



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

**Chairman:** Mark de Rooij, Universiteit Leiden, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken, Postbus 9555, 2300 RB Leiden, Nederland ([rooijm@fsw.leidenuniv.nl](mailto:rooijm@fsw.leidenuniv.nl))

**Secretary:** Katrijn Van Deun, Universiteit van Tilburg, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken van Onderzoek, Postbus 90153, 5000 LE Tilburg, Nederland ([K.VanDeun@uvt.nl](mailto:K.VanDeun@uvt.nl))

**Treasurer:** Tom Wilderjans, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, Postbus 9555, 2300 RB Leiden, Nederland ([t.f.wilderjans@fsw.leidenuniv.nl](mailto:t.f.wilderjans@fsw.leidenuniv.nl))

**Editor:** Pieter Schoonees, Erasmus Universiteit Rotterdam, Rotterdam School of Management, Department of Marketing Management, Postbus 1738, 3000 DR Rotterdam, Nederland ([schoonees@rsm.nl](mailto:schoonees@rsm.nl))

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

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

## 10<sup>th</sup> VOC Conference

**14 October 2022 Tilburg University**

Warandelaan 2 5037 AB Tilburg

### Room MDZ20 (10:00 – 12:30), Meerkoldreef Building

10:00-10:30 Members meeting  
10:30-11:30 Contributed paper session 1  
11:30-11:45 Break  
11:45-12:30 Keynote address: Bennett Kleinberg

*Using text data for the study of human behaviour:  
potentials and blind spots*

### Room AZ017 (12:30 – 17:00), Academia Building

12:30-13:45 Lunch  
13:45-14:45 Contributed paper session 2  
14:45-15:00 Break  
15:00-16:00 Contributed paper session 3  
16:00-16:15 Break  
16:15-17:00 Contributed paper session 4  
17:00- Closing and drinks

## In this issue:

Short program 10 <sup>th</sup> VOC Conference	1
Registration details for the 10 <sup>th</sup> VOC Conference	1
From the President	2
Publications	2
Book Announcements	4
Annual Report of the Secretary for 2019-2021	4
Minutes of the Members Meeting 2019 (Groningen)	5
Agenda Annual Members Meeting 2022 (Tilburg)	6
Financial Reports for 2019 - 2021	7
Programme 10th VOC Conference	10
Meeting Locations	11
Book of Abstracts 10th VOC Conference	13

## Registration details for the 10<sup>th</sup> VOC Conference

Those who would like to join the 10<sup>th</sup> VOC Conference are welcome and are kindly requested to register through our website <http://www.voc.ac>. Details are provided through the website.

## From the President

Dear VOC members,

Finally, we can have a VOC conference again. As a matter of fact, our last conference was the jubilee meeting, celebrating the 30th birthday of the VOC, in Wageningen. In the middle of organizing the 2020 conference, the corona virus started and postponed our organizing efforts.

It promises to be a very nice conference with a keynote by Bennett Kleinberg, who will present about the use of text data to study aspects of human behaviour. Besides this keynote address, we have 11 contributed talks on a variety of topics. The detailed program for the meeting can be found in this newsletter.

During the conference, we will also have our 'annual' members meeting. As we did not have annual meetings during the corona years, the agenda for the member meeting consists of an overview of the years since 2019. The financial report and the report by the secretary can be found in this newsletter.

During the annual meeting three members will leave the board: Katrijn van Deun, Hilde Tobi, and Jeroen Jansen. I would like to thank all three for their contributions to our society. Now, we are looking for new board members. If you are interested to play an active role in our society, please send an email to either Katrijn van Deun or myself.

I hope to see you in Tilburg this coming Friday.

*Mark de Rooij*

---

## Publications

Achterberg, H.C., de Rooij, J.J., Vernooij, M.W., Ikram, M.A., Niessen, W.J., Eilers, P.H.C., & de Bruijne, M. 2020. Spatially Regularized Shape Analysis of the Hippocampus Using P-Spline Based Shape Regression. *IEEE Journal of Biomedical and Health Informatics*, 24.  
<https://doi.org/10.1109/JBHI.2019.2926789>.

Albertsson-Wikland K, Niklasson A, Gelerander L, Holmgren A, Nierop AFM, Novel type of references for BMI aligned for onset of puberty – using the QEPS growth model, *BMC Pediatrics* 2022; 22:238, <https://doi.org/10.1186/12887-022-03304-3>.

Baldassarre, A., Dusseldorp, E., D'Ambrosio, A., de Rooij, M., & Conversano, C. (2022). The Bradley-Terry Regression Trunk approach for modelling preference data with small trees. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09882-6>.

Bongiorno, R., Bain, P. G., Ryan, M. K., Kroonenberg, P. M., & Leach, C. W. Think leader—think (immoral, power-hungry) man: an expanded

framework for understanding stereotype content and leader-gender bias. *PsyArXiv Preprints* doi = [10.31234/osf.io/p5uya](https://doi.org/10.31234/osf.io/p5uya).

