VOC Newsletter 62
May 2023

Vereniging voor Ordinatie en Classificatie / Dutch-Flemish Classification Society

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

Secretary: Matthijs Warrens, Rijksuniversiteit Groningen, Faculteit Gedrags- & Maatschappijwetenschappen, Afdeling Pedagogische en Onderwijswetenschappen, GION onderwijs/onderzoek, Grote Rozenstraat 3, 9712 TG Groningen, Nederland (m.j.warrens@rug.nl)

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

Editor: Pieter Schoonees, Erasmus Universiteit Rotterdam, Erasmus School of Economics, Econometric Institute, Postbus 1738, 3000 DR Rotterdam, Nederland (schoonees@ese.eur.nl)

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

11th VOC Conference

2 June 2023 Erasmus University Rotterdam

Burgemeester Oudlaan 50 3062 PA Rotterdam

Room C1-2, Theil Building, Woudestein Campus

10:00-10:30	Members meeting
10:30-11:15	Contributed paper session 1
11:15-11:45	Break
11:45-12.30	Keynote address: Mark van de Wiel *Co-data learning*
12:30-13:45	Lunch
13:45-14:45	Contributed paper session 2
14:45-15:15	Break
15:15-16:00	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:

Registration details for the 11th VOC Conference

Those who would like to join the 11th VOC Conference are welcome and are kindly requested to register through our website **https://www.voc.ac/meeting**. Details are provided through the website.

From the President

Dear VOC members,

After a successful VOC conference last October in Tilburg, I am looking forward to our meeting in Rotterdam on June 2nd. The meeting is organised by our board members Kathrin Gruber and Pieter Schoonees. It promises to be a very nice conference with a keynote by Mark van de Wiel who is professor in statistics for high-dimensional data at the Amsterdam University Medical Center. Mark van de Wiel will present about co-data learning, a learning process that takes into account complementary information on the variables. Besides this keynote address, we have 10 contributed talks from different disciplines on a variety of topics. The detailed program for the meeting can be found in this newsletter.

During the conference, we will have our annual members meeting, where we discuss the financial report and the report by the secretary (both can be found in this newsletter) and other matters that you would like to discuss. As decided in our previous annual meeting, this year no board members will leave the board. Nevertheless, if you are interested to play an active role in our society, please send me an email.

At this point, I would like to draw your attention to two events in 2024. First, the VOC will celebrate its 35th birthday with a jubilee meeting (probably in October or November) for which we started to make the preparations. We will keep you posted. Second, the International Federation of Classification Societies, of which VOC is a daughter, will have its conference in San Jose, Costa Rica. In short, 2024 will be an exciting year!

Before that, I hope to see you in Rotterdam at our annual conference.

Mark de Rooij

Conference Announcements

The **Channel Network Conference** (CNC) is a biennial conference organized by the International Biometric Society regions of Belgium, France, Great-Britain/Ireland and the Netherlands. This conference aims at gathering Statisticians and Data Scientists to discuss the newest methodology for the analysis of agricultural, biological and medical data. The 2023 CNC will be organized at Wageningen University and Research, from Wednesday August 23rd to Friday August 25th 2023 as an in-person event. More information can be found on the conference website: https://cnc23.sciencesconf.org/!

The joint **DSSV-ECDA 2023** conference takes place in Antwerp from July 5 to 7 2023. Data Science, Statistics & Visualisation (DSSV) and the European Conference on Data Analysis (ECDA) are a joint conference aimed at bringing together researchers and practitioners interested in the interplay of statistics, computer science, and visualization, and to build bridges between these fields for interdisciplinary research. Registration (with a discount for VOC members) is possible at https://iasc-isi.org/dssv-ecda2023/.

The international **Psychoco workshops** aim at bringing together researchers from statistics, psychology, and related disciplines working on modern techniques for the analysis of data from psychology and the social sciences. The next international workshop on psychometric computing (PsychoCo 2024) will take place at Erasmus University Rotterdam from February 29th to March 1st 2024. More information will be available at https://www.psychoco.org/.

