 81-91.

Eilers, PHC (2017). Uncommon penalties for common problems. *Journal Of Chemometrics*, 31 (4).

Ernst, A, Albers, C (2017). Regression assumptions in clinical psychology research practice - A systematic review of common misconceptions. *PeerJ*, 5:e3323.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (in press). Detecting treatment-subgroup interactions in clustered data with

generalized linear mixed-effects model trees.  
*Behavior Research Methods*.

Hoendervanger, J.G., Ernst, A., Albers, C., Mobach, M., van Yperen, N. (2018). Individual differences in satisfaction with activity-based work environments. *PLoS ONE*, 13(3): e0193878

Holmgren, A., Niklasson, A., Gelerander, L., Aronson, A. S., Nierop, A. F., & Albertsson-Wikland, K. (2017). Insight into human pubertal growth by applying the QEPS growth model. *BMC Pediatrics*, 17(1), 107.

Lakens, D., Adolphi, F., Albers, C., et al (2018). Justify your alpha. *Nature Human Behaviour* 2, 168-171.

Slob, E. A., Groenen, P. J., Thurik, A. R., & Rietveld, C. A. (2017). A note on the use of Egger regression in Mendelian randomization studies. *International Journal of Epidemiology*, 46(6), 2094-2097.

Tenenhaus, M., Tenenhaus, A., & Groenen, P.J.F. (2017). Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. *Psychometrika*, 82 (3), 737-777

Sikorska, K., Lesaffre, E., Groenen, P. J., Rivadeneira, F., & Eilers, P. H. (2018). Genome-wide Analysis of Large-scale Longitudinal Outcomes using Penalization—GALLOP algorithm. *Scientific Reports*, 8(1), 6815.

Spruit, I. M., Wilderjans, T. F., & van Steenbergen, H. (2018). Hear work after errors: Behavioral adjustment following error commission involves cardiac effort. *Cognitive, Affective, & Behavioral Neuroscience*, 18, 375-388.

Van Herk, H., Schoonees, P. C., Groenen, P. J., & van Rosmalen, J. (2018). Competing for the same value segments? Insight into the volatile Dutch political landscape. *PLoS ONE*, 13(1).

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.

Voncken, L., Albers, C., Timmerman, M (2017). Model selection in continuous test norming with GAMLSS. *Assessment*.

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.

Wilderjans, T. F., Vande Gaer, E., Kiers, H. A. L., Van Mechelen, I., & Ceulemans, E. (2017). Principal covariates clusterwise regression (PCCR): Accounting for multicollinearity and population heterogeneity in hierarchically organized data. *Psychometrika*, 82, 86-111.

## Other Meetings

### Chemometrics: Introduction to Advanced Data Analysis

The course Chemometrics: Introduction to Advanced Data Analysis (Chemometrics-RSS18) which will be held on the 13th-17th of August 2018 in Nijmegen. A week-long programme will cover the major aspects of Chemometrics with an introduction to the most commonly used data analysis method. These include methods for Data pre-processing, exploratory data analysis methods such as Principal Component Analysis (PCA) and regression/classification based on Partial Least Squares (PLS) method. Each subject will be introduced with a lecture, after which you will gain hands-on experience with the methods during (computer) exercises. For more information please do not hesitate to contact [r.folcarelli@science.ru.nl](mailto:r.folcarelli@science.ru.nl). More information is available at:

<http://www.ru.nl/radboudsummerschool/courses/2018/chemometrics-introduction-advanced-data-analysis/>

### Correspondence Analysis and Related Methods 2019

Cape Town/Stellenbosch 3-6 February 2019

(excursion to Cape Peninsula 3 February, conference 4-6 February)

This is a popular series of conferences every four years, starting in Cologne in 1991, where it also took place in 1995 and 1999, then Barcelona in 2003, Rotterdam in 2007, Rennes in 2011 (where 50 years of correspondence analysis was celebrated, in the city here it originated) and Naples in 2015. The conference has attracted 100-200 participants each time and has led to four edited volumes of books and two special journal issues. For the first time, CARME will take place outside Europe, in Stellenbosch & Cape Town, South Africa, organized by the Multivariate Data Analysis Group of the South African Statistical Association. If you have never visited South Africa, this is your chance to come, at the end of the South African summer, in what is the most beautiful region of the country, the winelands of the Cape and the Cape Peninsula where the cold Atlantic and warm Indian oceans meet in a spectacular setting. The website of the conference is: <https://carmesa2019.wixsite.com/conference>

Invited speaker include Cajo ter Braak (Wageningen), Peter Filzmoser (Vienna University of Technology), Frédéric Lebaron (École Normale Supérieure de Cachan), John Gower (via Skype, The Open University), Angela Montanari (Università di Bologna), Riaan de Jongh (North West University, South Africa) and Hervé Abdi (to be confirmed, University of Texas at Dallas).