De Rooij, M. Karch, J.D. Fokkema, M. Bakk, Z., Citra Pratiwi, B. & Kelderman, H. (2022). SEM-based out-of-sample predictions. *Structural Equation Modelling: A Multidisciplinary Journal*. <https://doi.org/10.1080/10705511.2022.2061494>.

De Rooij, M. and Weeda, W. (2020). Cross Validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, 3, 248–263. <https://doi.org/10.1177/2515245919898466>.

Eilers, P.H.C. 2020. Comments on: Inference and computation with Generalized Additive Models and their extensions. *TEST*, 29. <https://doi.org/10.1007/s11749-020-00715-1>.

Eilers, P.H.C. and Ruckebusch, C. 2022. Fast and simple super-resolution with single images.

Scientific Reports. <https://doi.org/10.1038/s41598-022-14874-8>.

Elsten, T. and De Rooij, M. (2021). SUBiNN: A stacked uni- and bivariate kNN sparse ensemble. *Advances in Data Analysis and Classification*. <https://doi.org/10.1007/s11634-021-00462-7>.

Frasso, G., & Eilers, P.H.C. 2022. Direct Semi-Parametric Estimation of the State Price Density Implied in Option Prices. *Journal of Business & Economic Statistics*, 40. <https://doi.org/10.1080/07350015.2021.1906686>.

Holmgren A, Niklasson A, Nierop AFM, Butler G, Albertsson-Wikland K, Growth pattern evaluation of the Edinburgh and Gothenburg cohorts by QEPS height model, *Pediatric Research* (2021), [doi.org/10.1038/s41390-021-01790-2](https://doi.org/10.1038/s41390-021-01790-2)

Hugelier, S., Vandenberg, W., Lukes, T., Grussmayer, K.S., Eilers, P.H.C., Dedecker, P., & Ruckebusch, C. 2021. Smoothness correction for better SOFI imaging. *Scientific Reports*. <https://doi.org/10.1038/s41598-021-87164-4>.

Kroonenberg, P. M. (2020). Multiway extensions of the SVD. In Imaizumi, T., Nakayama, A. & Yokoyama, S. (Eds.), *Advanced Research in Behaviormetrics and Data Science* (pp.141-157). Singapore: Springer Nature. ISBN 978-981-15-2699-2.

Kroonenberg, P. M. & Harshman, R. A. (2020). Cattell's parallel proportional profiles. The triumph of a prodigal rotation. *Journal of Chemometrics*, 35,000-000. doi=10.1002/cem.3235

Lombardo, R., Beh, E.J. & Kroonenberg, P.M. (2021). Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables. *Statistical Science*, 36, 542-561. (doi: 10.1214/20-sts814).

Muggeo, V.M.R., Torretta, F., Eilers, P.H.C., Sciandra, M., & Attanasio, M. 2021. Multiple smoothing parameters selection in additive regression quantiles. *Statistical Methods*, 21. <https://doi.org/10.1177/1471082X20929802>.

Rodriguez-Alvarez, M.X., Durban, M., Eilers, P.H.C., Lee, D. & Gonzalez, F. 2022. Multidimensional adaptive p-splines with application to neurons' activity studies. *Biometrics*. <https://doi.org/10.1111/biom.13755>.

Schoonees, P. C., Groenen, P. J. F., & van de Velden, M. (2021). Least-squares bilinear clustering of three-way data. *Advances in Data Analysis and Classification*. <https://doi.org/10.1007/s11634-021-00475-2>.

Van Ginkel, J.R., & Kroonenberg, P.M. (2021). Multiple imputation to balance unbalanced designs for two-way analysis of variance. *Methodology*, 17, 39-57.

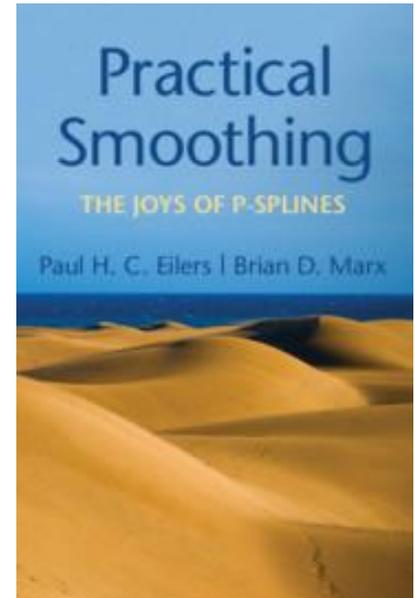
Van Loon, W., de Vos, F., Fokkema, M., Szabo, B., Koini, M., Schmidt, R., & de Rooij, M. (2022). Analyzing hierarchical multi-view MRI data with StaPLR: An application to Alzheimer's disease classification. *Frontiers in neuroscience*, 525. <https://doi.org/10.3389/fnins.2022.830630>.

Yuan, B., Heiser, W.J. and De Rooij, M. (2021). A comparison of two dissimilarity functions for mixed type predictor variables in the  $\delta$ -machine. *Advances in Data Analysis and Classification*. <https://doi.org/10.1007/s11634-021-00463-6>.

## Book Announcements

Eilers, P., & Marx, B. (2021). *Practical Smoothing: The Joys of P-splines*. Cambridge: Cambridge University Press.

