



Vereniging voor Ordinatie en Classificatie / Dutch-Flemish Classification Society

Chairman: Mark de Rooij, Universiteit Leiden, Faculteit Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, 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 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.

12th VOC Conference 8 March 2024 Radboud University, Nijmegen

Heyendaalseweg 141 6525 AJ Nijmegen

Room EOS N 01.150, Elinor Ostrum Building

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: Jeroen Jansen

From Chemical Fingerprints to Environmental Footprints: Advancing Feed Production Through Near-infrared Spectroscopy, Life Cycle Assessment and Novel Artificial Intelligence

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

In this issue:

| | |
|--|----|
| Short program 12 th VOC Conference | 1 |
| Registration details for the 12 th VOC Conference | 1 |
| From the President | 2 |
| Conference Announcements | 2 |
| Publications | 3 |
| Annual Report of the Secretary for 2023 | 4 |
| Minutes of the Members Meeting 2023 (EUR) | 4 |
| Agenda Annual Members Meeting 2024 (RU) | 5 |
| Financial Report for 2023 | 6 |
| Programme 12 th VOC Conference | 7 |
| Meeting Location | 8 |
| Book of Abstracts 12 th VOC Conference | 10 |

Registration details for the 12th VOC Conference

Those who would like to join the 12th 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 colleagues,

In 2024 the VOC celebrates its 35th birthday, and as we did previously we will also this year organise a jubilee meeting with invited national and international speakers. The organising committee is putting the last pieces of the program together and hope to inform you about the program in March. The meeting will be on November 21 and 22, this year. More information will follow by email.

Another exciting event we got engaged in is a joint conference of the Italian Classification Society (CLADAG) and the Dutch/Flemish classification society (VOC). The meeting will be held in Cagliari and takes place in September 2025. We just had a few calls and started thinking about the set-up of the joint conference. I am very much looking forward to this event, but I need many of you to make it a success. If you like to be involved in the organisation, just let me know.

Prior to the jubilee meeting and this joint CLADAG-VOC, we have our usual annual conference. This year the conference will be organised in Nijmegen on March 8th. The program of the meeting can be found in this newsletter. As you can see the presentations have a wide range of topics. We will hear about a comparison of Dutch data science and AI master programs, individual decision making, measurement with handheld spectroscopic devices, the use of stacking to form new theories, pollutants in the Rhine, distances for mixed variables, but also reduced rank regression for mixed variables, monitoring of polymer degradation for recycling processes, interpretable kernels, and robust matrix completion for rating scale data. It promises to be a very nice and inspiring day. So, don't hesitate to register and come to Nijmegen!

We start the annual conference with a short members meeting, where we will discuss current activities of the society. One of our board members, Matthijs Warrens, will leave the board. Matthijs has been a board member since 2018. I would like to thank Matthijs for his 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 me an email.

I hope to see many of you during our conference in Nijmegen!

Mark de Rooij

Conference Announcements

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 **PsychoCo 2024** workshop on psychometric computing takes place at Erasmus University Rotterdam from February 29th to March 1st 2024. The programme features keynote talks by Marco Riani and the VOC's very own Michel van de Velden: see <https://www.psychoco.org/2024/index.html>. If you are quick, you may still be able to sign up!

The 18th conference of the **International Federation of Classification Societies** will take place in San José, Costa Rica, on 15-19 July 2024. For 2024, the University of Costa Rica has been chosen to host the IFCS Conference, that will be organized by the Central American and Caribbean Society for Classification and Data Analysis (SoCCCAD). The conference will feature keynote speakers, invited and contributed sessions, and organized streams on specific topics. Preceding the conference itself, tutorials on up-to-date fields will be organized on July 15. The Local Organizing Committee is working enthusiastically to make sure IFCS 2024 will be a successful conference and is looking forward to welcoming all of you in Costa Rica. Do not forget to add IFCS 2024 to your agenda! Visit <https://ifcs.ucr.ac.cr/index.php/en/> for more information.

