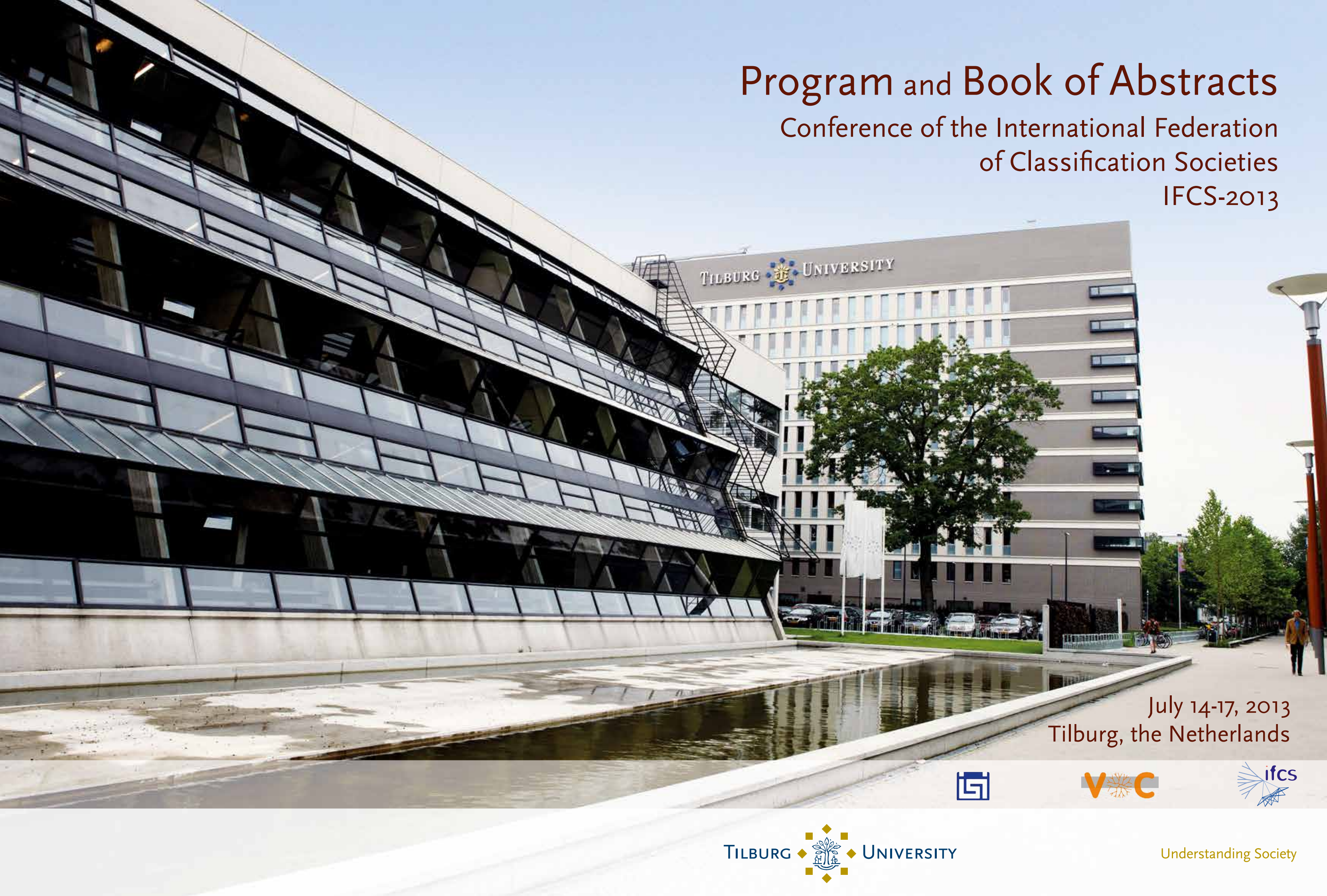


# Program and Book of Abstracts

Conference of the International Federation  
of Classification Societies  
IFCS-2013



July 14-17, 2013  
Tilburg, the Netherlands



Understanding Society

# Scientific Program

## Monday, July 15

### Plenary Invited Sessions

**Time: 09:00-10:30**

**Room: CZ115**

**Chair: Groenen** .....

- Critical Issues and Developments in High-dimensional Prediction with Biomedical Applications** 1  
Anne-Laure Boulesteix
- Flexible Model Based Clustering via the Cluster-Weighted Approach** 2  
Salvatore Ingrassia

## Monday, July 15

### Concurrent Session 1a

**Topic: Applications in marketing and social policy**

**Time: 11:00-12:20**

**Room: CZ6**

**Chair: Okada** .....

- Latent Class Models in Marketing: Trading off Classification Certainty and Costs of Data Collection.** 3  
Maurits Kaptein
- Market Segmentation based on Stated Preferences using Latent Class Models and R** 4  
Andrzej Bąk, Aneta Rybicka, and Marcin Pełka
- Multi-layer Cluster Analysis of Brand Switching Among Coffee Brands** 5  
Akinori Okada and Satoru Yokoyama
- Polish Households' Pharmaceutical Expenditures in Years 2010 – 2020 – Microsimulation Analysis with FARMMES** 6  
Agata Zoltaszek, M.A.