The eighteenth biennial meeting of the International Federation of Classification Societies, **IFCS 2024**, will take place in San Jose, Costa Rica, July 2024. Keep an eye on https://ifcs.boku.ac.at/site/doku.php for more information.

Publications

Babapour Mofrad, R., del Campo, M., Peeters, C.F.W., Meeter, L.H.H., Seelaar, H., Koel-Simmelink, M., Ramakers, I.H.G.B., Middelkoop, H.A.M., de Deyn, P.P., Claessen, J.A.H.R., van Swieten, J.C., Bridel, C., Hoozemans, J.J.M., Scheltens, P., van der Flier, W.M., Pijnenburg, Y.A.L., & Teunissen, C.E. (2022). Plasma proteome profiling identifies changes associated to AD but not to FTD. Acta Neuropathologica Communications, 10:148. https://doi.org/10.1186/s40478-022-01458-w

Bartzis, G., Peeters, C.F.W. & van Eeuwijk, F. (2022). psBLUP: incorporating marker proximity for improving genomic prediction accuracy. *Euphytica*, 218: 54. https://doi.org/10.1007/s10681-022-03006-y

del Campo, M., Peeters, C.F.W., Johnson, E.C.B., Vermunt, L., Hok-A-Hin, Y.S., van Nee, M., Chen-Plotkin, A., Irwin, D.J., Hu, W.T., Lah, J.J., Seyfried, N.T., Dammer, E.B., Herradon, G., Meeter, L.H., van Swieten, J., Alcolea, D., Lleó, A., Levey, A.I., Lemstra, A.W., Pijnenburg, Y.A.L., Visser, P.J., Tijms, B.M., van der Flier, W.M., & Teunissen, C.E. (2022). CSF proteome profiling across the Alzheimer's disease spectrum reflects the multifactorial nature of the disease and identifies specific biomarker panels. *Nature Aging*, 2:1040-1053. https://doi.org/10.1038/s43587-022-00300-1

Durieux, J., Rombouts, S. A., de Vos, F., Koini, M., & Wilderjans, T. F. (2022). Clusterwise Independent Component Analysis (C-ICA): Using fMRI resting state networks to cluster subjects and find neurofunctional subtypes. *Journal of Neuroscience Methods*, 382, 109718. https://doi.org/10.1016/j.jneumeth.2022.109718

Münch, M.M., van de Wiel, M.A., van der Vaart, A.W., & Peeters, C.F.W. (2022). Semi-supervised empirical Bayes group-regularized factor regression. *Biometrical Journal*, 64:1289-1306. https://doi.org/10.1002/bimj.202100105

Peeters, C.F.W., Bilgrau, A.E., & van Wieringen, W.N. (2022). rags2ridges: A One-Stop-ℓ2-Shop for Graphical Modeling of High-Dimensional Precision Matrices. *Journal of Statistical Software*, 102(4): 1–32. https://doi.org/10.18637/jss.v102.i04

Peeters, C.F.W., & van Wieringen, W.N. (2022). Ridge learning of static and dynamic graphical models using the rags2ridges and ragt2ridges packages. *Biometric Bulletin*, 39(2):10-13. https://github.com/softwarecorner/2022-39-2

Scarioni, M., Gami-Patel, P., Peeters, C.F.W., de Koning, F., Seelaar, H., Mol, M.O., van Swieten, J.C., Netherlands Brain Bank, Rozemuller, A.J.M., Hoozemans, J.J.M., Pijnenburg, Y.A.L., & Dijkstra, A.D. (2023). Psychiatric symptoms of frontotemporal dementia and subcortical (co-)pathology burden: New insights. *Brain*, 146:307-320. https://doi.org/10.1093/brain/awac043

Annual Report of the Secretary for the years 2022

1. Number of members

The VOC started 2022 with 83 members and counted 101 members at the end of 2022. One membership was terminated and there were 19 new members registered. In 2022 no fee was asked.