The **International Conference on Robust Statistics (ICORS) and Data Science, Statistics & Visualisation (DSSV)** is a joint conference aimed to bring together researchers and practitioners interested in the interplay of robust statistics, data analysis, computer science, and visualization, and to build bridges between these fields for

interdisciplinary research. This includes theoretical and applied statisticians as well as data analysts from other fields, and leading experts as well as junior researchers and graduate students. The conference is hosted by George Mason University (Fairfax, VA, USA) from 29 July to 1 August 2024. The conference welcomes contributions to applied statistics as well as theoretical statistics, in particular new problems related to robust statistics, machine learning and statistical learning, outlier detection, visualization and verbalization of data, data analysis and other important themes. See <https://icors2024.statistics.gmu.edu> for information.

Publications

Ernst, A. F., Timmerman, M. E., Ji, F., Jeronimus, B. F., & Albers, C. J. (2023). Mixture multilevel vector-autoregressive modeling. *Psychological Methods*. Advance online publication.

<https://doi.org/10.1037/met0000551>

Kroonenberg, P.M. & Stoltenborgh, M. (2023). Handbooks as a means of mapping a discipline. An exploration of the World of Attachment. In A. Okada, K. Shigemasu, R. Yoshino, S. Yokoyama (eds.) *Facets of Behaviormetrics: The 50th Anniversary of the Behaviormetric Society* (pp. 123-143). Singapore: Springer.

Kroonenberg, P.M. (2024). Nishisato's psychometric world. In: Beh, E.J., Lombardo, R., Clavel, J.G. (Eds.) *Analysis of categorical data from historical perspectives*. Behaviormetrics: quantitative approaches to human behavior. Springer, Singapore. <https://doi.org/10.1007/978-981-99-5329-5>

Schoonees, P. C., Groenen, P. J., van de Velden, M., & van Herk, H. (2023). Effects of Visual Priming on Rating Scale Usage. In *Multivariate scaling methods and the reconstruction of social spaces: Papers in honor of Jörg Blasius* (pp. 213-225). Verlag Barbara Budrich.

Van Mechelen, I., Boulesteix, A.-L., Dangl, R., Dean, N., Hennig, C., Leisch, F., Steinley, D., & Warrens, M. J. (2023). A white paper on good research practices in benchmarking: The case of cluster analysis. *WIREs Data Mining and Knowledge Discovery*, 13, e1511, 1-20. <https://doi.org/10.1002/widm.1511>

Vecco, M., Georgantzis, N., & Kroonenberg, P. M. (2022). Is it the firm, the innovator, or the innovation? Determinants of perceived non-imitability leading to unprotected Intellectual Property. *International Review of Law & Economics*, 2(8):106095. <https://doi.org/10.1016/j.irl.2022.106095>

Annual Report of the Secretary for the years 2023

1. Number of members

The VOC started 2023 with 101 members and counted 105 members at the end of 2023. One membership was terminated and there were 5 new members registered.

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 once in 2023 (October). The main topics were the organization of the 12th VOC conference at the Radboud University Nijmegen (March 8, 2024) and the VOC Jubilee Conference (November, 2024).

3. Activities

The main activity of the VOC in 2023 was the 11th VOC conference. The conference took place at Erasmus University Rotterdam on June 2, 2023 with a full day program, including ten contributions on a range of topics. A keynote contribution was given by Mark van de Wiel (Amsterdam University Medical Centers) on 'Co-data learning'.

4. Publicity

The newsletter appeared once (number 62) in 2023. The VOC conference was also announced to non-VOC members, using various channels, including the IFCS newsletter.

Minutes from the VOC Annual Members Meeting (2 June 2023, Rotterdam)

1. Opening of the Members Meeting

Mark opens the meeting. Introduces himself as president. Asks if someone wants to add something to the agenda. No one.

2. Minutes of the Members Meeting 14 October 2022

Mark introduces the minutes from the last meeting. Asks if there are any questions.