## Monday, July 15

### Concurrent Invited Session 1b

**Topic: Analysis of symbolic data**

**Time: 11:00-12:20**

**Room: CZ7**

**Organizer and chair: Brito** .....

---

<b>Clustering for Aggregated Symbolic Data</b>	7
Nobuo Shimizu and Junji Nakano	
<b>Factor Analysis of Distributional Data using Quantiles</b>	8
Rosanna Verde and Antonio Irpino	
<b>A Hierarchical Clustering Algorithm applied to Modal Ordinal Symbolic Data</b>	9
Carmen Bravo and José M. García-Santesmases	
<b>Constrained Clustering of Temporal Beanplot Data</b>	10
Carlo Drago	
<b>Monday, July 15</b>	
<b><u>Concurrent Invited Session 1c</u></b>	
<b>Topic: Reconsidering methodologies in inequalities indicators: the case of gender studies(session sponsored by European Association of Methodology)</b>	
<b>Time: 11:00-12:20</b>	
<b>Room: CZ8</b>	
<b>Organizer and chair: Crippa .....</b>	
<b>Gender Gap: towards a Measurement with Chain Graphical Models</b>	11
Federica Nicolussi and Fulvia Mecatti	
<b>Time To Graduation: Does Gender Make A Difference? An Analysis Of A Greek University</b>	12
Adele H. Marshall Aglaia Kalamatianou and Mariangela Zenga	
<b>Beyond indicators: a Causal Approach to Gender Statistics</b>	13
Silvia Caligaris and Fulvia Mecatti	
<b>Gender Differentials In Higher Education: Hints From A Fuzzy States Analysis</b>	14
Franca Crippa, Marcella Mazzoleni and Mariangela Zenga	
<b>Monday, July 15</b>	
<b><u>Concurrent Session 1d</u></b>	
<b>Topic: Correspondence analysis</b>	
<b>Time: 11:00-12:20</b>	
<b>Room: CZ9</b>	
<b>Chair: Le Roux .....</b>	
<b>Analysing Categorical Variables With Similar Categories: Constrained Multiple Correspondence Analysis</b>	15
Véronique Cariou and El Mostafa Qannari	
<b>Constrained Dual Scaling of Successive Categories for Detecting Response Styles</b>	16
Pieter C. Schoonees, Michel van de Velden, and Patrick J.F. Groenen	
<b>ORTHOMALS: Orthogonal Projection Of A Multiple Correspondence Solution On A Design Space</b>	17
Ralph C.A. Rippe and Willem J. Heiser	

---

<b>Squared Covariances Or Chi-Squared Statistics Based Distances</b>	18
Antoine de Falguerolles	
<b>Monday, July 15</b>	
<b><u>Concurrent Session 1e</u></b>	
<b>Topic: Latent class analysis</b>	
<b>Time: 11:00-12:20</b>	
<b>Room: CZ109</b>	
<b>Chair: Oberski .....</b>	
<b>A New Constant Memory Recursion For Hidden Markov Models</b>	19
Francesco Bartolucci and Silvia Pandolfi	
<b>Detecting Local Dependence In Binary Data Latent Class Models: Some Developments</b>	20
Daniël Oberski	
<b>Power and Sample Size Determination for Latent Class Models</b>	21
Dereje W. Gudicha, Jeroen K. Vermunt, and Fetene B. Tekle	
<b>The Bias-Adjusted Three-Step Approach To Latent Class Modeling With External Variables</b>	22
Zsuzsa Bakk, Daniel Oberski, and Jeroen K. Vermunt	
<b>Monday, July 15</b>	
<b><u>Concurrent Invited Session 1f</u></b>	
<b>Topic: Recent clustering techniques and their applications</b>	
<b>Time: 11:00-12:20</b>	
<b>Room: CZ114</b>	
<b>Organizer and chair: Kurihara .....</b>	
<b>Comparative Analysis on LDA-based Classification and Subject Categories of the Japanese Awards Database of Grants-in-Aid for Scientific Research, KAKEN</b>	23
Kei Kurakawa, Yuan Sun, and Yasumasa Baba	
<b>Prototype Identification through Archetypes</b>	24
Giancarlo Ragozini	
<b>Spatial Clustering based on Hierarchical Structure of Multidimensional Lattice Data</b>	25
Koji Kurihara and Fumio Ishioka	
<b>Research Literature Analytics through Mapping Narratives</b>	26
Fionn Murtagh	
<b>Monday, July 15</b>	
<b><u>Plenary Invited Sessions</u></b>	
<b>Time: 13:20-14:50</b>	
<b>Room: CZ115</b>	
<b>Chair: Jajuga .....</b>	
<b>Effects of Moment-to-moment Likeability Patterns on the Virality of Online Ads</b>	27
Tammo Bijmolt	

---

<b>Formal Concepts for Classification</b>	28
Bernhard Ganter	
<b>Monday, July 15</b>	
<u>Concurrent Invited Session 2a</u>	
Topic: Biostatistics & psychometrics	
Time: 15:20-16:40	
Room: CZ6	
Organizer and chair: Lee .....	
<b>Multinomial Logistic Regression Ensembles</b>	29
Hongshik Ahn	
<b>Age-specific Disease Network For The Major Disease In Korea</b>	30
Taerim Lee and Hongseok Kim	
<b>Analysis of Questionnaire Survey with Ordinal-polytomous Using the Binomial Confidence Limits</b>	31
Ueno, T., Tatsunami, S., Otaki, M., and Kuwabara, R.	
<b>Comparison Of Methods For Handling Missing Data In A Multi-Item Instrument</b>	32
I. Eekhout, H.C.W. de Vet, J.W.R. Twisk, J.P.L. Brand, M.R. de Boer, and M.W. Heymans	
<b>Monday, July 15</b>	
<u>Concurrent Session 2b</u>	
Topic: Reduced rank clustering	
Time: 15:20-16:40	
Room: CZ7	
Chair: Wilderjans .....	
<b>Common and Cluster-specific Simultaneous Component Analysis</b>	33
Kim De Roover, Marieke E. Timmerman, Batja Mesquita and Eva Ceulemans	
<b>Extending Clusterwise non-negative matrix factorization (NMF) to hierarchically organized data</b>	34
Joke Heylen, Philippe Verduyn, Iven Van Mechelen and Eva Ceulemans	
<b>Generalized Reduced Clustering Analysis</b>	35
Michio Yamamoto	
<b>Mixtures Of Factor Analyzers And Unobserved Heterogeneity In Questionnaire Data</b>	36
Robert Kapłon	
<b>Monday, July 15</b>	
<u>Concurrent Invited Session 2c</u>	
Topic: Research of IOPS Ph.D.-students	
Time: 15:20-16:40	
Room: CZ8	
Organizer and chair: Kuijpers .....	