2. Board

From the member meeting on October 14, 2022 onwards, the board of the VOC is composed as follows:

Mark de Rooij	President
Matthijs Warrens	Secretary
Tom Wilderjans	Treasurer
Pieter Schoonees	Newsletter Editor
Carel Peeters	Webmaster
Kathrin Gruber	Member
Gerjen Tinnevelt	Member

The board met two times in 2022. The main topics were the organization of the 10th VOC conference in Tilburg (October 14, 2022) and the 11th VOC conference in Rotterdam (June 2, 2023).

3. Activities

The main activity of the VOC in 2022 was the 10th VOC conference. It was the first in person conference after a long period of social distancing and online meetings.

The 10th VOC Conference took place at Tilburg University (the Netherlands) on October 14, 2022 with a full day program, including eleven contributions on a range of topics. A keynote contribution was given by Bennett Kleinberg (Tilburg University, the Netherlands) on 'Using text data for the study of human behaviour: potentials and blind spots'. The conference had approximately 36 participants.

4. Publicity

The newsletter appeared once (number 61) in 2022. The VOC conferences were also announced to non-VOC members, using various channels, including the IFCS newsletter.

Minutes from the VOC Annual Members Meeting (14 October 2022, Tilburg)

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 years 2019-2021*

Katrijn mentions that she is proud that the scientific society is doing as well as it is (surprisingly good).

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

Mark mentions that we did not ask contributions in 2020 and 2021 because we did not do anything. Cash committee approved the bookkeeping, thanking Tom.

New members of cash committee are Michel van de Velden and Jeroen Vermunt.

Decharge of treasurer without complaints.

5. *Composition of the Board*

Mark mentions that the tenure of the board members is not clear. Katrijn was on the board since 2010 but is stepping down now. She is still on IFCS board. Thank you, Katrijn!

Tom, on the board since, 2011 remains as treasurer. Pieter and Matthijs, on the board since 2017, remains. Matthijs takes over from Katrijn as Secretary. Both Hilde and Jeroen, on the board since 2017, will leave. Mark, on the board since 2016, will remain (having taken over from Jeroen).

Three new potential board members are presented. Gerjen speaks about himself: from Radboud University, chemometrics, group of Jeroen Jansen. PhD in 2018. Flow cytometry data analysis. Gerjen joins the board.

Kathrin speaks about herself: Erasmus School of Economics, Bayesian and approximate Bayesian methods. Currently, she also interested in connections with neural networks. Kathrin joins the board.

Third, Carel Peeters is introduced by Mark: Wageningen biometrics group (WUR involved with the VOC since its founding in 1989). Carel works on a variety of topics including classification in higher dimensional data. A vote is held for Carel, in absentia. Carel joins the board.

Mark plans to discuss remaining terms in the next board meeting.

Tom joined the meeting, but the finance discussion was already done. Tom is thanked for taking care of treasury duties.

6. *Miscellaneous*

Advice and/or ideas were sought for how to proceed with the Society in the future. Katrijn mentions that the cost for arranging has increased. We may have to run meetings which is more expensive than membership income.

Mark states that in his opinion the Society should go back to having meetings in May. A plan to raise membership fees can be constructed then, or a fee can be charged for the meeting. This matter is to be discussed at the board meeting.

7. *Questions before closure of the meeting*

Pieter mentions that there are several universities which are unfortunately not represented at the current meeting. He makes a general call that the Society should try to involve other groups more.

Mark states that he agrees, and adds that he wants to keep it interdisciplinary. The present members are asked to share thoughts with the board later on.

8. *Closure of the Members Meeting*

Meeting closed.

Agenda for the VOC Annual Members Meeting (2 June 2023, Erasmus University Rotterdam)

1. *Opening of the Members Meeting*

2. *Minutes of the Members Meeting 14 October 2022*

The Minutes of this Meeting are included in this Newsletter (see p. 4).

3. *Annual Report of the Secretary on the year 2022*

The Annual Report is included in this Newsletter (see p. 4).

4. *Financial report of the treasurer on the year 2022*

The Financial Report is included in this Newsletter (see p. 6).