Make your plans for a trip and conference of a lifetime!



### Data Science, Statistics & Visualisation (DSSV 2018)

9 – 11 July 2018, TU Wien, Austria

The conference of Data Science, Statistics & Visualisation (DSSV 2018) will take place July 9-11 in TU Wien, Austria, see <http://iasc-isi.org/dssv2018/>. We aim to bring together researchers and professionals interested in the interplay of computer science, statistics and visualisation, hence building bridges between these fields. The conference will feature the keynote speaker Jerome H. Friedman (Stanford University).

The deadline for abstract submission is due to 9 April, 2018 subject to no other postponement. The DSSV2018 scientific committee encourages and welcomes you to submit your abstract. We look forward to receive many exciting contributions!



## Annual Report of the Secretary for the year 2017

### 1. Number of members

The VOC started 2017 with 120 members and counted 113 members at the end of 2017. Nine memberships were terminated and there were two new members registered. In 2017, 56 members paid contribution.

### 2. Board

The Board of the VOC was composed as follows in 2017:

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

The Board met once in 2017. The main topic was the organization of the 6<sup>th</sup> VOC conference.

### 3. Activities

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

The sixth VOC Conference took place at Leiden University (the Netherlands) on the 19<sup>th</sup> of May 2017 with a full day program, including eight contributions by VOC members on a range of topics. A keynote contribution was given by Boudewijn Lelieveldt (Leiden University, the Netherlands) on 'Fast and scalable non-linear embedding techniques for high-dimensional data'. 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 (19 May 2017, 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

Non-active and non-paying members are no longer considered members. Meetings were well attended. The board has a new chairman.

4. Financial report of the treasurer on the year 2016

The 7000 euro reserve means the society is in good shape. This will be useful for jubilee meetings. We do not need more. It is now possible to invite speakers from outside.

Dues in arrears are more or less something of the past, except for 100 euro from two people. Some members do not respond after multiple emails. After some time we will need to end their membership. 48 people have paid by now; it should be more like 65. Yearly income is about 1000. We have about 200 euro fixed costs, plus costs of the yearly meeting.

Report of the cash committee, read by Katrijn

Marieke Timmerman and Jeroen Vermunt have controlled the accounts. The accounts are in proper order. The cash committee was pleased with the well-organized administration and the adverse balance of about 5000 euros. Further, the measures taken to facilitate the membership payments have had their positive effects. The cash committee would like to compliment the treasurer Tom Wilderjans.

Most money is currently in the current account, a part of that will be transferred to the savings account: 2000 to 3000 euro.

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

The terms of Ralph Rippe and Katrijn Van Deun end. Candidates for these position can register up to 24 hours before the meeting with the President.

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

Ralph wants to leave the board partly because many people from Leiden are involved. This was agreed upon. Efforts were made to recruit replacements; Hilde Tobi (Wageningen) and Pieter Schoonees (Rotterdam School of Management) were approached. It was agreed that Katrijn wil stay on and that Hilde and Pieter wil join the board.

#### 6. Miscellaneous

Ralph shared that the newsletter was sent with a link to a short survey for members. There were only about 6 replies. Many newsletters had this in the beginning; it could be a good idea to keep doing this in the future to get to know new members.

There was a discussion about potential changes, but nothing concrete was offered.

#### 7. Questions before closure of the meeting

Eeke stated that attendance of 40 is still not many. Earlier the meetings were much more popular. Reaction

of the board: it was even worse before, in terms of members attending it is quite good. Much more is being organized now.

#### 8. Closure of the Members Meeting

Meeting closed.

---

### Agenda for the VOC Annual Members Meeting (25 May 2018, Utrecht)

#### 1. Opening of the Members Meeting

#### 2. Minutes of the Members Meeting 19<sup>th</sup> May 2017

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

#### 3. Annual Report of the Secretary on the year 2017

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

#### 4. Financial report of the treasurer on the year 2017

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 (1)
- Katrijn van Deun, Secretary (2)
- Tom Wilderjans, Treasurer (0)
- Pieter Schoonees, Newsletter Editor (2)
- Jeroen Jansen, board member (0)
- Tobi Hilde, board member (1)

Both Katrijn and Jeroen are willing to serve another term of three years. The board suggests adding Matthijs Warrens as a new board member.