---

<b>Estimation Methods for Categorical Marginal Models: Comparing MAEL, GEE, and GSK.</b>	37
Renske E. Kuijpers, Wicher P. Bergsma, L. Andries van der Ark, and Marcel A. Croon	
<b>Applying Multilevel Latent Class Analysis To Large-Scale Educational Assessment Data: Predicting Students' Mathematical Strategy Choices From Teachers' Instructional Practice</b>	38
Marije F. Fagginger Auer, Marian Hickendorff, and Cornelis M. van Putten	
<b>A Tuning Strategy for COSA</b>	39
Maarten M.D. Kampert and Jacqueline J. Meulman	
<b>Accuracy Of Reliability Estimates</b>	40
Pieter R. Oosterwijk, Klaas Sijtsma, and L. Andries van der Ark	
<b>Monday, July 15</b>	
<b><u>Concurrent Session 2d</u></b>	
<b>Topic: Symbolic data clustering and regression</b>	
<b>Time: 15:20-16:40 Room: CZ9</b>	
<b>Chair: Brito</b> .....	
<b>A Big Data Intensive Application System with Symbolic Data Analysis and its Implementation</b>	41
Hiroyuki Minami and Masahiro Mizuta	
<b>An Generalization Of Centre And Range Method For Fitting A Linear Regression Model To Symbolic Interval Data Using Ridge Regression, Lasso And Elastic Net Methods</b>	42
Oldemar Rodríguez	
<b>Symbolic Data Clustering. A Review</b>	44
Justyna Wilk	
<b>The Ensemble Conceptual Clustering of Symbolic Data</b>	45
Marcin Pełka	
<b>Monday, July 15</b>	
<b><u>Concurrent Invited Session 2e</u></b>	
<b>Topic: Applications in economics and business</b>	
<b>Time: 15:20-16:40</b>	
<b>Room: CZ109</b>	
<b>Organizer and chair: Pocięcha</b> .....	
<b>The Hierarchy Test Of Geographic Units based on Border Lengths</b>	46
Andrzej Sokołowski, Danuta Strahl, Małgorzata Markowska, and Marek Sobolewski	
<b>Statistical Modeling the Optimal Level of FX Reserves for Poland</b>	47
Eugeniusz Gatnar	
<b>Latent Transitions with Mixture Rasch Model of Bankruptcy Risk in the Classification of Polish Firms</b>	48
Barbara Pawełek, Józef Pocięcha, and Adam Sagan	

---

<b>Automatic Determination The Number Of Clusters In Spectral Clustering</b>	49
Marek Walesiak and Andrzej Dudek	
 <b>Monday, July 15</b>	
<b><u>Concurrent Session 3a</u></b>	
<b>Topic: Clustering algorithms</b>	
<b>Time: 17:10-18:30</b>	
<b>Room: CZ6</b>	
<b>Chair: Hennig</b> .....	
<b>A Spectral-Mean Shift Algorithm for Clustering of Symbolic Data</b>	50
Andrzej Dudek and Marcin Pełka	
<b>Asymptotics of Reduced <math>K</math>-means Clustering</b>	51
Yoshikazu Terada	
<b>Non-hierarchical Clustering Algorithm For Mixed Numerical And Categorical Three-Way Three-Mode Data</b>	52
Takahiro Umei and Hiroshi Yadohisa	
<b>Using Simulation Strategies to Test Clustering Algorithm Performances</b>	53
Marina Marino and Cristina Tortora	
 <b>Monday, July 15</b>	
<b><u>Concurrent Invited Session 3b</u></b>	
<b>Topic: recursive partitioning and application (session sponsored by IASC)</b>	
<b>Time: 17:10-18:30</b>	
<b>Room: CZ7</b>	
<b>Organizer and chair: Wilhelm</b> .....	
<b>Random Forest Variable Importance Measures: Current Developments</b>	54
Anne-Laure Boulesteix and Silke Janitza	
<b>Detecting Threshold Interactions In Binary Classification: STIMA</b>	55
Claudio Conversano and Elise Dusseldorp	
<b>A Recursive Partitioning-Based Method To Balance Covariates When Estimating Causal Effects</b>	56
Massimo Cannas, Claudio Conversano and Francesco Mola	
<b>Recursive Partitioning for Hybrid Image Classification using Captions and Image Features</b>	57
Adalbert Wilhelm	
 <b>Monday, July 15</b>	
<b><u>Concurrent Session 3c</u></b>	
<b>Topic: Applications in economics</b>	
<b>Time: 17:10-18:30</b>	
<b>Room: CZ8</b>	
<b>Chair: Markos</b> .....	
<b>Change of Aspects of Industrial Classification System from Hierarchical Structure to Network Structure</b>	58
Hiroki Furuzumi, Yoshiro Matsuda, and Yasumasa Baba	

---

<b>Econometric Models of Durable Goods' Prices: A Hedonic Approach</b>	59
Anna Król	
<b>Smart Growth Versus Economic And Social Cohesion – Econometric Panel Analysis</b>	60
Beata Bal-Domańska and Elżbieta Sobczak	
<b>Workflow Classification Based On The K-Means Partitioning</b>	61
Etienne Lord, Abdoulaye Baniré Diallo, and Vladimir Makarenkov	
<b>Monday, July 15</b>	
<b><u>Concurrent Session 3d</u></b>	
<b>Topic: R packages</b>	
<b>Time: 17:10-18:30</b>	
<b>Room: CZ9</b>	
<b>Chair: Leisch</b> .....	
<b>Functional Principal Component Analysis with R</b>	62
Malgorzata Sej-Kolasa and Mirosława Sztemberg-Lewandowska	
<b>Implementation of Time Series Methods of Forecasting in TSprediction R Package</b>	63
Tomasz Bartłomowicz	
<b>Latest developments of the RSDA: An R package for Symbolic Data Analysis</b>	64
Oldemar Rodríguez and Johnny Villalobos	
<b>Microeconometrics Multinomial Logit Models and their Implementations in MMLM R Package</b>	65
Andrzej Bąk and Tomasz Bartłomowicz	
<b>Monday, July 15</b>	
<b><u>Concurrent Session 3e</u></b>	
<b>Topic: Latent variable &amp; multilevel analysis</b>	
<b>Time: 17:10-18:30</b>	
<b>Room: CZ109</b>	
<b>Chair: Montanari</b> .....	
<b>Latent Spaces of the Product Baskets - A Hybrid Model of On-line Shopping</b>	66
Adam Sagan and Mariusz Łapczyński	
<b>Multilevel Principal Covariates Regression</b>	67
Marlies Vervloet, Wim Van den Noortgate, Katrijn Van Deun and Eva Ceulemans	
<b>Three-step Estimation Method For Discrete Micro-Macro Multilevel Models</b>	68
M. Bennink, M. A. Croon and J. K. Vermunt	
<b>Single-array SNP Genotype Classification With Semi-Parametric Log-Concave Mixtures</b>	69
Paul H.C. Eilers and Ralph C.A. Rippe	
<b>Monday, July 15</b>	
<b><u>Concurrent Invited Session 3f</u></b>	
<b>Topic: Least-squares clustering</b>	
<b>Time: 17:10-18:30</b>	