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)
- Tom Wilderjans, Treasurer (3)
- Matthijs Warrens, Secretary (1)
- Pieter Schoonees, Newsletter Editor (3)
- Carel Peeters, Webmaster (2)
- Gerjen Tinnevelt, Member (2)
- Kathrin Gruber, Member (2)

6. *Miscellaneous*

7. *Questions before closure of the meeting*

8. *Closure of the Members Meeting*

Financial Report for 2022

Revenue		Expenditure	
membership fees (8 paying members)	200	Transaction costs ING	160.99
overdue membership fees	50	Website hosting	131.89
interest savings account	0	VOC meeting Tilburg	754.36
		Chamber of commerce	2.36
Total	250	Total	1049.6

Debit		Credit	
Balance ING account	1791.9	Accounts payable	0
Balance savings account	5877.71	Equity	7669.61
Total	7669.61	Total	7669.61

Notes to the balance sheet

(1) Contributions from 8 members have been collected in 2022 (membership for 2022 was included in fees for 2021 or 2020; most members paid in 2020)

(2) In 2022 we collected 50 euro's of overdue membership fee's

(3) 10 euro was transferred from the current account to the savings account

(4) Compared to the previous year, there is decrease in the equity (because we did not asked for member fees but organized a VOC meeting)

(5) An overview of the evolution of the equity:

2022	€ 7,669.61	2009	€ 8,189
2021	€ 8,469.21	2008	€ 6,248
2020	€ 8,591.11	2007	€ 5,914
2019	€ 7,979.30	2006	€ 6,869
2018	€ 7,254.52	2005	€ 6,057
2017	€ 7,264.34	2004	€ 5,019
2016	€ 5,432.34	2003	€ 6,795
2015	€ 3,913.66	2002	€ 6,408
2014	€ 4,019.92	2001	€ 5,898
2013	€ 5,444.46	2000	€ 5,731
2012	€ 5,524.70	1999	€ 4,871
2011	€ 6,194	1998	€ 5,100
2010	€ 7,621		

Programme: 11th VOC Conference
Rotterdam, 2 June 2023

Room C1-2, Theil Building, Burgemeester Oudlaan 50, 3062 PA Rotterdam

10:00 – 10:30	**Members meeting**
10:30 – 11:15	**Contributed paper session 1**
	Tim Offermans (RU) – Process expert knowledge is essential in creating value from data-driven industrial soft sensors
	Aurore Archimbaud (EUR) – Tandem clustering with invariant coordinate selection (ICS)
11:15 – 11:45	**Break**
11:45 – 12:30	**Keynote address**
	Mark van de Wiel (AUMC) – Co-data learning
12:30 – 13:45	**Lunch**
13:45 – 14:45	**Contributed paper session 2**
	Max Welz* (EUR) – Quantifying effects of rating-scale response bias in correlational analyses
	Kevin Kloos* (LEI) – Continuous Sweep: a new, binary quantifier
	Paul Eilers (EMC) – Super-fast image deconvolution for super-resolution
14:45 – 15:15	**Break**
15:15 – 16:00	**Contributed paper session 3**
	Berber Postma* (RU) – Chemometric model adaptation for spectroscopic sensors in industry
	Nuria Senar* (AUMC) – A framework for interpretation and testing of sparse canonical correlations
16:00 – 16:15	**Break**
16:15 – 17:00	**Contributed paper session 4**
	Carlo Cavicchia (EUR) – Ultrametric Gaussian mixture models with parsimonious structures
	Carel Peeters (WUR) – Representation learning: shallowed be thy name
17:00 –	**Closing and drinks**

* PhD student presentation

Meeting Locations

The meeting will take place in **Room C1-2** on the first floor of the **Theil building** on the **Woudestein campus** of Erasmus University Rotterdam. Lunch and coffee breaks will be served in Room C1-2. The drinks at 17:00 will take place in the Erasmus Pavilion building on campus.