#### 6. Miscellaneous

#### 7. Questions before closure of the meeting

#### 8. Closure of the Members Meeting

---

## Financial Report for 2017

Revenue		Expenditure	
Membership fees (56 paying members)	1400	Transaction costs ING	108,37
Overdue membership fees	720	Hosting website	121
Interest savings account	6,46	Domain registration	52,74
		Drinks board meeting	12,35
<b>Total</b>	<b>2126,46</b>	<b>Total</b>	<b>294,46</b>

Debit		Credit	
Balance ING account	1402,91	Accounts payable	0
Balance savings account	5861,43	Equity	7264,34
<b>Total</b>	<b>7264,34</b>	<b>Total</b>	<b>7264,34</b>

### Notes to the balance sheet

- (1) Contributions from 56 members have been collected in 2017
- (2) In 2017 we collected 720 euros of overdue membership fees
- (3) Compared to the previous year, there is a substantial increase in the equity
- (4) An overview of the evolution of the equity

2017	€ 7.264,34
2016	€ 5.432,34
2015	€ 3.913,66
2014	€ 4.019,92
2013	€ 5.444,46
2012	€ 5.524,70
2011	€ 6.194
2010	€ 7.621
2009	€ 8.189
2008	€ 6.248
2007	€ 5.914
2006	€ 6.869
2005	€ 6.057
2004	€ 5.019
2003	€ 6.795
2002	€ 6.408
2001	€ 5.898
2000	€ 5.731
1999	€ 4.871
1998	€ 5.100

## Programme: 7<sup>th</sup> VOC Meeting

Utrecht, 25 May 2018

Room Ruppert A, Marinus Ruppertgebouw

10:15 - 11:00	Welcome and registration	
10:30 - 11:00	VOC Annual Members Meeting	
11:00 - 12:00	<b>Ellen Hamaker</b> (keynote)	Intensive longitudinal data in the social and behavioral sciences: Challenges and solutions
12:00 - 12:45	Contributed paper session 1	
12:00	<b>Shuai Yuan</b>	The development of a novel method for clustering analysis on multi-source data
12:20	<b>Paul Eilers</b>	New ways to present the ternary diagram for compositional data
12:45 - 13:45	Lunch	
13:45 - 15:15	Contributed paper session 2	
13:45	<b>Jungjeon Choi</b>	How to handle missing values in propensity score analysis: a simulation study
14:05	<b>Matthijs Warrens</b>	On the negative bias of the Gini coefficient due to grouping
14:25	<b>Mark Schoot</b>	Improvement of NIR Calibration models and their value for animal food producers
14:45	<b>Marjolein Fokkema</b>	Prediction rule ensembles: An accurate and interpretable method for prediction
15:15 - 15:45	Coffee break	
15:45 - 16:45	<b>Johannes Textor</b> (keynote)	How the immune system learns to tell apart self from foreign
16:45 - 16:55	Announcement: PhD Presenter Award Winner	
16:55 -	Closing and Drinks	

---

### **Route description to meeting location**

Room Ruppert A, Marinus Ruppertgebouw  
Leuvenlaan 21, 3584 CE Utrecht, Nederland

### **Work activities**

Many construction projects will take place at Utrecht Science Park the Uithof. Please, check the diversions and closures before you visit USP the Uithof.

### **Public transportation**

You can reach the Utrecht Science Park by bus from Utrecht Central station (lines 28 and 12), and from railway stations Amersfoort (line 72), Bunnik (line 242), Bilthoven (line 31), Driebergen-Zeist (line 71, 271 and 371), Utrecht Overvecht (line 30) and Utrecht Lunetten (line 31). Busses ride directly to the Utrecht Science Park from Houten (line 281), IJsselstein (line 283) and Vianen (line 287). Plan your trip via the website of the NS for all train traffic or for all regional public transportation. Exit the bus at the bus-stop: Heidelberglaan.

### **Bike**

The Utrecht Science Park is easily reached by bike from Utrecht (20 minutes from Central Station) and from the surrounding towns and railway stations (10 minutes from Bunnik). You can plan your bike trip using the Cycling Association's bike route planner.