---

<b>Room: CZ114</b>	
<b>Organizer and chair: Mirkin</b> .....	
<b>On Featureless K-Means Clustering</b>	70
Sergey D. Dvoenko	
<b>Two Major Least-squares Divisive Clustering Methods: Bisecting K-Means, PDDP and in between</b>	71
E. Kovaleva and B. Mirkin	
<b>Scoring Dissimilarity between Binary Images by Aligning Series of Skeleton Primitives</b>	72
Olesya A. Kushnir and Oleg S. Seredin	
<b>Least-squares Consensus Clustering versus: (a) other Consensus Approaches and (b) K-Means</b>	73
A. Shestakov and B. Mirkin	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Session 4a</u></b>	
<b>Topic: Clustering methods</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ6</b>	
<b>Chair: Bertrand</b> .....	
<b>Combination of Several Control Charts using Dynamic Weighted Majority Algorithm</b>	74
Dhouha Mejri, Claus Weihs and Mohamed Limam	
<b>Multiplicity Within Clustering: Challenges And Unifications</b>	75
Jacques-Henri Sublemontier	
<b>Non-Isometric Transforms in Time Series Classification using DTW</b>	76
Tomasz Górecki and Maciej Łuczak	
<b>Performance of the Accelerated Hyperbolic Smoothing Clustering Method</b>	77
Adilson Elias Xavier and Vinicius Layter Xavier	
<b>STATIS Based Multiblock Clustering</b>	78
Ndèye Niang and Mory Ouattara	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Invited Session 4b</u></b>	
<b>Topic: New trends in analyzing multi-set and three-way data</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ7</b>	
<b>Organizers: Wilderjans and Ceulemans (Chair)</b> .....	
<b>Identifying Common And Distinctive Processes Underlying Multiset Data</b>	79
Katrijn Van Deun, Age K. Smilde, Henk A.L. Kiers, and Iven Van Mechelen	
<b>Fuzzy Clustering of Three-way Proximity Arrays</b>	80
Paolo Giordani and Henk A.L. Kiers	

---

<b>Principal Covariates Clusterwise Regression</b>	81
Eva Ceulemans, Eva Vande Gaer, Henk A. L. Kiers, Iven Van Mechelen, and Tom F. Wilderjans	
<b><i>Clusterwise PARAFAC To Identify Heterogeneity In Three-Way Data</i></b>	82
Tom F. Wilderjans and Eva Ceulemans	
<b>Structure-Revealing Data Fusion Model</b>	83
Evrin Acar, Anders J. Lawaetz, Morten A. Rasmussen, and Rasmus Bro	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Session 4c</u></b>	
<b>Topic: Distances and similarities</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ8</b>	
<b>Chair: Okada</b> .....	
<b>Effects of Resampling Schemes on Stability of Cluster Validation Indices</b>	84
Rainer Dangl and Friedrich Leisch	
<b>Functional Canonical Correlation Analysis</b>	85
Mirosław Krzyśko and Łukasz Waszak	
<b>Pearson's Product-Moment Correlation is a Special Case Of Cohen's Weighted Kappa</b>	86
Matthijs J. Warrens	
<b>Ternary Diagrams Based On A Probabilistic Ideal Point Model</b>	87
Mark de Rooij and Paul Eilers	
<b>The Matter Of Scale: Perceiving Distances And Proximities In The Bi-Partial Clustering Setting</b>	88
Jan W. Owsinski	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Session 4d</u></b>	
<b>Topic: Algorithms for clustering and classification</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ9</b>	
<b>Chair: Sokołowski</b> .....	
<b>Comparing Direct Estimators of the Mode</b>	90
Andrzej Sokołowski and Kamil Fijorek	
<b><i>k</i>-NN Algorithm for Instantaneous Classification</b>	91
Carmen Villar-Patiño and Carlos Cuevas-Covarrubias	
<b>Flexible Multiclass Support Vector Machines: An Approach using Iterative Majorization and Huber Hinge Errors</b>	92
G.J.J. van den Burg and P.J.F. Groenen	
<b>Power-Stress for Multidimensional Scaling</b>	93
Patrick J.F. Groenen and Jan de Leeuw	

---

<b>Variable Selection in Cluster Analysis Using Resampling Techniques: a Proposal</b>	94
Hans-Joachim Mucha and Hans-Georg Bartel	

**Tuesday, July 16**

**Concurrent Session 4e**

**Topic: Applications in risk analysis and finance**

**Time: 08:30-10:10**

**Room: CZ109**

**Chair: Cuevas** .....

<b>Adversarial Risk Analysis in Auctions</b>	95
David Banks	

<b>Gaussian Process Classification And Duration Models For Credit Risk</b>	96
Silvia Figini and Aki Vehtari	

<b>Model Averaging For Credit Risk Modelling</b>	97
Silvia Figini and Marika Vezzoli	

<b>Multiobjective Optimization Of Financing Household Goals With Multiple Investment Programs</b>	98
Lukasz Feldman, Radoslaw Pietrzyk, and Pawel Rokita	

<b>Power Of Skewness Tests In The Presence Of Fat Tailed Financial Distributions</b>	99
Krzysztof Piontek	