Detailed information about reaching and navigating the campus is available at https://www.eur.nl/en/campus/locations/campus-woudestein. Below is a summary.

Public transportation

From Rotterdam Central Station

There are multiple ways of reaching Woudestein from Rotterdam Central Station:

- Tram 7, direction Woudestein, stop 'Erasmus Universiteit'.
- Tram 21 or 24 direction De Esch. Stop 'Woudestein'.
- Subway (metro), direction Slinge. Transfer at Beurs, to subway in the direction of Binnenhof, Nesselande or De Terp. Get off the subway at station Kralingse Zoom. This subway is within walking distance (10 minutes) of the campus.

From Rotterdam Alexander

- Subway (metro), direction Schiedam Centrum or De Akkers. Get off the subway at station Kralingse Zoom. This subway is within walking distance (10 minutes) of the campus.

By car

The Woudestein campus is situated near the A16 and easily accessible by car.

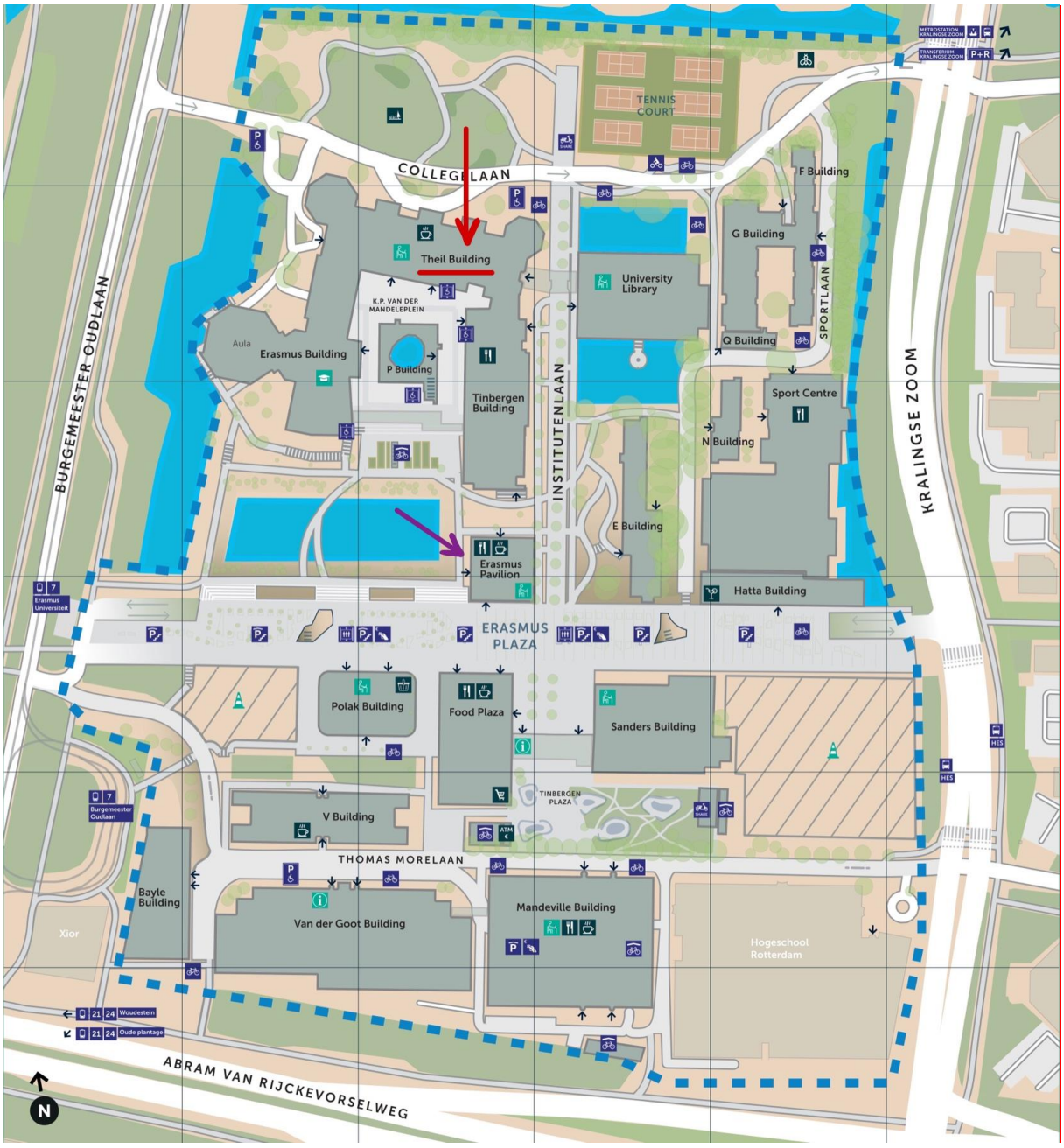
- On the A16, from both directions, take exit 25 (Centrum).
- Follow the signs 'Centrum'.
- Turn right at the traffic lights (Burgemeester Oudlaan).
- You will find the Woudestein campus on the right-hand side.