OV-fiets (public transport bike): you can rent a public transport bike from the following stations: Utrecht Central Station, Utrecht Overvecht, Bunnik, Bilthoven and Driebergen-Zeist.

### **Car**

#### **From Amsterdam (A2)**

- A2, exit Utrecht Noord
- N230 follow to the A27
- A27, Rijnsweerd junction towards Utrecht/De Uithof A28, towards Amersfoort / De Uithof
- A28, first exit: exit 2 De Uithof

#### **From Rotterdam/Den Haag or Arnhem (A12)**

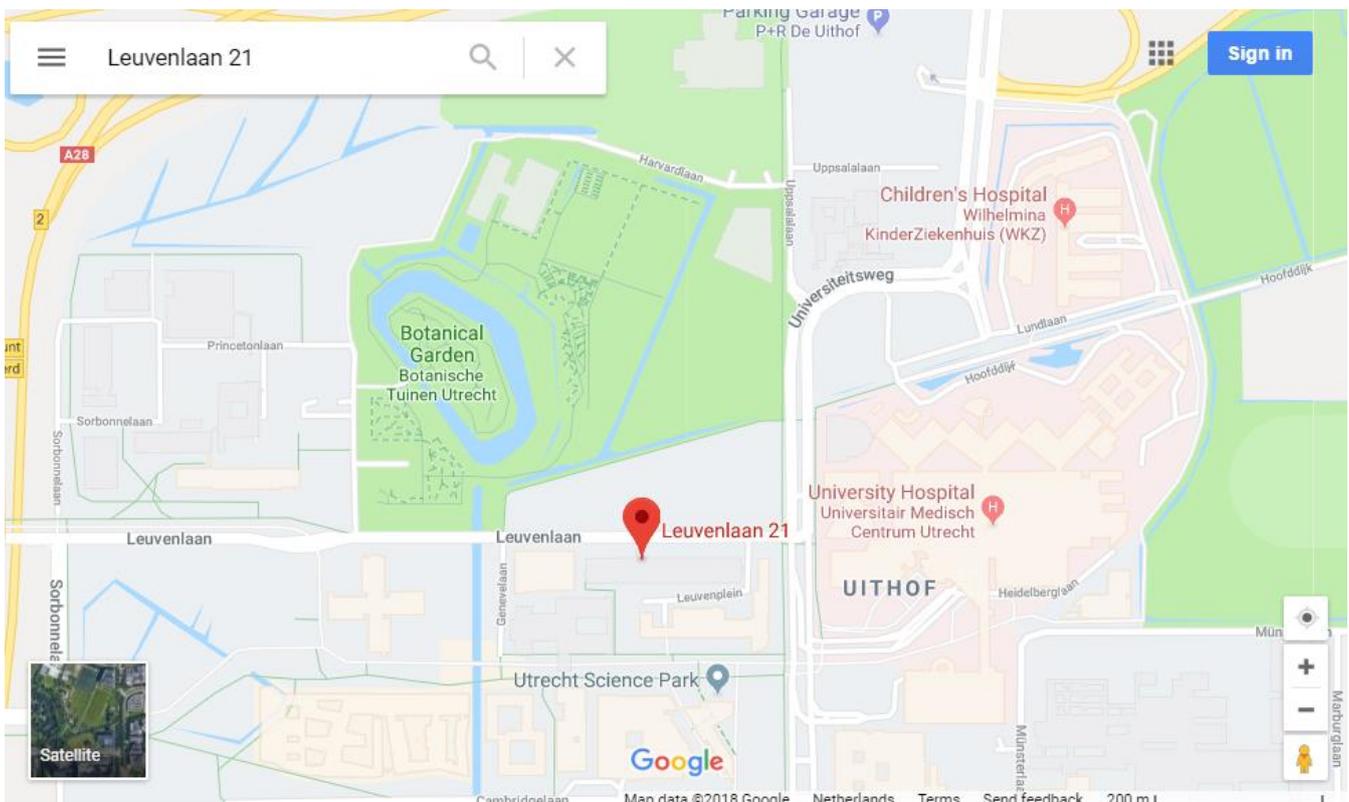
- A12, Lunetten junction towards Amersfoort
- A27, Rijnsweerd junction towards Utrecht/De Uithof A28, towards Amersfoort/De Uithof
- A28, first exit: exit 2 De Uithof