This is a practical guide to P-splines, a simple, flexible and powerful tool for smoothing. P-splines combine regression on B-splines with simple, discrete, roughness penalties. They were introduced by the authors in 1996 and have been used in many diverse applications. The regression basis makes it straightforward to handle non-normal data, like in generalized linear models. The authors demonstrate optimal smoothing, using mixed model technology and Bayesian estimation, in addition to classical tools like cross-validation and AIC, covering theory and applications with code in R. Going far beyond simple smoothing, they also show how to use P-splines for regression on signals, varying-coefficient models, quantile and expectile smoothing, and composite links for grouped data. Penalties are the crucial elements of P-splines; with proper modifications they can handle periodic and circular data as well as shape constraints. Combining penalties with tensor products of B-splines extends these attractive properties to multiple dimensions. An appendix offers a systematic comparison to other smoothers.



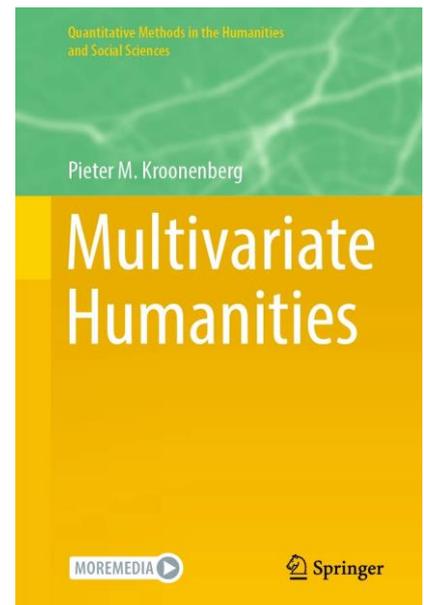
Additional materials are available at <https://psplines.bitbucket.io/>. The JOPS R package is available on CRAN at <https://cran.r-project.org/web/packages/JOPS/index.html>.

Kroonenberg, P.M. (2021). *Multivariate Humanities*. New York: Springer.

This case study-based textbook in multivariate analysis for advanced students in the humanities emphasizes descriptive, exploratory analyses of various types of data sets from a wide range of sub-disciplines, promoting the use of multivariate analysis and illustrating its wide applicability. Fields featured include, but are not limited to, historical agriculture, arts (music and painting), theology, and stylometrics (authorship issues). Most analyses are based on existing data, earlier analysed in published peer-reviewed papers.

Four preliminary methodological and statistical chapters provide general technical background to the case studies. The multivariate statistical methods presented and illustrated include data inspection, several varieties of principal component analysis, correspondence analysis, multidimensional scaling, cluster analysis, regression analysis, discriminant analysis, and three-mode analysis.

The bulk of the text is taken up by 14 case studies that lean heavily on graphical representations of statistical information such as biplots, using descriptive statistical techniques to support substantive conclusions. Each study features a description of the substantive background to the data, followed by discussion of appropriate multivariate techniques, and detailed results interpreted through graphical illustrations. Each study is concluded with a conceptual summary.



More information is available from the publisher at <https://www.springer.com/gp/book/9783030691493>.

## Annual Report of the Secretary for the years 2019 - 2021

### 1. Number of members

The VOC started beginning 2019 with 93 members and counted 83 members at the end of 2021. Thirteen memberships were terminated and there were three new members registered. In 2019, 39 members paid contribution in 2020, 30 members paid contribution, and in 2021 9 members paid contribution (although no fee was asked).

### 2. Board

The Board of the VOC was composed as follows in 2019 - 2021:

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

The Board met a few times in 2019. The main topics were the organization of the 8th and 9th VOC conferences and the VOC jubilee meeting at the occasion of the 30th Anniversary of the VOC.

### 3. Activities

The main activities of the VOC were the 8th VOC conference and the jubilee meeting.

The 8th VOC Conference took place at Groningen University (the Netherlands) on the 5th of April 2019 with a full day program, including nine contributions by VOC members on a range of topics. A keynote contribution was given by Casper Albers (Groningen University, the Netherlands) on 'Intensive longitudinal data analysis in practice – Possibilities and limitations'. The conference had approximately 25 participants.

The Dutch/Flemish classification society (VOC) celebrated its 30th anniversary with a two-day Jubilee Meeting 21 to 22 November 2019 at the Wageningsche Berg, Wageningen, The Netherlands. This meeting was organized by Paul Eilers together with the VOC board. The programme covered exciting topics related to classification and ordination reflecting the width of the field, all presented by distinguished speakers who were invited for a presentation. The speakers were Ron Wehrens, Eva Ceulemans, Peter Bühlmann, Laura

Bringmann, Niël Le Roux, Tim Offermans (best PhD presenter), Aurélie Lemmens, Anders Skrondal, Sophie Swinkels, Michael Greenacre, Tom Snijders, and Anne-Laure Boulesteix. The conference had 47 participants.

Because of the pandemic, the ninth VOC conference did not take place.