Kathrin mentions the timing of the conference. Asks if it is better to have the conferences in autumn. Mark says that it is usually in May/June. Agrees that there is a lot of holidays and conferences start. Mark says we can think about better timing. Kathrin mentions start of July. Carel mentions that fall may have issues for mathematics.

Carlo mentions that summer is difficult. Mark wants to avoid July and August. Mark mentions that it is a challenge to find a day. Mark suggests end of March / beginning of April (Before King's day). Kathrin mentions for her it would be better.

Carlo suggests to be far from IFCS, ECDA and so on. Mark suggest we could try and plan it earlier (March). We should remember to do this, Mark mentions.

Minutes approved.

3. Annual Report of the Secretary on the year 2022

Mark handles this in the absence of Matthijs. Mentions that we grew a bit. Mentions that we started up last year after Covid. Nice that we have new members. Mark asks whether there are any questions. None. Approved.

4. Financial report of the treasurer on the year 2022

Word given to Tom. Contributions were taken in again for this year. Mentions that we need some back up funds for the jubilee conference. In his opinion the back up is enough. Mark mentions we are doing quite well. Michel and Jeroen were the cash committee. Michel mentions it looks all fine. Mentions that it could be discharged.

Paul mentions that a written statement is required. Mentions that he can supply one. The official way is that there is a written statement.

Mark asks Tom to take that into account for next time.
Tom agrees.

Mark mentions that a new cash committee is needed.
Michel agrees to do it again. Tim is asked by Mark, and agrees. Handled. No complaints, so accepted (discharged).

5. Composition of the Board

Mark introduces this point. As mentioned earlier, with the Covid pandemic, the remaining term was a bit of a mess. Pieter mentions that it is from the board meeting. Mark says he does not think we need to approve it. Mark says we will look for new members next year. Pieter asks whether Mark will continue. Mark replies that indeed he has been around for 7 years and will think about it for the next board meeting.

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

Mark asks for input.

Mentions that Gerjen, Carel and Mark has started preparing for the jubilee meeting and has made a shortlist of locations. Mentions that there will be communication to block your calendar.

7. Questions before closure of the meeting

None.

8. Closure of the Members Meeting

Mark mentions that the meeting is closed. Thanks everyone for being present.

Agenda for the VOC Annual Members Meeting (8 March 2024, Radboud University, Nijmegen)

1. Opening of the Members Meeting

2. Minutes of the Members Meeting 2 June 2023

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

3. Annual Report of the Secretary on the year 2023

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

4. Financial report of the Treasurer on the year 2023

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 (0)
- Tom Wilderjans, Treasurer (2)
- Matthijs Warrens, Secretary (0)
- Pieter Schoonees, Newsletter Editor (2)
- Carel Peeters, Webmaster (1)
- Gerjen Tinnevelt, Member (1)
- Kathrin Gruber, Member (1)

6. Miscellaneous

7. Questions before closure of the meeting

8. Closure of the Members Meeting

Financial Report for 2023

| Revenue | | Expenditure | |
|-------------------------------------|----------------|-----------------------|----------------|
| membership fees (45 paying members) | 1125 | transaction costs ING | 235.73 |
| overdue membership fees | 0 | website hosting | 146.41 |
| interest savings account | 46.96 | VOC meeting Rotterdam | 209.79 |
| | | IFCS dues 2023-2024 | 221.38 |
| Total | 1171.96 | Total | 813.31 |
| | | | |
| Debit | | Credit | |
| Balance ING account | 2103.59 | Accounts payable | 0 |
| Balance savings account | 5924.67 | Equity | 8028.26 |
| Total | 8028.26 | Total | 8028.26 |

Notes to the balance sheet

(1) Contributions from 45 members have been collected in 2023

(2) In 2023 we collected 0 euro's of overdue membership fee's

(3) Compared to the previous year, there is small increase in the equity (because we again asked for member fees -after not doing this during Corona- and kept the costs low for the VOC meeting)

(4) An overview of the evolution of the equity

| | |
|------|------------|
| 2023 | € 8,028.26 |
| 2022 | € 7,669.61 |
| 2021 | € 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 |