**Tuesday, July 16**

**Concurrent Session 4f**

**Topic: Applications in social sciences**

**Time: 08:30-10:10**

**Room: CZ110**

**Chair: Palumbo** .....

<b>Robust Clustering for Anti-Fraud Analysis</b>	100
Andrea Cerioli and Domenico Perrotta	

<b>An Extended Gravity Approach To Examining Internal Migrations. The Case Of Poland</b>	101
Justyna Wilk and Michał Pietrzak	

<b>Clustering of US counties based on their demographic structures</b>	102
Simona Korenjak-Černe, Vladimir Batagelj, Nataša Kejžar	

<b>Strategic, Motivational And Emotional Aspects Of University Study. A Latent Class Approach</b>	103
Anna Giraldo, Silvia Meggiolaro, and Elisa Visentin	

<b>The Comparative Log-Linear Analysis Of Unemployment In Poland In 2004-2011</b>	104
Justyna Brzezińska	

**Tuesday, July 16**

**President's Invited Session**

**Time: 10:40-12:10**

---

<b>Room: CZ115</b>	
<b>Chair: Dean</b> .....	
<b>Measurement of Quality in Cluster Analysis</b>	105
Christian Hennig	
<b>Resampling Methods for Exploring Cluster Stability</b>	106
Friedrich Leisch	
<b>The Effect Of Data Generation On Our Understanding Of Clustering Algorithms</b>	107
Doug Steinley	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Session 5a</u></b>	
<b>Topic: Clustering and multilevel analysis of symbolic data</b>	
<b>Time: 13:10-14:30</b>	
<b>Room: CZ6</b>	
<b>Chair: McNicholas</b> .....	
<b>CLustering Constrained Symbolic Objects Constrained By Rules</b>	108
Marc Csernel	
<b>Conceptual Clustering with Interval Representation</b>	109
Paula Brito and Géraldine Polailon	
<b>Hierarchical Symbolic Cluster Analysis with Quantile Function Representation</b>	110
Yusuke Matsui, Hiroyuki Minami, and Masahiro Mizuta	
<b>Multilevel Consumer Preference Model on Symbolic Data</b>	111
Adam Sagan, Marcin Pełka, and Aneta Rybicka	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Invited Session 5b</u></b>	
<b>Topic: advances in clustering and classification</b>	
<b>Time: 13:10-14:30</b>	
<b>Room: CZ7</b>	
<b>Organizer and chair: Nugent</b> .....	
<b>The Variance of the Adjusted Rand Index (and other properties)</b>	112
Doug Steinley	
<b>Identifying Clusters Bayesian Disease Mapping</b>	113
Nema Dean, Craig Anderson, and Duncan Lee	
<b>Classification Boundary Mapping</b>	114
Yuning He and Herbert Lee	
<b>Deduplicating Text Records by Clustering the Results of Aggregated Conditional Classifiers</b>	115
Rebecca Nugent and Samuel L. Ventura	
<b>Tuesday, July 16</b>	
<b><u>Concurrent Session 5c</u></b>	
<b>Topic: Applications in behavioral sciences</b>	

---

**Time: 13:10-14:30**  
**Room: CZ8**  
**Chair: Yamaguchi** .....

<b>Classifications of Baseball Pitching Strategies and Exploring Effects of the New Official Balls in the Japanese Professional Baseball League</b>	116
Kazunori Yamaguchi	
<b>Life Long Learning Idea on Background of Poles' Needs</b>	117
Marta Dziechciarz-Duda and Klaudia Przybysz	
<b>Migration Of Population - The Analysis With The Use Of Log-Linear Models</b>	118
Justyna Brzezińska	
<b>The Influence of Emotion Recognition and Academic Performance on Group Popularity</b>	119
Ivan Loredana	

**Tuesday, July 16**

**Concurrent Invited Session 5d**  
**Topic: Formal concept analysis**  
**Time: 13:10-14:30**  
**Room: CZ9**  
**Organizer and chair: Ganter** .....

<b>Hierarchical Classes Analysis vs. Formal Concept Analysis</b>	120
Bernhard Ganter and Cynthia V. Glodeanu	
<b>The Diversity of Pattern Structures in Formal Concept Analysis</b>	121
Aleksy Buzmakov, Sergei O. Kuznetsov, and Amedeo Napoli	
<b>Decision Aiding Software And Consensus Theory</b>	122
Florent Domenach and Ali Tayari	
<b>Experimental Comparison of Some Triclustering Algorithms</b>	123
Dmitry V. Gnatyshak, Dmitry I. Ignatov, and Sergei O. Kuznetsov	

**Tuesday, July 16**

**Concurrent Invited Session 5e**  
**Topic: Interactions in bi- and tri-additive models**  
**Time: 13:10-14:30**  
**Room: CZ109**  
**Organizers: Albers and Gower (Chair)** .....

<b>A Framework For Modeling Covariances</b>	124
Age K. Smilde, M.E. Timmerman, H.C.J. Hoefsloot, J.J. Jansen, and E. Saccenti	
<b>Biadditive Models, Alternative Estimation Procedures And Better Biplots</b>	125
Fred A. van Eeuwijk, Gerrit Gort, Sabine K. Schnabel, and Paul H.C. Eilers	
<b>Triadditive Models for Three-way Tables</b>	126
John C. Gower, Casper J. Albers, and Steffen Unkel	
<b>Three-way Candecomp/Parafac And The Diverging Components Problem</b>	127
Alwin Stegeman	