### 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 (5 April 2019, Groningen)

### 1. Opening of the Members Meeting

### 2. Minutes of the Members Meeting on 5 April 2019

The minutes of this meeting were approved.

### 3. Annual Report of the Secretary on the year 2018

Tom explained the reduction in membership numbers. After three years of non-payment, membership is terminated. So the decline should be less next year.

Only 40 members paid. Some are still outstanding, some are PhD students (non-paying members). Tom suggested mentioning the winner of the presentation prize in the newsletter / minutes / website.

Additional comments: Jorien was not aware that the group existed. Publicity could be better. An idea is to use IOPS to distribute the newsletter. Mark suggests that the website is not easy to find. Casper suggests at least announcing events on the IOPS mailing list. Marieke Timmerman has contact at other graduate schools. Cajo and Casper suggests trying to get announcements in other newsletters (of other societies).

### 4. Financial report of the treasurer on the year 2018

It was a break-even year. The jubilee conference will bring some expenses due to international speakers. Tom announces it to attendees. We need enough speakers for the jubilee meeting. It is to be located at Wageningen. Matthijs and Tom will soon put information on the

website and start advertising. Cajo suggests that many people from Wageningen would be interested.

The cash committee agreed with Tom's work. Marieke retired as cash committee. Mark thanks Tom for keeping the books balanced. A replacement is needed for Marieke. Casper agrees to do it.

#### 5. *Composition of the Board*

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

- Mark de Rooij, President (0)
- Katrijn Van Deun, Secretary (1)
- Tom Wilderjans, Treasurer (3)
- Pieter Schoonees, Newsletter Editor (1)
- Jeroen Jansen, board member (3)
- Hilde Tobi, board member (1)

Mark's term ends. Nobody reported to Katrijn to step in. Mark is willing to do it for 3 more years. Approved. Mark notes that board is spread well. But contributions from Vlaanderen is desirable.

#### 6. *Miscellaneous*

Mark says the board is still thinking about a potential fall meeting on a theme. It would also be possible to add a theme to the annual meetings. Mark asks for ideas to contact companies and to get them involved.

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

It is asked whether master students can also attend: Mark says yes. Tom says for students, membership is free. Andrei asked whether someone in a company would be welcome. Mark says, indeed. Such persons would also be very welcome to give a talk. Thanks to Matthijs for the website.

#### 8. *Closure of the Members Meeting*

Meeting closed.

## **Agenda for the VOC Annual Members Meeting (14th October 2022, Tilburg University)**

### 1. *Opening of the Members Meeting*

### 2. *Minutes of the Members Meeting 5th April 2019*

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

### 3. *Annual Report of the Secretary on the years 2019-2021*

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

### 4. *Financial report of the treasurer on the years 2019-2021*

The Financial Report is included in this Newsletter (see p. 7 - 9).

### 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 (?)
- Katrijn Van Deun, Secretary (0)
- Tom Wilderjans, Treasurer (2)?
- Pieter Schoonees, Newsletter Editor (1)?
- Matthijs Warrens, Webmaster (2) ?
- Jeroen Jansen, board member (2)?
- Hilde Tobi, board member (1)?

Note that there is some uncertainty about the remaining tenure of the board members due to the suspension of activities during the coronavirus pandemic.

The term of Katrijn Van Deun ends. Jeroen Jansen and Hilde Tobi are also stepping down. Candidates for these positions can register up to 24 hours before the meeting with the Secretary.

### 6. *Miscellaneous*

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

### 8. *Closure of the Members Meeting*

## Financial Report for 2019

<b>Revenue</b>		<b>Expenditure</b>	
membership fees (39 paying members)	975	transaction costs ING	137,38
overdue membership fees	250	update website (with domain name and hosting)	698,17
interest savings account	1,58	VOC meeting Groningen	238,8
fee Jubilee meeting	13495	Jubilee meeting hotel and dinner	12568,3
sponsorship Jubilee meeting	1300	costs speakers	1439,8
		IFCS dues	214,35
<b>Total</b>	<b>16021,58</b>	<b>Total</b>	<b>15296,8</b>
<b>Debit</b>		<b>Credit</b>	
Balance ING account	2111,59	Accounts payable	0
Balance savings account	5867,71	Equity	7979,3
<b>Total</b>	<b>7979,3</b>	<b>Total</b>	<b>7979,3</b>

### Notes to the balance sheet

- (1) Contributions from 39 members have been collected in 2019
- (2) In 2019 we collected 250 euro's of overdue membership fee's
- (3) Compared to the previous year, there is a small increase in the equity
- (4) An overview of the evolution of the equity

2019	€ 7.979,30
2018	€ 7.254,52
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

## Financial Report for 2020

<b>Revenue</b>		<b>Expenditure</b>	
membership fees (30 paying members)	750	transaction costs ING	113,56
overdue membership fees	100	website hosting	124,63
interest savings account	0		
<b>Total</b>	<b>850</b>	<b>Total</b>	<b>238,19</b>
<b>Debet</b>		<b>Credit</b>	
Balance ING account	2723,4	Accounts payable	0
Balance savings account	5867,71	Equity	8591,11
<b>Total</b>	<b>8591,11</b>	<b>Total</b>	<b>8591,11</b>