Programme: 12th VOC Conference

Nijmegen, 8 March 2024

Room EOS N 01.150, Elinor Ostrom Building, Heyendaalseweg 141 6525 AJ Nijmegen

- 10:00 – 10:30 **Members meeting**
- 10:30 – 11:15 **Contributed paper session 1**
- Patrick Groenen (EUR) – Interpretable Kernels
- Berber Postma* (RU) – Trend Analysis of Pollutants in Rhine Water
- 11:15 – 11:45 **Break**
- 11:45 – 12:30 **Keynote address**
- Jeroen Jansen (RU) – From Chemical Fingerprints to Environmental Footprints: Advancing Feed Production Through Near-infrared Spectroscopy, Life Cycle Assessment and Novel Artificial Intelligence
- 12:30 – 13:45 **Lunch**
- 13:45 – 14:45 **Contributed paper session 2**
- Zino Brystowski* (LEI) – Stacking for Multidisciplinary Theory Development
- Andreas Alfons (EUR) – Robust Matrix Completion for Rating-Scale Data
- Sanwouly Marlene Yao* (RU) – Enhancing Analytical Measurements in Field and Industry: An Innovative Methodology Using Handheld Spectroscopic Devices
- 14:45 – 15:15 **Break**
- 15:15 – 16:15 **Contributed paper session 3**
- Jacqueline Zadelaar (LEI) – Developmental Differences in Informational Influence on Decision Making: Decision Strategies Differences in Late Childhood and Early Adolescence
- Lorenza Cotugno* (Naples) – A Generalized Mixed Regularized Reduced Rank Regression Model
- Diogo Goncalves* (RU) – Tracking the Degradation of Reprocessed Polypropylene by Relating NIR Spectroscopy and Chemiluminescence Using PLS2
- 16:15 – 16:30 **Break**
- 16:30 – 17:15 **Contributed paper session 4**
- Carlo Cavicchia (EUR) – Unbiased Mixed Variable Distance
- Zsuzsa Bakk (LEI) – Unraveling the Skillsets of Data Scientists: Text Mining Analysis of Dutch University Master Programs in Data Science and Artificial Intelligence
- 17:15 – **Closing and drinks**

* Student presentation

Meeting Location

The meeting will take place in **Room EOS N 01.150** of the **Elinor Ostrum Building** housing the Nijmegen School of Management on the campus of Radboud University in Nijmegen.

Detailed information about reaching and navigating the campus is available at <https://www.ru.nl/en/services/getting-to-the-campus-and-parking>. Below is a summary.

Train

Trains call at Nijmegen Central train station and Nijmegen Heyendaal train station. Heyendaal train station is within walking distance of the university. Walking from the Nijmegen Heyendaal station to the meeting location will take approximately 15 minutes (1km).

Bus

Direct buses run every five minutes during rush hour from Nijmegen Central train station to Radboud University. Bus 10 (Heyendaal Shuttle) calls at all bus stops. Get off at bus stop *Erasmus Building* for our meeting location.

An OV chip card top-up machine is found near the main entrance of the University Library. Plan your journey via www.ns.nl or www.9292.nl (or use the corresponding phone apps).

Bicycle

Radboud University has different indoor and outdoor bicycle parking facilities. These facilities are found in or near each building on campus.

Car

Radboud University is indicated on the ANWB road signs on all access roads in Nijmegen.

Parking

Paid parking is available from Monday to Friday (except on official holidays) from 7.00 to 18.00. Parking up to a half hour is free. Gymnasium car park (P7b) is the recommended parking location for the meeting location. You can charge your electric car at indoor parking Grotius, Gymnasium and Huygens-gebouw and parking Berchmanianum.

Campus map

On the campus map below, the **Elinor Ostrum Building** is marked with a red arrow in the middle of the map (number 15 on the map). Train station Nijmegen Heyendaal is located on the north-side of the campus, and is shown at the top of the map.

The full map can be downloaded from <https://www.ru.nl/en/services/getting-to-the-campus-and-parking/campus-map>.