---

**Tuesday, July 16**

**Concurrent Session 5f**

**Topic: Cluster-weighted modeling**

**Time: 13:10-14:30**

**Room: CZ114**

**Chair: Ingrassia .....**

**Cluster-weighted  $t$ -factor Analyzers for Clustering of High-dimensional Data** 128

Sanjeena Dang, Antonio Punzo, Salvatore Ingrassia, and Paul D. McNicholas

**Cluster-Weighted Modeling For Time To Event Data** 129

Utkarsh J. Dang and Paul D. McNicholas

**Modeling Bivariate Mixed-Type Data with the Generalized Linear Exponential Cluster-Weighted Model** 130

Salvatore Ingrassia and Antonio Punzo

**Tuesday, July 16**

**Plenary Invited Sessions**

**Time: 15:00-15:45**

**Room: CZ115**

**Chair: Vichi .....**

**Cluster Inference using Modes** 131

Surajit Ray

**Tuesday, July 16**

**Presidential address**

**Time: 15:45-16:30**

**Room: CZ115**

**Chair: Vichi .....**

**IFCS Presidential Address**

**Classipedia: A Road Map to Help Traverse the Classification Jungle** 132

Iven Van Mechelen

**Wednesday, July 17**

**Concurrent Session 6a**

**Topic: Clustering, including ultrametric approaches**

**Time: 08:30-10:10**

**Room: CZ6**

**Chair: Diatta .....**

**A Restricted ADCLUS Type Model for Transition Matrices** 133

Tadashi Imaizumi

**Clustering Of Time Series Via A Segmentation Approach** 134

Christian Derquenne

**Looking For A Best Compromise Between The Ultrametric Supremum-Norm Approximations** 135

B. Fichet

---

<b>Ultrametric Tree Representation For Three-Way Three-Mode Data With Weights Of Variables And Occasions</b>	136
Kensuke Tanioka and Hiroshi Yadohisa	
<b>Which Movie Shall I Watch? Ultrametric Based Recommendation System</b>	137
Pedro Contreras, Fionn Murtagh, and Javier Pereira	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Invited Session 6b</u></b>	
<b>Topic: Personalized medicine by treatment-subgroup interaction</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ7</b>	
<b>Organizer : Elise Dusseldorp</b>	
<b>Chair and discussant: Willem Heiser .....</b>	
<b>Model-Based Recursive Partitioning for Detecting Interaction Effects in Subgroups</b>	138
Achim Zeileis, Torsten Hothorn, and Kurt Hornik	
<b>Predicting Individual Causal Effects (ICE)</b>	139
Xiaogang Su and Joseph Kang	
<b>A New Tool For Identifying Qualitative Treatment-Subgroup Interactions: QUINT</b>	140
Elise Dusseldorp and Iven Van Mechelen	
<b>A Comparison Of Six Sequential Partitioning Methods To Find Subgroups Involved In Treatment-Subgroup Interactions</b>	141
Lisa Doove, Elise Dusseldorp, Katrijn Van Deun, and Iven Van Mechelen	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Session 6c</u></b>	
<b>Topic: Modeling distributions and associations</b>	
<b>Time: 08:30-10:10</b>	
<b>Room: CZ8</b>	
<b>Chair: Kiers .....</b>	
<b>Automatic Bayes Factors for Comparing Variances of Two Independent Normal Distributions</b>	142
Florian Böing-Messing and Joris Mulder	
<b>Bayesian Model Selection For Evaluating Equality And Order Constraints On Correlation Matrices</b>	143
Joris Mulder	
<b>Bivariate Dependence Patterns And Copulas: Model Discrimination And Robustness</b>	144
Lianne Ippel and Johan Braeken	
<b>Posterior Predictive checking as alternative to Asymptotics and Bootstrapping in Latent Class Analysis</b>	145
Geert H. van Kollenburg, Joris Mulder, and Jeroen K. Vermunt	
<b>Statistical Modeling Of The Distribution Of Financial Returns</b>	146
Cuevas-Covarrubias C., Iñigo-Martínez J. and Rosales-Contreras J.	

---

Wednesday, July 17

**Concurrent Session 6d**

**Topic: Classification trees**

**Time: 08:30-10:10**

**Room: CZ9**

**Chair: Lausen** .....

<b>Combining Decision Trees And Stochastic Curtailment For Assessment Length Reduction Of Test Batteries Used For Classification</b>	147
Marjolein Fokkema, Niels Smits Henk Kelderman	
<b>Gaussian Tree Models For Discrimination</b>	148
Gonzalo Perez-de-la-Cruz and Guillermina Eslava-Gomez	
<b>Stochastic Curtailment Of Questionnaires For Three Level Classification: Shortening The Ces-D For Assessing Low, Moderate, And High Risk Of Depression</b>	149
Niels Smits, Matthew Finkelman, and Henk Kelderman	
<b>Tree-Based Prediction with Missing Data</b>	150
Holger Cevallos Valdiviezo, Stefan Van Aelst	
<b>Sparse Classifier Ensembles for Improved Interpretability.</b>	151
Werner Adler, Zardad Khan, Sergej Potapov and Berthold Lausen	

Wednesday, July 17

**Concurrent Session 6e**

**Topic: Classification**

**Time: 08:30-10:10**

**Room: CZ109**

**Chair: Groenen** .....

<b>A ROC-Optimised Multi-Prototype Classifier</b>	152
Mario Ziller	
<b>Classification of Rounded Shapes with Penalized Signal Regression</b>	153
Johan J. de Rooi and Paul H.C. Eilers	
<b>Classification of Topics on Twitter in Consideration of Time Series Variation</b>	154
Atsuhiko Nakayamar, Hiroyuki Tsurumi, and Junya Masuda	
<b>Classifying Real-World Data With The <math>DD\alpha</math>-Procedure</b>	155
Pavlo Mozharovskiy, Karl Mosler, and Tatjana Lange	
<b>Comparing High-Dimensional Classifiers: Abuse and Dangers of Overall Accuracy</b>	156
A. Pedro Duarte Silva	