Paid parking is available on campus. The main parking is in the underground Erasmus Plaza garage at the center of the campus. The maximum day rate for parking is €10. Payment terminals are located at the center of the garage on both levels. More information is available at https://www.eur.nl/en/campus/locations/campus-woudestein/parking.

Campus map

On the campus map below, the Theil Building is marked with a red arrow in the top part of the map. The drinks at 17:00 will take place in the Erasmus Pavilion, which is marked with a purple arrow in the centre of the map. Metro station Kralingse Zoom is located to the north-east side of campus (off-campus), while the tram lines are on the (south-) west side of the campus.

The full map can be downloaded from https://www.eur.nl/en/campus/locations/campus-woudestein.



11th VOC Conference
2 June 2023

Erasmus University Rotterdam, the Netherlands
Room C1-2, Theil Building, Woudestein Campus

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

Co-data learning

Mark van de Wiel
Amsterdam University Medical Centers

Co-data learning is a framework that allows to incorporate complementary information on the features (co-data) into the learning process. Examples of co-data are p-values from previous studies or a predefined grouping of features such as chromosomes for genes. We start out by discussing how co-data learning relates to, but also differs from, transfer learning and historical priors. Here, our focus lies on supervised learning and feature selection. We formulate co-data learners such that the importance of the co-data data source is regulated by hyper-parameters, like penalties in penalized regression. The main challenge is to learn those hyper-parameters efficiently and stably. We show how to use (empirical) Bayes procedures for this purpose in both a dense (ridge) and sparse (horseshoe) regression setting. These methods will be illustrated on high-dimensional oncological data, which are used to build diagnostic and prognostic models. We end by shortly discussing extensions of co-data learning to factor analysis and tree-based learners.

CONTRIBUTED TALKS

Process expert knowledge is essential in creating value from data-driven industrial soft sensors

Tim Offermans
Radboud University

The objective of the current Industry 5.0 trend is to (re)centre the human operator amidst process automation tools. In this work we elucidate on why and how this can be achieved for production data-modelling technology. Specifically, we investigate and compare data- and expert knowledge-driven approaches for ordering and selecting process parameters for real-time product quality prediction using a soft-sensor. We provide a demonstration for three parallel dairy processing lines. We show that well-established data-driven selections are outperformed by expert-driven selections, but that the data-driven selection methods can improve an initial expert-driven selection, either in terms of accuracy or in terms of the number of parameters selected. The reason identified for this is instability of process parameters due to temporal variation, which is easier picked up by experts than by data-driven methods. Overall, this study shows that taking more effort to include human expert knowledge in data modelling technologies benefits more human-centred, high-quality and sustainable industrial production.