12th VOC Conference
8 March 2024

Radboud University, Nijmegen, the Netherlands
Room EOS N 01.150, Elinor Ostrom 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

From Chemical Fingerprints to Environmental Footprints: Advancing Feed Production Through Near-infrared Spectroscopy, Life Cycle Assessment and Novel Artificial Intelligence

Jeroen Jansen

Radboud University

Process Analytical Technologies has been the key technology of quality maintenance and improvement in process industry. Quality is however only an indicator of process excellence: Safety, Cost, Delivery, Maintenance and specifically Environment are strongly complementary determinants of process value. The rising societal demands on sustainability of contemporary process industry has made specifically environmental impact increasingly relevant, demonstrable by the implementation of CSRD into national legislation in the coming years. This however creates an interesting “collision of timelines” as the future predictions from large volumes of PAT data collide with the retrospective quantification of environmental with Life Cycle Assessment (LCA), of data that is available at time of production.

Aside from quality information, PAT data (e.g. NIR spectra) contain a wealth of information on aspects like provenance, which are the key inputs for LCA. The available sustainability data on ingredients may therefore also be used to predict the footprint of the end-product. In this way, both quality and environmental impact (and production cost) may be simultaneously predicted. This allows the producer to take control of the product footprint, like they already are used to take control of quality through PAT. We show on a case study of animal feeds, how NIR spectroscopy (1) adequately predicts all product outcomes, (2) likewise predicts ingredient provenance, thereby providing a paperless evidence basis for their origin and (3) makes transparent the economic balance underlying sustainable production.

CONTRIBUTED TALKS

Interpretable Kernels

Patrick Groenen

Erasmus University Rotterdam

The use of kernels for nonlinear prediction is widespread in machine learning. They have been popularized in support vector machines and used in kernel ridge regression, amongst others. These methods share three aspects. First, instead of the original $n \times p$ matrix of predictor variables, each row is mapped into a high dimensional feature space. Second, a ridge penalty term is used to shrink the weights (coefficients) on the predictors in the high-dimensional feature space. Third, the solution is not obtained in this feature space, but through solving a dual problem. A major drawback in the use of kernels is that the interpretation in terms of the original predictor variables is lost.

In this paper, we argue that in the case of a wide $n \times p$ matrix of predictor variables (with $p > n$), the kernel solution can be re-expressed in terms of a linear combination of the original matrix of predictor variables and a ridge penalty that involves a special metric.

Consequently, the exact same predicted values can be obtained as a weighted linear combination of the predictor variables in the usual manner and thus can be interpreted. In the case $p \leq n$, we discuss a least-squares approximation of the kernel matrix that still allows the interpretation in terms of a linear combination. It is shown that these results hold for any function of a linear combination that minimizes the weights and has a ridge penalty of these weights such as in kernel logistic regression and kernel Poisson regression. When the objective function is minus the log likelihood, standard likelihood theory can be used to estimate the standard deviations of the weights.

As an extension, it is possible to apply an approximation in a k -dimensional space with $k < p$, thereby enforcing dimension reduction in the predictor space.

Trend Analysis of Pollutants in Rhine Water

Berber Postma

Radboud University

The International Commission for Protection of the Rhine (ICPR) has set goals to reduce the median loads of a list of micropollutants by a certain amount in the year 2040. Currently, progress towards this is monitored by fitting a linear trend through the loads. However, there are indications that the decrease of pollutant loads is not in fact linear. In this study, we compared the current linear prediction models with more sophisticated exponential models. We also investigated the effect that seasonal factors (temperature and discharge) have on the loads.

Stacking for Multidisciplinary Theory Development

Zino Brystowski

Leiden University

The use of predictive modeling for research and theory development has increasingly come into focus in recent years (Yarkoni & Westfall, 2017; Shmueli, 2010). In predictive modeling, a model is selected based on its ability to predict data across samples. Accordingly, a pattern found in one sample should be of relevance for predicting data when a different sample is used. Building theories based on predictive ability therefore ensures that they are more robust to sampling variation and better generalizable to new data. This is beneficial in both applied and theoretical contexts and allows to assess the empirical relevance of a theory.