### Notes to the balance sheet

- (1) Contributions from 30 members have been collected in 2020
- (2) In 2020 we collected 100 euro's of overdue membership fee's
- (3) Compared to the previous year, there is a small increase in the equity
- (4) An overview of the evolution of the equity

2020	€ 8.591,11
2019	€ 7.979,30
2018	€ 7.254,52
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

## Financial Report for 2021

<b>Revenue</b>		<b>Expenditure</b>	
membership fees (9 paying members)	225	transaction costs ING	107,37
overdue membership fees	100	website hosting	124,63
interest savings account	0	IFCS dues	214,88
		chamber of commerce	0,02
<b>Total</b>	<b>325</b>	<b>Total</b>	<b>446,9</b>
<b>Debet</b>		<b>Credit</b>	
Balance ING account	2601,5	Accounts payable	0
Balance savings account	5867,71	Equity	8469,21
<b>Total</b>	<b>8469,21</b>	<b>Total</b>	<b>8469,21</b>

### Notes to the balance sheet

(1) Contributions from 9 members have been collected in 2021 (no fee was asked for 2021 or the 2021 fee included the 2020 fee)

(2) In 2021 we collected 100 euro's of overdue membership fee's

(3) Compared to the previous year, there is a very small decrease in the equity

(4) An overview of the evolution of the equity

2020	€ 8.469,21
2020	€ 8.591,11
2019	€ 7.979,30
2018	€ 7.254,52
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: 10<sup>th</sup> VOC Meeting

## Tilburg, 14 October 2022

**Room MDZ20, Meerkoldreef Building, Meerkoldreef 6, 5042 PN Tilburg**

- 10:00 – 10:30 **Members meeting**
- 10:30 – 11:30 **Contributed paper session 1**
- Marina Khismatullina (EUR) – Multiscale comparison of nonparametric trend curves
- Igor Custodio João\* (VU) – Dynamic nonparametric clustering of multivariate panel data
- Andres Felipe Perez Alonso\* (TIU) – Mixture multigroup SEM for comparing structural relations among many groups
- 11:30 – 11:45 **Break**
- 11:45 – 12:30 **Keynote address**
- Bennett Kleinberg (TIU) – Using text data for the study of human behaviour: potentials and blind spots

**Room AZ017, Academia Building, Prof. Cobbenhagenlaan 205, 5037 DB Tilburg**

- 12:30 – 13:45 **Lunch**
- 13:45 – 14:45 **Contributed paper session 2**
- Tim Offermans (RU) – Improved understanding of industrial process relationships through conditional path modelling with Process PLS
- Rick Willemsen\* (EUR) – A new mixed integer programming approach for inverse correspondence analysis
- Felix Clouth\* (TIU) – Causal inference for latent class analysis
- 14:45 – 15:00 **Break**
- 15:00 – 16:00 **Contributed paper session 3**
- Andrea Jr Carnoli\* (RU) – Accounting for dependence in modeling chemical data
- Max Welz\* (EUR) – Identifying periods of careless responding in rating-scale surveys
- Pieter Kroonenberg (LEI) – What a three-way analysis of cracks can tell us about the origin of paintings
- 16:00 – 16:15 **Break**
- 16:15 – 17:00 **Contributed paper session 4**
- Francisco Bahamonde-Birke (TIU) – Probabilistic market segmentation of transit users in Vancouver, Canada, during the COVID-19 pandemic
- Francisco Souza (RU) – Contextual mixture of experts: Integrating knowledge into predictive modeling
- 17:00 – **Closing and drinks**

\* PhD student presentation

## Meeting Locations

Please note that the meeting will take place in **two different locations** on the Tilburg University campus.

- From **10:00 to 12:30**, we will be in **Room MDZ20** of the **Meerkoldreef Building** (address: Meerkoldreef 6, 5042 PN Tilburg). This building is north of the train station Tilburg Universiteit. It is marked **MD** on the campus map (next page).
- From **12:30 to 17:00**, we will be in room **Room AZ017** of the **Academia Building** (address: Prof. Cobbenhagenlaan 205, 5037 DB Tilburg). This building is located to the south of the train station. It is marked **A** on the campus map (next page).

It is approximately a 10 minute walk between these two buildings. Building information is available at <https://www.tilburguniversity.edu/contact/campus-map>.

The drinks at 17:00 will take place on campus close to the Academia Building.

Detailed information about reaching and navigating the campus is available at <https://www.tilburguniversity.edu/contact/signpost>. Below is a summary.

### Public transportation

Visitors coming to Tilburg University by local train from the direction of Eindhoven, Breda or 's-Hertogenbosch can get off at train station Tilburg Universiteit. The University and the University Sports Center are within walking distance (5 minutes) from the station.

Please note that although there are elevators at this station, they could be temporarily out of service. If you are dependent on these, please read the information provided at <https://www.tilburguniversity.edu/contact/signpost/train> and consider making use of NS Travel Assistance.

### Car parking

Parking is available on campus. The main parking lot is located on Warandelaan on the west side of campus. This can be reached via the Bredaseweg. Visitors are asked to park on the campus as much as possible.