Wednesday, July 17

**Concurrent Session 6f**

**Topic: Model-based clustering**

**Time: 08:30-10:10**

**Room: CZ114**

**Chair: McLachlan** .....

---

<b>Divisive Latent Class Modeling as a Density Estimation Tool: The Estimation Algorithm and an Application to Incomplete Data.</b>	157
Daniel W. van der Palm, L. Andries van der Ark, and Jeroen K. Vermunt	
<b>Determining the Number of Clusters in Categorical Data</b>	158
Cláudia Silvestre, Margarida Cardoso, and Mário Figueiredo	
<b>Identifying Mixtures of Mixtures Using Bayesian Estimation</b>	159
Gertraud Malsiner-Walli, Sylvia Frühwirth-Schnatter, and Bettina Grün	
<b>Logratio Methodology Applied To Model-Based Clustering</b>	160
M. Comas-Cufí, G. Mateu-Figueras and J.A. Martín-Fernández	
<b>Model-based Clustering Of Multivariate Longitudinal Data</b>	161
Laura Anderlucchi, Angela Montanari, and Cinzia Viroli	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Session 7a</u></b>	
<b>Topic: Longitudinal and multilevel analysis</b>	
<b>Time: 10:40-12:00</b>	
<b>Room: CZ6</b>	
<b>Chair: Nugent</b> .....	
<b>A Bayesian Multilevel Modeling of Longitudinal data: Application to Hygroscopic Expansion in Composite Resins</b>	162
Nasim Vahabi, Mahmood Reza Gohari, and Ali Azarbar	
<b>A New Approach To Analyse Longitudinal Epidemiological Data With An Excess Of Zeros</b>	163
A.S. Spiensma, T.R.S. Hajos, M.R. de Boer, M.W. Heijmans, and J.W.R. Twisk	
<b>A Linear Mixed Model with a Mixture of Smooth Random Effects Distributions</b>	164
Berrie Zielman	
<b>Longitudinal IRT Modelling compared with Multilevel Analysis in estimating Development Over Time In Data From Three Likert-Item Questionnaires</b>	165
R. Gorter, M.R. de Boer, M.W. Heijmans, and J.W.R. Twisk	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Invited Session 7b</u></b>	
<b>Topic: Biclustering</b>	
<b>Time: 10:40-12:00</b>	
<b>Room: CZ7</b>	
<b>Organizer: Vichi</b>	
<b>Chair: Van Mechelen</b> .....	
<b>Mutual Information, Chi-Squared And Model-Based Clustering For Co-Clustering Of Contingency Tables</b>	166
Mohamed Nadif and Gérard Govaert	
<b>Parsimonious Estimation And Testing Of Two-Way Interaction By Means Of Two-Mode Clustering</b>	167
Jan Schepers	

<b>A general Model for Two-mode Clustering</b>	168
Maurizio Vichi	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Session 7c</u></b>	
<b>Topic: Applications in medicine</b>	
<b>Time: 10:40-12:00</b>	
<b>Room: CZ8</b>	
<b>Chair: Lausen</b> .....	
<b>Comprehensive Calculations of the Sensitivity and Specificity of Diagnosis Using Bile Cytological Data</b>	169
Tatsunami S., Hayakawa C., Koike J., Hoshikawa, M., and Ueno T.	
<b>Diagnostics for the Risk Prediction of Each Type of Endoleak Formation after TEVAR Using Statistical Discriminant Analysis</b>	170
Kuniyoshi Hayashi, Fumio Ishioka, Bhargav Raman, Daniel Y. Sze, Hiroshi Suito, Takuya Ueda, and Koji Kurihara	
<b>Extension Of A Multilingual Medical Lexicon By Combined Feature Extraction Methods</b>	171
Wiebke Petersen, Denis Anuschewski, Pascal Chave, and Philipp F. Zeitz	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Invited Session 7d</u></b>	
<b>Topic: Correspondence analysis and related methods</b>	
<b>Time: 10:40-12:00</b>	
<b>Room: CZ9</b>	
<b>Organizers: Groenen (Chair) and Greenacre</b> .....	
<b>The Joy of Fuzzy</b>	172
Michael Greenacre	
<b>Fast Iterative Implementation of Correspondence Analysis</b>	173
Alfonso Iodice D'Enza, Patrick J. Groenen and Michel van de Velden	
<b>Inverse Multiple Correspondence Analysis</b>	174
Michel van de Velden, Patrick Groenen, and Wilco van den Heuvel	
<b>Tracking Association Structures in Categorical Data Flows</b>	175
Alfonso Iodice D'Enza and Angelos Markos	
<b>Wednesday, July 17</b>	
<b><u>Concurrent Invited Session 7e</u></b>	
<b>Topic: finding the number of clusters</b>	
<b>Time: 10:40-12:00</b>	
<b>Room: CZ109</b>	
<b>Organizer and chair: Hennig</b> .....	
<b>Determining the Number of Clusters: a Problem of Definition or Estimation?</b>	176
Giovanna Menardi	
<b>Enhancing The Selection Of A Number Of Clusters In Model-Based Clustering With External Qualitative Variables</b>	177
AJ.-P. Baudry, M. Cardoso, G. Celeux, M.J. Amorim, and A.S. Ferreira	

---

<b>Choosing the Number of Clusters after, before, and while Clustering</b> B. Mirkin	178
<b>Wednesday, July 17</b>	
<b><u>Plenary Invited Sessions</u></b>	
<b>Time: 13:00-14:30</b>	
<b>Room: CZ115</b>	
<b>Chair: Nadif</b> .....	
<b>Competitions in Machine Learning: the Fun, the Art, and the Science</b> Isabelle Guyon	179
<b>Playing with Data—or How to Discourage Incorrect Data Analysis</b> Klaas Sijtsma	180
<b>Wednesday, July 17</b>	
<b><u>Concurrent Session 8a</u></b>	
<b>Topic: Applications</b>	
<b>Time: 15:00-16:20</b>	
<b>Room: CZ6</b>	
<b>Chair: Bassi</b> .....	
<b>A Study on Small-Area Geographical Analysis of Residential Characteristics after the Great Hanshin-Awaji Earthquake by two Individual Differences Model</b> Mitsuhiro Tsuji, Hiroshi Kageyama and Toshio Shimokawa	181
<b>Author Identification of Japanese Classical Literature by Quantitative Analysis</b> Gen Tsuchiyama and Masakatsu Murakami	182
<b>A Latent Class Approach for Estimating Labour Market Mobility in the Presence of Multiple Indicators and Retrospective Interrogation</b> Francesca Bassi, Marcel Croon, and Davide Vidotto	183
<b>Wednesday, July 17</b>	
<b><u>Concurrent Invited Session 8b</u></b>	
<b>Topic: Non-Gaussian model-based classification</b>	
<b>Time: 15:00-16:20</b>	
<b>Room: CZ7</b>	
<b>Organizer and chair: McNicholas</b> .....	
<b>On Finite Mixtures of Skew Distributions</b> Geoff McLachlan and Sharon Lee	184
<b>Classification via Mixtures of Shifted Asymmetric Laplace and Mixtures of Generalized Hyperbolic Distributions</b> Paul D. McNicholas, Ryan P. Browne, and Brian C. Franczak	185
<b>Gaussian And Distance Based Clustering In High-Dimensional Space: Differences And Common Aspects</b> Francesco Palumbo, Cristina Tortora, and Paul McNicholas	186
<b>Clustering and Dimension Reduction using Non-Gaussian Mixtures</b> Katherine Morris and Paul McNicholas	187

