

Report from the 30th Scientific Conference: ‘Classification and Data Analysis – Methodology and Practice’

Krzysztof Jajuga,^a Grażyna Dehnel,^b Marek Walesiak^c

The 30th Scientific Conference: ‘Classification and Data Analysis – Methodology and Practice’ (the 25th Taxonomic Conference) of the Classification and Data Analysis Section of the Polish Statistical Association (SKAD) took place on 8–10 September 2021, in Poznań, Poland. The conference was organised by the Department of Statistics of the Poznań University of Economics and Business and was held online through the Microsoft Teams communication platform. Grażyna Dehnel, PhD, DSc, Assoc. Prof. was the chairman of the Organising Committee, consisting of Maciej Beręsewicz, PhD, Tomasz Klimanek, PhD, DSc, Assoc. Prof., Wojciech Roszka, PhD, Marcin Szymkowiak, PhD, DSc, Assoc. Prof., Hanna Wdowicka, PhD and Kamil Wilak, PhD. Basic information about the conference is available at: <https://SKAD2021.ue.poznan.pl>.

The following topics were addressed during the conference:

- theoretical aspects (taxonomy, discriminant analysis, linear ordering methods, multivariate statistical analysis, methods of analysing continuous variables, methods of discrete variables analysis, symbolic data analysis, graphical methods);
- practical applications (financial data analysis, marketing data analysis, spatial data analysis, other areas of data analysis application – medicine, psychology, archaeology, etc., computer application of statistical methods).

The main objective of the SKAD conference was to present the current research and to create a platform for the exchange of ideas relating to theoretical and applied aspects of classification and data analysis. This annually held forum provides an opportunity to present and promote state-of-the-art research and to indicate possible directions for its further development.

The conference featured 102 participants who are faculty members and doctoral students of the following universities and institutions: AGH University of Science and Technology, Calisia University – Kalisz, Poland, Statistics Poland, the Institute of Rural and Agricultural Development of the Polish Academy of Sciences,

^a Wrocław University of Economics and Business, ul. Komandorska 118/120, 53-345 Wrocław, Poland, e-mail: krzysztof.jajuga@ue.wroc.pl, ORCID: <https://orcid.org/0000-0002-5624-6929>.

^b Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: grazyna.dehnel@ue.poznan.pl, ORCID: <https://orcid.org/0000-0002-0072-9681>.

^c Wrocław University of Economics and Business, ul. Komandorska 118/120, 53-345 Wrocław, Poland, e-mail: marek.walesiak@ue.wroc.pl, ORCID: <https://orcid.org/0000-0003-0922-2323>.

Iowa State University, Gdańsk University of Technology, Cracow University of Technology, RTI International, Warsaw University of Life Sciences, SGH Warsaw School of Economics, Medical University of Silesia, University of Agricultural Sciences Bangalore, University of Economics in Katowice, Cracow University of Economics, Poznań University of Economics and Business, Wrocław University of Economics and Business, University of Gdańsk, Adam Mickiewicz University in Poznań, Łódź University, Poznań University of Life Sciences, University of Szczecin, University of Białystok, the Statistical Office in Łódź, the Statistical Office in Poznań, Utrecht University, University of Southampton, West Pomeranian University of Technology in Szczecin.

54 presentations introducing research results relating to the theory and applications of classification and data analysis were delivered during 3 plenary sessions and 15 parallel sessions. The sessions were chaired by Andrzej Dudek (two sessions), Elżbieta Gołata, Iwona Markowicz, Beata Bieszk-Stolorz, Dominik Rozkrut, Kamila Migdał-Najman, Andrzej Sokołowski, Marek Walesiak, Marcin Szymkowiak, Barbara Pawełek, Jacek Kowalewski, Paweł Lula, Krzysztof Najman, Agnieszka Stanimir, Grażyna Dehnel, Ewa Roszkowska and Józef Pocięcha.