#### **From Amersfoort (A28)**

- A28: exit 2 De Uithof

#### **From Den Bosch and Eindhoven (A2)**

- A2, Oudenrijn junction towards Amersfoort A27, towards Utrecht/Almere
- A27, Lunetten junction towards Amersfoort A28, towards Amersfoort/De Uithof
- A28, first exit: exit 2 De Uithof





**7<sup>th</sup> VOC Conference**  
**25 May 2018**

**Utrecht University, the Netherlands**  
**Marinus Ruppertgebouw, Room Ruppert A**

## **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 in the social and behavioral sciences:  
Challenges and solutions****Ellen Hamaker***Social Sciences Methodology & Statistics  
Faculty of Social and Behavioural Sciences,  
Utrecht University, the Netherlands*

In the social and behavioral sciences there is a strong increase in the number of studies based on intensive longitudinal data. These data consist of a large number of repeated measurements that are closely spaced in time, and they are valuable for getting at the underlying dynamics of a process as it unfolds over time. A possible modeling approach is single subject time series analysis, but this is hard to reconcile with the aim of scientific research to find general laws. In this talk I will discuss two possible solutions to this dilemma: A bottom-up approach in which we analyze data per person, and then look for similarities, and a top-down approach in which we use a multilevel model that allows for quantitative differences between individuals in the parameters that describe the dynamics of a process.

**KEYNOTE****How the Immune System Learns to Tell Apart Self from Foreign****Johannes Textor***Department of Tumor Immunology,  
Radboud University Medical Center,  
Nijmegen, the Netherlands*

Our immune system needs to distinguish foreign intruders from the body's own cells and molecules, a classification problem that to immunologists is known as "self-nonself discrimination". I will explain how the immune system manages to do that, and how we can build a computer model of such immunological learning. Such 'artificial immune systems' can be used for two purposes: (1) to understand how the real immune system behaves, and (2) as a general-purpose 'one-class' classification algorithm. I will show some results from both of these application domains.

## The development of a novel method for clustering analysis on multi-source data

**Shuai Yuan**

*Department of Methodology and Statistics  
Tilburg University,  
Tilburg, the Netherlands*

Psychological studies more and more often yield multi-source data, which consists of novel blocks of data (e.g. genetic data) and traditional blocks of data (e.g. survey data) collected from the same sample. Multi-source data could offer researchers valuable insights into the complex social mechanisms where several influences act together and the differences in such mechanisms between unknown subgroups. Fully revealing the composite mechanisms underlying multi-source data is challenging, however, since the appropriate clustering methods should simultaneously detect both the subgroups and the cluster-specific associations or mechanisms. Additionally, in empirical practice, the methods should also be able to handle high-dimensional datasets which might include huge amounts of irrelevant information. The existing methods could not adequately address the clustering problems because they cannot advise cluster-specific mechanisms and (or) because they cannot deal with high-dimensional data. We present the Cluster-wise Sparse Simultaneous Component Analysis (CSSCA), which groups the observations that possess the same mechanisms while extracts cluster-specific linked variables that collectively suggest these mechanisms. Furthermore, with built-in penalty function, the CSSCA method is also capable of handling high-dimensional data. Two versions of CSSCA algorithms will be discussed here. The old version naturally adapts the cluster-wise approach developed in other component-based methods (e.g. Cluster-wise Simultaneous Component Analysis and Cluster-wise Parafac). Although it has demonstrated convincing performance in simulation studies, it suffers severely from its low computation efficiency. The developing new version departs from an interesting influential statistic (which is derived from the famous PRESS statistic) proposed for PCA, and integrates the solutions obtained from Sparse Simultaneous Component Analysis. We further discuss the results of the simulation studies carried out with the first version of CSSCA method.

## New ways to present the ternary diagram for compositional data

**Paul Eilers**

*Erasmus Medical Centre,  
Rotterdam, the Netherlands*

Compositional data consist of fractions or proportions that add up to 1. An example is election results, where the absolute numbers of votes are not relevant, but their ratios to the total. When there are three components, the ternary diagram is popular and useful to display data. It exploits the fact that for all points in an equilateral triangle the sum of the distances to the sides is constant. When one or two of the proportions are small, the data will be presented by points close to one side of the triangle, or in a corner. That makes it hard to detect systematic patterns. A solution is to present logarithms of ratios of fractions.