---

Wednesday, July 17

**Concurrent Session 8c**

**Topic: Applications in social sciences**

**Time: 15:00-16:20**

**Room: CZ8**

**Chair: Dean** .....

**Comparison of Spatial Clusters between Suicide Data and Its Increase-decrease Rates in Japan** 188

Makoto Tomita, Takafumi Kubota, Fumio Ishioka and Toshiharu Fujita

**Detection of Spatial Clusters for High and Low Suicidal Risk Areas in Japan** 189

Takafumi Kubota, Makoto Tomita, Fumio Ishioka, Tomokazu Fujino and Hiroe Tsubaki

**Patterns of Cultural Practices and Characteristics of the Cultural Omnivore** 190

Miki Nakai

**The Structure Of Subjective Social Status In Japan: An Approach Based On Latent Class Model** 191

Yusuke Kanazawa

Wednesday, July 17

**Concurrent Invited Session 8d**

**Topic: Biplot-based visualisations and classification**

**Time: 15:00-16:20**

**Room: CZ9**

**Organizer and chair: Le Roux** .....

**Reference Set Selection for Multivariate Statistical Process Monitoring with Biplots** 193

RF Rossouw, RLJ Coetzer, and NJ Le Roux

**PLS Biplot: Another Graphical Tool for Multivariate Data** 194

Opeoluwa V.F. Oyedele and Sugnet Lubbe

**Variable Selection for Regression and PLS using Generic Algorithms and Particle Swarm Optimization: A Comparison between the Two Methods** 195

Martin Philip Kidd and Martin Kidd

**Classification with Hyperspheres** 196

Morné Lamont

Wednesday, July 17

**Concurrent Invited Session 8e**

**Topic: Combinatorial methods for hierarchical and non-hierarchical clustering**

**Time: 15:00-16:20**

**Room: CZ109**

**Organizer and chair: Brucker** .....

**Separation And Convexity Properties Of Hierarchical And Non Hierarchical Clustering** 197

Patrice Bertrand and Jean Diatta

---

<b>Latticial Approach for Perfect Phylogeny Problems</b>	198
François Brucker and Pascal Pr�ea	
<b>Some Aspects of Formal Concept Analysis in Hierarchical Classification and Data Analysis</b>	199
Mehdi Kaytoue, Sergei O. Kuznetsov, and Amedeo Napoli	
<b>Which Movie Shall I Watch? Ultrametric Based Recommendation System</b>	200
Pedro Contreras, Fionn Murtagh, and Javier Pereira	

**Wednesday, July 17**

**Concurrent Session 8f**

**Topic: Applications in genetics**

**Time: 15:00-16:20**

**Room: CZ114**

**Chair: Van Deun** .....

<b>Automatic Annotation and Classification of new Papillomavirus genomes</b>	201
Mohamed Amine Remita, Ahmed Halioui and Abdoulaye Banir� Diallo	

<b>Different Approaches To Modeling Family Data In GWAS: Application To Cannabis Use</b>	202
Camelia C. Minic�, Conor V. Dolan, Jouke-Jan Hottenga, Dorret I. Boomsma and Jacqueline M. Vink	

<b>Utilization Of Machine-Learning Methodologies In Order To Understand Complex Evolutionary And Functional Links Among Bacterial Genomes</b>	203
Olivier Poiron and Benedicte Lafay	

<b>Application of a Bayesian Artificial Neural Network to the Breast Cancer Survival Data</b>	204
Masoud Salehi and Mahmood Reza Gohari	

**Wednesday, July 17**

**Plenary Invited Sessions**

**Time: 16:50-17:50**

**Room: CZ115**

**Chair: McLachlan** .....

<b>Achieving Near-perfect Classification for Functional Data</b>	205
Peter Hall (and Aurore Delaigle)	

# Comparing High-Dimensional Classifiers: Abuse and Dangers of Overall Accuracy

A. Pedro Duarte Silva

## Abstract

Statistical classification has a respected tradition in the support of medical diagnosis. Early applications relied on classical methodologies that assumed training samples with more patients than disease predictors and understood that simple performance measures, that do not take into account disease prevalence and the different costs of negative and positive predictions, have serious limitations.

More recently, new classification methodologies have been applied to large genomic data bases where thousands of genes are measured on a few dozen patients. However, many of the studies that have evaluated these proposals employed only overall accuracy measures. This practice is potentially misleading, as it is known that changing prior probabilities and/or cost assumptions can strongly affect the relative standing of traditional classification rules.

This presentation describes a study on the consequences of comparing high-dimensional classification rules by different performance measures. It will be argued that measures based on expected utilities or decision curves, that focus on the precision of risk estimates near the optimal threshold, should be preferred to overall accuracy. Furthermore, it will be shown that when samples proportions are not close to true disease probabilities corrected by misclassification costs, the use of overall accuracy can indeed lead to incorrect rankings of high-dimensional classifiers.

## References

BAKER, S.G.; COOK, N.R., VICKERS, A. and KRAMER, B.S. (2009): Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society. A*, 172, 729–748.

DUARTE SILVA, A.P.; STAM, A. and NETER, J. (2002): The Effects of Misclassification Costs and Skewed Distributions in Two-Group Classification. *Communications in Statistics: Simulation and Computation* 31, 401–423.

## Keywords

CLASSIFIER EVALUATION, DECISION CURVES, HIGH DIMENSIONAL CLASSIFICATION, MISCLASSIFICATION COSTS