Tandem clustering with invariant coordinate selection (ICS)

Aurore Archimbaud
Erasmus University Rotterdam

Tandem clustering is a well-known technique for dealing with high-dimensional or noisy data to better identify clusters. This is a sequential approach based on first reducing the dimension of the data and then performing the clustering. The most common method, based on principal component analysis (PCA), has been criticized for only focusing on maximizing inertia and not necessarily preserving the structure of interest for clustering. Therefore, we suggest a new tandem clustering approach based on invariant coordinate selection (ICS). This multivariate method is designed to identify the structure of the data by jointly diagonalizing two scatter matrices, while maintaining the affine invariance of the new coordinates. More specifically, some theoretical results proved that under some elliptical mixture models, the first and/or last components are carrying the information regarding the clustering structure. However, despite the attractive properties of ICS, the method has not been studied much in the context of clustering but mostly for outlier detection purposes. The issues of choosing the pair of scatter matrices and the components to keep are the two challenges that must be addressed. For clustering purposes, we suggest that the best scatter pairs consist of one matrix which captures the within-cluster structure and another which captures the global structure. To this end, we find the local shape or pairwise scatters to be good choices for estimating the within-structure. In addition, we also investigate the use of the well-known minimum covariance determinant (MCD) estimator based on a smaller-than usual subset size. The performance of ICS as a dimension reduction method is evaluated to determine its ability to preserve the cluster structure of the data. We conducted a large simulation study and applied it to benchmark data sets. We tested various combinations of scatter matrices, component selection criteria, and the effects of the presence of outliers. Results indicate that the ICS-based tandem clustering method has superior performance over PCA, and thus is a promising approach.

Quantifying effects of rating-scale response bias in correlational analyses

Max Welz
Erasmus University Rotterdam

Responses to rating-scale items are often plagued by biases stemming from content-responsive faking (such as malingering or socially desirable responding) or content nonresponsitivity (particularly careless responding). While there is consensus that response biases can jeopardize the validity of survey measures through a variety of psychometric issues, their exact effects are yet to be statistically quantified. Leveraging robustness theory, we study the statistical properties of response biases in survey data. In particular, we derive bias curves and breakdown values of survey measures, with a focus on correlational measures due to their key role in factor analyses and structural equation models. Furthermore, we study how the adverse effects of response biases can be mitigated by survey design, for instance through the number of answer categories, number of items in a

measure, and construct reliability. We find that already low prevalence of response biases can render survey measures fundamentally invalid. In addition, we show how comparatively short survey measures with a balanced number of negatively-worded items can enhance the robustness of survey measures against response biases. Furthermore, we provide freely available software in R for computation and visualization of bias curves in survey measures.

Continuous Sweep: a new, binary quantifier

Kevin Kloos
Leiden University

Quantification is a supervised machine learning task, focused on estimating the class prevalence of a dataset rather than labeling its individual observations. We introduce Continuous Sweep, a new parametric binary quantifier inspired by the well-performing Median Sweep. Median Sweep is currently one of the best binary quantifiers, but we have changed this quantifier on three points, namely 1) using parametric class distributions instead of empirical distributions, 2) optimizing decision boundaries instead of applying discrete decision rules, and 3) calculating the mean instead of the median. We derive analytic expressions for the bias and variance of Continuous Sweep under general model assumptions. This is one of the first theoretical contributions in the field of quantification learning Moreover, these derivations enable us to find the optimal decision boundaries. Finally, our simulation study shows that Continuous Sweep outperforms Median Sweep in a wide range of situations.

Super-fast Image Deconvolution for Super-resolution

Paul Eilers
Erasmus University Medical Center

The resolution of an optical microscope is limited by the wavelength of light. The observed image can be seen as the convolution of the true image and the spread function of the optical instrument. One way to improve the resolution of a digitized image is deconvolution by penalized regression. The principle is simple, but practical implementation is challenging. To increase the resolution of an image of 250 by 250 pixels with a factor 4, a linear system with one million unknowns has to be solved. A straightforward implementation would be very demanding in computation time and computer memory.

The conjugate gradients method for solving (large) linear systems of equations uses only products of matrices and vectors (Hestenes and Stiefel, 1952). The optical spread function can be written as the outer product of two vectors. The two-dimensional convolution can thus be written as a one-dimensional convolution of the rows of the source image, followed by a convolution of the columns of the result. Combining these ideas leads to a very fast and compact algorithm.