This way the regions in the corners can be expanded without bounds. It is convenient to visualize a scatterplot as a density of points, using a gray-scale or color scale. In the transformed domain this is easy. To display the result in the ternary diagram, the density has to be modified by the Jacobian of the transformation. This introduces interesting computational challenges. We derived the transformation when developing a display for log-linear models for election results, based on distances to ideal points. Later it turned out that we had rediscovered an existing transformation, but improved the display.

## How to handle missing values in propensity score analysis: a simulation study

**Jungyeon Choi**

*Leiden University Medical Center  
Leiden, the Netherlands*

### Background

Propensity score analysis is a popular method to control for confounding in observational studies. A challenge in propensity methods are missing values in confounders and several strategies for handling missing values exist.

### Study Design and Setting

In this simulation study, we compared four strategies of handling missing values in propensity score matching and propensity score weighting analyses, namely complete case analysis, missing indicator, multiple imputation and combining multiple imputation and missing indicator, and provide guidance in choosing the optimal strategy. Simulated scenarios varied regarding missing mechanism, presence of effect modification or unmeasured confounding. Additionally, we demonstrated how missingness graphs (m-graph) help clarifying the missing structure.

### Results

Complete case analysis yielded unbiased estimates even when missing was not at random as long as there was no unmeasured confounding and effect modification was absent. Multiple imputation worked well if the data were missing (completely) at random and the imputation model is correctly specified. In the presence of effect modification, more complex imputation models than default options implemented in commonly used statistical software are required. Multiple imputation failed when data were missing not at random. Here in some circumstances, combining multiple imputation and the missing indicator method reduced the bias as the missing indicator variable can be a proxy for unobserved confounding.

### Conclusion

The optimal way to handle missing values in propensity score model depends on the missing data structure and the presence of effect modification. In case of effect modification, default settings of imputation methods will yield biased results even if data are missing at random.

## On the negative bias of the Gini coefficient due to grouping

**Matthijs Warrens**

*University of Groningen,  
Groningen, the Netherlands*

The Gini coefficient is a measure of statistical dispersion that is commonly used as a measure of inequality of income, wealth or opportunity. Empirical research has shown that the coefficient may have a non-negligible downward bias when data are grouped. It is unknown under which grouping conditions the downward bias occurs. It turns out that the Gini coefficient strictly decreases if the data are partitioned into equal sized groups. A sketch of the proof of this property will be presented, as well as various examples that illustrate the property.

## Improvement of NIR Calibration models and their value for animal food producers

**Mark Schoot**

*Nutricontrol and Radboud University,  
Nijmegen, the Netherlands*

Near infrared (NIR) spectroscopy has applications for process control if the performance of calibration models is adequate. We investigated potential improvements for making and updating NIR calibration models, as well as the (potential) value of NIR measurements for animal food (feed) producers. Results show that the method resulting in the best calibration model is highly dependent on the product. Important factors for optimizing calibration models for feed producers are the number of latent variables and the preprocessing strategy. Suitable methods for making and updating prediction models are cross-validation (to determine the number of latent variables) and a design of experiments (to determine a preprocessing strategy). Simulated NIR measurements on feed ingredients in feed factories show great potential for saving costs (4%) and improving end product stability. Better predictions improve these benefits even more, highlighting the importance of good calibration models.

## Prediction rule ensembles: An accurate and interpretable method for prediction

**Marjolein Fokkema**

*Leiden University,  
Leiden, the Netherlands*

Most statistical prediction methods provide a trade-off between accuracy and interpretability. For example, single classification trees are easy to interpret, but provide lower predictive accuracy than methods like tree ensembles. Tree ensembles, on the other hand, are more difficult to interpret, sometimes even termed black boxes. Prediction rule ensembles (PREs) aim to strike a balance between accuracy and interpretability. PREs consist

of a set of prediction rules, which can be depicted as very simple decision trees, and as such are easy to interpret and apply. Friedman and Popescu (2008) proposed the RuleFit algorithm for deriving PREs, which has been found to provide accuracy competitive with random forests and boosted tree ensembles. The R package 'pre' provides an open-source implementation of the RuleFit algorithm, with several extensions and improvements. For example, it supports continuous, factor, count and multivariate response variables and employs unbiased recursive partitioning methods. The latter has been found to not only yield ensembles with equal or better predictive accuracy than RuleFit, but also yields much sparser ensembles. As such, 'pre' provide a highly accurate and interpretable prediction model. In this talk, I will illustrate the methodology and package with real data examples, and discuss current developments, like PREs for multilevel data.