More information is available at <https://www.tilburguniversity.edu/contact/signpost/car/parking>. Parking is available for disabled persons in a variety of locations. Please see the link provided above for more information.

### Campus map

The Meerkoldreef Building (10:00 – 12:30) is marked **MD** at the top right of the map. The Academia Building (12:30 – 17:00) is marked **A** in the middle of the lower half of the map.



### Legenda

 2x oplaadpunt auto

 2x oplaadpunt fiets  
 Alle laadpalen voor fietsen staan in een overdekte fietsenstalling.  
 De elektrische laadpalen bij gebouw Montesquieu (M) bevinden zich in de parkeergarage.



**10<sup>th</sup> VOC Conference**  
**14 October 2022**

**Tilburg University, the Netherlands**  
**10:00 – 12:30: Room MDZ20, Meerkoldreef Building**  
**12:30 – 17:00: Room AZ017, Academia Building**

## **Book of Abstracts**

### **Scope**

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

## KEYNOTE

### Using text data for the study of human behaviour: potentials and blind spots

**Bennett Kleinberg**

*Tilburg University*

Over the past decade, two trends have fundamentally impacted how we do social science research: the availability of "found data" and advancements in statistical natural language processing. Unsurprisingly, text data are increasingly being adopted to study human behaviour through a new lens. These developments bring two exciting questions: what can we learn from text data about human behaviour, and what are the blind spots? This talk shows why text data hold some of the most exciting potentials for psychological research, but it will also make a case for a more cautious approach when using text data as a proxy measure. We will zoom in on recent findings on adversarial machine learning that merit attention for text classification models. We then focus on generative language models, which are currently pushing the boundaries of artificial intelligence research, and highlight the role of psychometric approaches in trying to understand the workings of these powerful models.

## CONTRIBUTED TALKS

### Multiscale comparison of nonparametric trend curves

**Marina Khismatullina**

*Erasmus University Rotterdam*

We develop new econometric methods for the comparison of nonparametric time trends. In many applications, practitioners are interested in whether the observed time series all have the same time trend. Moreover, they would often like to know which trends are different and in which time intervals they differ. We design a multiscale test to formally approach these questions. Specifically, we develop a test which allows to make rigorous confidence statements about which time trends are different and where (that is, in which time intervals) they differ. Based on our multiscale test, we further develop a clustering algorithm which allows to cluster the observed time series into groups with the same trend. We derive asymptotic theory for our test and clustering methods. We illustrate the use of the multiscale test and the clustering algorithm by analysing the house price data.

## Dynamic nonparametric clustering of multivariate panel data

**Igor Custodio João**

*Vrije Universiteit Amsterdam*

We introduce a new dynamic clustering method for multivariate panel data characterized by time-variation in cluster locations and shapes, cluster compositions, and possibly the number of clusters. To avoid overly frequent cluster switching (flickering), we extend standard cross-sectional clustering techniques with a penalty that shrinks observations towards the current center of their previous cluster assignment. This links consecutive cross-sections in the panel together, substantially reduces flickering, and enhances the economic interpretability of the outcome. We choose the shrinkage parameter in a data-driven way and study its misclassification properties theoretically as well as in several challenging simulation settings. The method is illustrated using a multivariate panel of four accounting ratios for 28 large European insurance firms between 2010 and 2020.

## Mixture multigroup SEM for comparing structural relations among many groups

**Andres Felipe Perez Alonso**

*Tilburg University*

Social scientists often examine the relationships between two or more latent variables or constructs, and Structural Equation Modeling (SEM) is the state-of-the-art for doing so. When comparing these structural relations among many groups, they likely differ across the groups. However, it is equally likely that some groups share the same relations, and that clusters of groups emerge in terms of the relations between the latent variables. For validly comparing the latent variables' relations among groups, the measurement of the latent variables should be invariant across the groups (i.e., measurement invariance), whereas often at least some measurement parameters differ across the many groups. Restricting these measurement parameters to be equal across groups, when they are not, causes the structural relations to be estimated incorrectly and thus invalidates their comparison. Therefore, to capture differences and similarities in structural relations while accounting for the reality of measurement non-invariance, we propose Mixture Multigroup SEM (MMG-SEM). MMG-SEM obtains a clustering of groups focused entirely on the structural relations by making them cluster-specific, while allowing for the measurement parameters to be (partially) group-specific. In this way, MMG-SEM disentangles differences in structural relations from differences in measurement parameters. We present an expectation-maximization estimation procedure, built around the R-'lavaan', as well as an evaluation of MMG-SEM's performance in terms of recovering the group-clustering and the cluster-specific parameters.