Selecting models based on predictive performance provides further advantages. It enables the comparison of competing theories, different measurement concepts such as instruments or operationalizations, and different methodological approaches (Shmueli, 2010). This helps in the development of measurement and analysis methods. Another advantage is that predictive modeling methods are well suited for the analysis of large data sets. Features can be selected on the basis of their predictive relevance and complex patterns can be

uncovered that lead to the generation of new hypotheses, which in turn can be incorporated into theory development (Shmueli, 2010).

In this work, a predictive modeling framework for theory development in multidisciplinary research is presented. In research contexts where data are collected using measurements from different research domains, data analysis and theory development pose particular challenges. Simply aggregating all data and fitting a single model ignores the differences in data properties between measurements from different research domains. For example, genetic data have many features with high redundancy, while questionnaire data are more compact. Another difficulty when merging data is that the theories developed in the respective fields cannot be easily integrated. This leaves scientific advances unconsidered and limits the testing and development of field-specific theories.

These challenges can be addressed by using a multi-step procedure based on the method stacked generalization (Breiman, 1996; Wolpert, 1992). In a first step, models are fitted separately to each set of predictors that define a research domain. This enables to build models specific to each research domain, allowing the integration and comparison of field-specific theories. In a second step, these models are used to generate cross-validated predictions per research domain. In a third step, the predictions obtained for each research domain are entered as predictors in a new model, called a meta-model, that uses them to predict the outcome of interest. The use of the meta-model makes it possible to evaluate and select research domains based on their relevance for the prediction of new sample data.

This combination of integrating field-specific theories (domain specificity), and the selection of research domains based on predictive relevance (domain selection) makes this approach especially useful for theory development in a multidisciplinary context.

The demonstration of the method is illustrated by an example application to data from the ABCD study, a multidisciplinary research project on health and brain development in children and adolescents (Volkow et al., 2018). The method is used to analyze the relevance of different research domains such as brain imaging and mental health for the prediction of a measure of self-regulation.

The method, its application, and future directions will be discussed.

Robust Matrix Completion for Rating-Scale Data

Andreas Alfons

Erasmus University Rotterdam

Low-rank matrix completion (LRMC) has gained considerable interest in recent years. The goal of LRMC is to predict the unknown entries of a partially observed matrix using its known elements and a low-rank rank constraint while minimizing a specific criterion such as the mean squared error. Although common applications feature discrete rating-scale data, such as the well-known Netflix Prize competition on recommender systems, methods for LRMC are almost always designed for and studied in the context of continuous data matrices. Little is therefore known on the statistical properties of LRMC methods on discrete rating-scale data. Furthermore, while ample work on LRMC exists, only a relatively small subset of the literature has considered matrix completion in the presence of corrupted observations. Yet corrupted observations may widely occur in user-product rating matrices, for instance via

malicious users who deliberately manipulate ratings in so-called attacks in order to influence a recommender system to their advantage. We contribute to the literature on LRMC by introducing a novel matrix completion algorithm that is tailored towards the discrete nature of rating-scale data and that is robust to the presence of corrupted observations. In addition, we study the performance of the proposed method and several competitors in simulation studies with discrete rating-scale (rather than continuous) data and under various attack scenarios.

Enhancing Analytical Measurements in Field and Industry: An Innovative Methodology Using Handheld Spectroscopic Devices

Sanwouly Marlene Yao

Radboud University

In contemporary industrial settings, operators frequently encounter challenges in conducting analytical measurements. The conventional approach involves various trips to the laboratory or quality control lab, necessitating the transportation of samples, sometimes even resorting to cycling if the location is remote. The subsequent waiting period for analysis consumes valuable time, incurs significant costs, and demands considerable effort presenting a pervasive issue.