The algorithm needs less than ten iterations for a sharp result. Already after three or four iterations a useful image is obtained. On a PC with an I5-8400 processor and Matlab 2012,

one iteration of the algorithm takes one second when increasing the resolution of a 250 by 250 pixels image by a factor 8.

I will discuss and illustrate the theory and show applications to biological images. The software is freely available.

Reference
Hestenes MR and Stiefel E] (1952). Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49: 409--436.

Chemometric model adaptation for spectroscopic sensors in industry

Berber Postma
Radboud University

In the food industry, quality control is vital not just for the satisfaction, but also the safety of consumers. In-line spectroscopy is a quick, cost-efficient and informative way to measure quality indicators during production. To monitor product quality, chemometric models need to be calibrated to predict product composition from raw spectral data. Traditionally, a separate model is constructed and maintained for each production line and measuring instrument, which can be time-consuming, costly and labour-intensive. However, chemometric model adaptation can be used to combine data from several instruments into a single model, and to optimise the process of model calibration and maintenance. In this study, we critically compare the effectiveness of model adaptation methods for this purpose. Using NIR-spectra of a dairy product as a demonstration, we found calibration transfer to be the most potent method. We also present a sample matching technique that overcomes the traditional necessity of this method to have calibration samples measured on all instruments of interest. This opens up the possibility to apply calibration transfer much more broadly. Datasets from many different instruments and product lines, potentially from all around the world, can be combined into one unifying model.

A framework for interpretation and testing of sparse canonical correlations

Nuria Senar
Amsterdam University Medical Centers

In clinical and biomedical research, multiple high-dimensional datasets are nowadays routinely collected from omics and imaging devices. Multivariate methods, such as Canonical Correlation Analysis (CCA), integrate such datasets to discover and understand underlying biological mechanisms. For an explorative method such as CCA, interpretation is key. Sparse CCA alternatives use penalties for dimension reduction in high-dimensional settings to help estimation and interpretation. However, choosing these penalties brings computational burdens and may render unintelligible results.

We present a sparse CCA method that improves interpretability, produces near-orthogonal components, allows for browsing over various sparsity levels, and implements permutation-based hypothesis testing. We use the Nonlinear Iterative Partial Least Squares algorithm which can efficiently recover relationships linking high-dimensional datasets. We impose sparsity through a soft-thresholding approach to avoid tuning a penalty parameter by, instead, fixing the number of nonzero variables. In addition, this approach shows less dependency on the initialisation values as well as consistency in variable selection for different choices of nonzero variables.

We evaluated our approach on simulations and illustrate its use on cancer genomics data from drug sensitivity screens. Throughout, its performance is compared to that of a popular sparse CCA alternative, Penalised Matrix Analysis (PMA). Our method improves interpretability of subsequent components, while not compromising, or even improving, signal discovery.

Ultrametric Gaussian mixture models with parsimonious Structures

Carlo Cavicchia
Erasmus University Rotterdam

Multidimensional phenomena are usually characterized by nested latent dimensions associated, in turn, with observed variables. These phenomena, for instance, poverty, well-being, and sustainable development, can often differ across countries, or cities within countries, in terms of dimensions, other than in their relationships to each other, on the one hand, and their importance in the definition of the general concept, on the other hand. This paper discusses several parsimonious structures of the covariance matrix reconstructing relationships among variables which can be implemented in Gaussian mixture models to study complex phenomena in heterogeneous populations.

Representation learning: shallowed be thy name

Carel Peeters
Wageningen University & Research

We will focus on representation learning (RL), i.e., learning representations of data that make subsequent learning tasks easier. From a probabilistic perspective RL can be viewed as the recovery of a low-dimensional set of latent random variables that captures the information contained in the observed data. We will treat the case in which a one-layer generative neural network concurs with the classic factor analytic model. We will see that, for high-dimensional data, such shallow representations are sufficient for data representation and downstream modeling support.