## Improved understanding of industrial process relationships through conditional path modelling with Process PLS

**Tim Offermans**

*Radboud University*

Understanding how different units of an industrial production plant are operationally related is key to improving production quality and sustainability. Chemometrics has proven indispensable in obtaining such understanding from vast amounts of historical process data. Path modelling is a valuable statistical tool to obtain such information from historical production data, and allows for a high level of structural knowledge on the plant and data to be incorporated in the analysis. Investigating how relationships within a process are affected by multiple production conditions and their interactions can however an even deeper understanding of the plant's daily operation. We therefore propose conditional path modelling as an approach to obtain such improved understanding, demonstrated for a milk protein powder production plant. For this plant we studied how the relationships between different production units and steps are dependent on factors like production line, different seasons and product quality range. We show how the interaction of such factors can be quantified and interpreted in context of daily plant operation. Process PLS, which is a path modelling method recently developed to be optimally suited for analyzing industrial data, is used for this study. Our analysis revealed an augmented insight into the process that can be readily placed in the context of the plant's structure and behavior. Such insights can be vital to identify and improve upon shortcomings in current plant-wide monitoring and control routines.

## A new mixed integer programming approach for inverse correspondence analysis

**Rick Willemsen**

*Erasmus University Rotterdam*

Correspondence analysis (CA) is a dimension reduction technique for categorical data in a two-way contingency matrix. The aim is to optimally depict the relationship between categories for both variables in a low-dimensional representation. We investigate inverse correspondence analysis (ICA), which uses the low-dimensional CA solution to retrieve the original data matrix. We propose a new mixed-integer programming formulation for the ICA problem based on so-called transition formulae, which link the row and column coordinates in a CA solution. We show that this new formulation has better theoretical characteristics than the existing formulation in the ICA literature. In addition, we introduce a novel iterative method, which uses a measure of how well a point is represented in the low-dimensional space. This iterative method is compatible with both the existing and our newly proposed formulation. By incorporating known CA results into our methodology, we are able to solve larger ICA instances. In our computational experiments we always retrieve a unique data matrix, corresponding to the original data. Since the original data set may contain sensitive information, researchers should be careful with disclosing published CA results.

## Causal inference for latent class analysis

**Felix Clouth**

*Tilburg University*

Causal inference techniques such as inverse propensity weighting (IPW) are becoming increasingly popular in medical, social, and behavioral research. When data is collected with an observational study design rather than in a randomized controlled trial, treatment effect estimates will be confounded. However, causal inference provides a toolbox for accounting for these confounding effects and for estimating average treatment effects (ATE) based on observational data. IPW can easily be combined with standard statistical models such as generalized linear models or survival analysis. However, sometimes the outcome of interest is not directly observable and a measurement model is needed, e.g., when analyzing patient-reported outcome measures data. Latent class analysis (LCA) and its extensions have gained popularity for analyzing such data as it explicitly models the multidimensionality of these constructs. Recently, we proposed a stepwise approach to incorporate IPW in LCA. First, the measurement model (latent class model without auxiliary variables) is estimated on the unweighted dataset and individuals are classified. Next, the structural model (the effect of the treatment on class membership) is estimated taking into account the classification errors from the first step and IPW. While this approach works well when the latent class model is correctly specified, differential item functioning (DIF) poses a major challenge to the correct estimation of the ATE. DIF can occur when an auxiliary variable has a direct effect on at least one of the indicator variables. In this case, the assumption that the indicator variables and the auxiliary variables are independent conditional on class membership is violated which can lead to biased estimates of the ATE. In this talk, I will present our analysis strategy for incorporating IPW in LCA and a modification of this approach that allows for the correct estimation of the ATE in the presence of DIF.

## Accounting for dependence in modeling chemical data

**Andrea Jr Carnoli**

*Radboud University*

A common goal in chemistry is to study the relationship between a measured signal and a set of factors. To this end, researchers often use Design of Experiments, an approach where they decide the settings of certain variables a priori and manipulate them across experiments. The relationship between the measured chemical signals and the manipulated factors is then studied by analyzing the data with (Multiple) Linear Regression and/or Analysis of Variance. The validity of this strategy is based on the assumptions that all the experiments are independent, conditional on the settings of the experimental factors. Unfortunately, real-life studies are more complex, with several uncontrollable factors potentially influencing the outcome and causing dependencies in the data. As a result, the assumptions are violated and thus the conclusions derived from the results may not be valid. Mixed effect modeling offers a convenient framework to account for dependence. Here we will provide an introduction to this type of modeling and demonstrate their application with data on the garment contamination from fire-smoke exposure. We will discuss the

advantages of these models in the context of this example, hoping to facilitate their use in the field of chemistry.

## Identifying periods of careless responding in rating-scale surveys

**Max Welz**

*Erasmus University Rotterdam*

Rating-scale datasets collected from surveys are paramount to empirical research. However, due to deficiencies in survey design or lack of motivation, some respondents may not comply with the instructions of the survey's questions. This phenomenon is known as careless responding. Careless responding has been identified as a major threat to the internal validity of survey-based studies and should therefore be screened for. From a robustness perspective, respondents who engage in careless responding constitute (shape) outliers. Notwithstanding, common outlier detection methods from robust statistics are not suitable for rating-scale data due to their discrete and bounded nature as well as their evident non-Gaussianity.