Our research addresses this challenge by introducing a more efficient method leveraging handheld spectroscopic devices. These devices, varying in spectral ranges, are user-friendly, lightweight, and easy to operate. Several handheld devices have been used to scan different samples already and the data resulting from it have proven efficient in predictions and classifications, showcasing the potential in revolutionizing current practices. Notably, these devices can be used to predict diverse parameters, from the color of a sample to the aramid content in a bulletproof vest. The methodology involves employing chemometric models, encompassing preprocessing tailored to the specific device, model development for prediction, and training/testing through techniques such as Partial Least Squares (PLS), or random forest for non-linear data.

Beyond mere predictive capabilities, our research advocates a holistic approach. We aim to use the data from the handheld spectroscopic devices to enhance the environmental sustainability of companies. Our ultimate goal is to transition from fingerprinting to footprinting, thereby contributing to a more conscientious and eco-friendly industrial landscape. This paradigm shift not only streamlines analytical processes but also aligns with broader sustainability goals, marking a significant advancement in the field.

Developmental Differences in Informational Influence on Decision Making: Decision Strategies Differences in Late Childhood and Early Adolescence

Jacqueline Zadelaar

Leiden University

When making decisions, individuals may take into account the decision behavior of others for either performance-related reasons (i.e., informational social influence) or socio-normative reasons (i.e., normative influence). The current study examines developments in decision strategy use between late childhood and early adolescence for both informational and normative influence. A computerized perceptual decision task was administered to 255 participants, both children (ages 9-10) and adolescents (ages 13-14), where “the other” was either a computer (informational version) or a supposed classmate (normative version).

A Bayesian hierarchical mixture model assigned decision strategies (i.e., how available information is translated into decision behavior) and estimated the subjective weights of different pieces of decision-relevant information on an individual level. As such, each individual was assigned to a strategy (a latent class/category) and the extent to which cognitive mechanisms corresponding to that strategy influenced decision making (continuous parameters) were estimated simultaneously. By separating these categorical and continuous individual differences, we aim to prevent mistaking one for the other as well as changes in one obscuring changes in the other.

Decision strategy assignment indicated that both children and adolescents consistently ignored the supposed classmate’s decisions. The majority of children also ignored the computer’s decisions but the majority of adolescents integrated these into their own decision-making process. On average, adolescents were 12% more likely to use a strategy that integrates personal observations and the computer’s decisions than children. Among those using such an integrative strategy, the extent to which the computer’s decisions weighed into the decision making did not change meaningfully with age. Equal performance confidence was reported across age groups and task versions.

Adolescents using the integrative strategy more often than children, but only in the informational version of the task, was explained by the strategy’s computational complexity. Presumably, children lack the cognitive capacities required to incorporate both their own observations and the computer’s decisions given the narrow time restrictions of the task. As such, the majority of children simply ignored the computer’s decisions. The supposed classmate’s decisions being consistently ignored may indicate that a classmate was deemed less competent than a computer at the task, thus providing no incentive to heed the classmate.

We conclude that informational influence increases between late childhood and early adolescence as adolescents are more likely to use decision strategies that consider the decisions of another on top of their own observations, at least under certain circumstances. Additionally, computers may be perceived as more competent than human beings at certain types of decision making. This may render adolescents more vulnerable to, for example, misinformation by AI and other digital sources. Further research is needed for insights concerning the development of normative influence on decision making. Sensitivity of

strategy assignment may be improved by incorporating covariates like age and task version into the Bayesian hierarchical mixture model rather than making these separate analyses.

A Generalized Mixed Regularized Reduced Rank Regression Model

Lorenza Cotugno

University of Naples Federico II

This paper proposes a Generalized Mixed Regularized Reduced Rank Regression model (GMR4) for mixed response and predictor variables. An algorithm will be developed and implemented in R, tested using simulation studies, and applied to an empirical data set. We will use data from the European Commission's Eurobarometer Surveys (Jan-Feb 2023), with a specific focus on residents of the Netherlands.