While some methods for detecting careless responding have been proposed in the social science literature, these methods are designed to identify respondents who respond carelessly throughout the survey. However, recent work suggests that the more questions a survey contains, the higher the probability that a large proportion of all respondents will eventually start responding carelessly. Thus, we are interested in identifying when a respondent becomes careless (if at all) rather than detecting respondents who respond carelessly throughout the survey. Correspondingly, we propose a novel method for identifying the periods of carelessness (or a lack thereof) of each respondent.

Our proposed method is based on the deep learning technique of auto-associative neural networks (autoencoders) in combination with response times and longstring indices. Survey datasets typically give rise to a low-dimensional representation due to multiple questions measuring the same variable(s). The autoencoder learns this representation and projects the observed rating-scale responses onto it. The autoencoder then reconstructs the observed responses based on the projections. We demonstrate that, depending on the type of careless behavior, the reconstruction errors of careless responses are either substantially higher or lower than those of regular responses. In other words, the periods of careless responding are reflected by structural breaks in the reconstruction errors. We propose to identify these breaks with a specifically tuned version of Self-Normalization. In addition to the observed responses, our proposed method can be extended with response times, which often provide additional evidence of carelessness.

By means of extensive numerical experiments, we find that our proposed method achieves high reliability in correctly identifying periods of careless responding and discriminates well between careless and regular respondents. Our method seems to perform particularly well in long surveys, which are common in psychology and health sciences, where it is likely that a large proportion of all respondents eventually responds carelessly due to fatigue.

## What a three-way analysis of cracks can tell us about the origin of paintings

**Pieter Kroonenberg**

*Leiden University*

Most paintings acquire cracks over time, but the way these become manifest, varies with the surface on which they were painted, and the materials that were used. In this presentation cracks in paintings from four different traditions in Europe were characterised by both experts and laymen. Crack characteristics were scored based on 50 cm<sup>2</sup> parts of each painting. These parts did not contain information on a painting's subject.

The central research question was whether art-historical traditions from different countries can be distinguished via judging subjective characterisations of the cracks? An additional question is whether experts and laymen judge alike.

## Probabilistic market segmentation of transit users in Vancouver, Canada, during the COVID-19 pandemic

**Francisco Bahamonde-Birke**

*Tilburg University*

Before the COVID-19 pandemic, Metro Vancouver was the North American leader in transit ridership growth, with its most popular routes being severely crowded during peak travel times. As the region gradually recovers from the effects of the pandemic, it is expected that the crowding will also return to the transit system and will have to be tackled. With the focus on the opportunities for effective transit demand management policies to reduce in-vehicle crowding, this study applies a probabilistic market segmentation technique to identify distinct behavioral profiles of transit riders in Metro Vancouver, Canada. The results can be used to develop targeted policy interventions to influence travel patterns, specifically to nudge some riders to travel using less-congested routes or at off-peak times.

The data arises from two waves of a survey distributed in December 2020 and May 2021 by a marketing research company that used hard age and gender quotas to represent the population of Metro Vancouver. This rich dataset includes 2,397 full records with questions covering a range of attitudes towards crowding, safety, flexibility, and actions to avoid crowding on transit, as well as participants' demographics. Participants identified as transit users and had either never stopped riding TransLink, or had not used it since March 2020.

Categorization of respondents into behavioral classes was performed using unobserved latent variables as proposed by Bahamonde-Birke and Ortúzar (2020). Unlike the other methods aiming at a probabilistic segmentation of individuals, this approach aims at the identification of latent classes on the basis of underlying unobserved attitudinal traits. Along these lines, this approach allows identifying how observed characteristics of the individuals affect the likelihood of exhibiting a given underlying trait, which in turn results in a likelihood of belonging to a given population segment.

Principal Component Analysis was employed to explore the underlying relationships between the attitudinal indicators. This information was used as a starting point to specify the latent variables and the structural equations structure within an HDC framework. For the purpose of the analysis, a subset of the indicators was considered to be directly explained by the underlying latent variables, while the remaining indicators were considered to be linked to the underlying latent variables via latent classes allowing for the probabilistic segmentation.

The final model produced six probabilistic classes based on the estimates for two latent variables that accounted for respondents' concerns regarding crowding and safety, as well as personal flexibility to travel to and from work via public transit. Based on the results, a policy framework is developed that suggest that Metro Vancouver's transit agency might already be affecting the travel choices of those riders who are most concerned and flexible through the provision of information on crowding levels. On the other hand, to affect the choices of those riders who are least flexible, it is recommended that the agency develop partnerships with large regional employers.

## Contextual mixture of experts: Integrating knowledge into predictive modeling

**Francisco Souza**

*Radboud University*

In this talk, I want to discuss a new data-driven model devised to integrate expert process knowledge into its structure to increase human-machine synergy in the process industry. The proposed Contextual Mixture of Experts (cMoE) explicitly uses process knowledge along the model learning stage to mold the historical data to represent operators' context related to the process through possibility distributions. This model is evaluated in two real case studies for quality prediction, including a sulfur recovery unit and a polymerization process. The contextual mixture of experts was employed to represent different contexts in both experiments. The results indicate that integrating process knowledge has increased predictive performance while improving interpretability by providing insights into the variables affecting the process's different regimes.