Reduced Rank Regression models are regression models for multiple outcome variables. These models over time have been used for different types of response variables: numeric (Izenman, 1975), binary (Yee and Hastie, 2003; De Rooij 2022), and ordinal response variables (De Rooij, Breemer, Woestenburg, and Busing, 2023). To deal with different types of predictor variables, Optimal Scaling will be used (Meulman, et al. 2019). Furthermore, this model includes a penalty on the coefficients to address challenges associated with high-dimensional data, such as having an excessive number of predictors compared to observations and the presence of multicollinearity among the predictors. Penalties can be applied using Ridge, Lasso, or Elastic Net. For the estimation of the parameters, the Majorization Minimization algorithm will be presented. Furthermore, a small Monte Carlo simulation study is set up to investigate how well the algorithm retrieves population parameter value.

Tracking the Degradation of Reprocessed Polypropylene by Relating NIR Spectroscopy and Chemiluminescence Using PLS2

Diogo Goncalves

Radboud University

Incorporating recycled polymer into plastic production is key towards the transition to a circular economy. Ideally, reprocessed polymer is integrated into new productions until it is no longer viable. Consecutive recycling cycles diminish the quality of the polymer which may jeopardize its application down the line.

Spectroscopy offers the possibility to differentiate between reprocessed and virgin polymer incorporated in new plastic materials in an inexpensive and easily automatable way. Chemiluminescence is a well-established technique to monitor the oxidative stability of polymers. Despite this, there is a lack of solutions for a practical monitoring of polymer degradation resulting from recycling.

Here we show the application of PLS2 to link NIR and chemiluminescence data, enabling the tracking of reprocessed polypropylene degradation up to the seventh recycling cycle. Through the analysis of important variables via VIP interpretation, we found that the

combination of these two analytical methods offer the possibility to evaluate both the quality of recycled polypropylene prior to its incorporation in new plastic materials and the regions of the NIR spectra that explain this observation.

Our findings corroborate the use of handheld NIR spectroscopy in a plastic sorting facility, rendering valuable insight into the material resilience to reprocessing. This approach holds promise for future applications, potentially offering additional confirmation by assessing the recycling viability of specific materials within clearly defined applications such as bags or the automotive industry.

Unbiased Mixed Variable Distance

Carlo Cavicchia

Erasmus University Rotterdam

Gower's general coefficient of similarity (Gower, 1971) provides an elegant and simple way to measure similarity between observations based on measurement on multiple variables that are of different types. That is, variables can be either numerical, binary, ordinal or categorical. We refer to the presence of variables of different types as mixed variables. Although alternative proposals exist that allow distance calculations in mixed variable contexts, Gower's proposal remains a popular choice and it is typically used with basic settings. However, the original paper actually allows for quite some implementation flexibility. Here, we use this flexibility and propose alternatives that overcome some of the shortcomings of the default implementation. In particular, using a very general framework for implementing distances for categorical data, we propose a highly adaptable measure for dissimilarity for mixed variables that can easily be implemented and customized.

Unraveling the Skillsets of Data Scientists: Text Mining Analysis of Dutch University Master Programs in Data Science and Artificial Intelligence

Zsuzsa Bakk

Leiden University

The growing demand for data scientists in the global labour market and the Netherlands has led to an increase in data science and artificial intelligence (AI) master programs offered by universities. However, there is still a lack of clarity regarding the specific skill sets of data scientists. This study aims to address this issue by employing Correlated Topic Modeling (CTM) to analyze the content of 41 master programs offered by 11 Dutch universities and an interuniversity combined program. We assess the differences and similarities in the core skills taught by these programs, determine the subject-specific and general nature of the skills, and provide a comparison between the different types of universities offering these programs. Our analysis reveals that data processing, statistics, research, and ethics are the core competencies in Dutch data science and AI master programs. General universities tend to focus on research skills, while technical universities lean more towards IT and electronics skills. Programs with a broad data science and AI focus generally concentrate on data

processing, information technology, electronics, and research. In contrast, those with a subject-specific focus prioritize statistics and ethics. This research contributes to a better understanding of the diverse skill sets of Dutch data science graduates, providing insights for employers, academic institutions, and prospective students.