Selected papers presented at the conference will be published in a book entitled 'Classification and Data Analysis – Methodology and Practice', edited by K. Jajuga, G. Dehnel, M. Walesiak, published by Springer. Below is a list of all papers presented during the conference:

Peter van der Heijden, *Multiple system estimation using covariates having missing values and measurement error: estimating the size of the Maori population in New Zealand*;

Dominik Rozkrut, *Digital transformation, data governance, data stewardship and official statistics*;

Stephanie Eckman, *Using Passive Data to Supplement or Replace Survey Data*;

Jae-Kwang Kim, *Propensity score estimation using the density ratio model under item nonresponse*;

Grażyna Trzpiot, *Longevity risk versus longevity dividend*;

Elżbieta Gołata, *Changes in demographic structures – consequences for the labour market in Central and Eastern Europe*;

Marcin Szymkowiak, Mirosław Krzyśko, Waldemar Wołyński, *Using functional distance correlation and functional Hilbert-Schmidt correlation to select features for cluster analysis of the labour market in Poland*;

Andrzej Sokołowski, Małgorzata Markowska, *Measure of cluster stability in dynamic classification*;

Dorota Rozmus, *Stability studies in an aggregated approach in taxonomy*;

Cyprian Kozyra, *Problems with exploratory and confirmatory factor analysis*;

- Lilia Karpinska, Sławomir Śmiech, *Mapping regional vulnerability to energy poverty in Poland*;
- Dwijendra Nath Dwivedi, Katarzyna Wójcik, Anilkumar Guntipalli Vemareddy, *Identification of key concerns and sentiments towards data quality and data strategy challenges using Sentiment Analysis and Topic Modelling*;
- Mateusz Marszałkowski, Maciej Beręsewicz, *Extending the Demand for Labour survey by administrative data using data integration methods*;
- Marek Walesiak, Grażyna Dehnel, Andrzej Dudek, *Dynamic approach in relative taxonomy and robust measures of central tendency*;
- Marta Kusterka-Jefmańska, Bartłomiej Jefmański, Ewa Roszkowska, *Application of the Intuitionistic Fuzzy Synthetic Measure in subjective quality of life measurement based on aggregated data*;
- Marcin Pełka, *Outlier identification for symbolic data with the application of the DBSCAN algorithm*;
- Victor Shevchuk, *Determinants of the real estate prices in Poland*;
- Anna Gdakowicz, Ewa Putek-Szeląg, *Attributes affecting exposure time of a residential property*;
- Radosław Trojanek, *Are transaction prices the best source of information for constructing apartment price indexes on the primary market?*;
- Iwona Markowicz, Beata Bieszk-Stolorz, *Changes in the share prices of macro-sector companies on the Warsaw Stock Exchange in response to the COVID-19 pandemic*;
- Barbara Batóg, Katarzyna Wawrzyniak, *Comparison of the results of linear ordering of companies listed on the Warsaw Stock Exchange by means of different propositions of transformations of nominants into stimulants*;
- Dominik Krężolek, *The impact of the COVID-19 pandemic on extreme risk in the metals market – application of GARCH-type models with alpha-stable distributed error*;
- Michał Pietrzak, Tomasz Józefowski, Tomasz Klimanek, Andrzej Młodak, *Optimisation of the selection of statistical disclosure control methods on the example of microdata from the Polish survey of accidents at work*;
- Romana Głowicka-Wołoszyn, Andrzej Wołoszyn, *Spatial effects in regional inequality analysis of own income potential among Polish gminas*;
- Radosław Murkowski, *Multidimensional analysis of the excessive number of deaths related to the COVID-19 pandemic in European countries by gender and age*;
- Agata Majkowska, Kamila Migdał-Najman, Krzysztof Najman, Katarzyna Raca, *Graphic characters as Twitter age group identifiers*;
- Beata Bieszk-Stolorz, Krzysztof Dmytrów, Sebastian Majewski, Wojciech Zbaraszewski, *Application of random forests in the study of the differences in the perception of the neighbourhood of national parks in the Pomerania Euroregion*;

- Iwona Bąk, Katarzyna Cheba, *Fuzzy cognitive maps as a tool for structuring new research problems*;
- Izabela Michalska-Dudek, Andrzej Dudek, *Deep neural network model for predicting ROPO (Research Online, Purchase Offline) behaviour of tourists*;
- Michel Voss, Maciej Beręsewicz, *Detection of solar panels using deep learning methods*;
- Urszula Cieraszevska, Paweł Lula, Magdalena Talaga, *Successes and failures of scientific journals and their determinants*;
- Urszula Cieraszevska, Anna Drabina, Ewelina Paluch, Janusz Tuchowski, *The co-authorship study among the Cracow University of Economics employees in years 2010–2020*;
- Hanna Wdowicka, *Multilevel modelling of the foot placement control law proposed by Hof in different conditions of walk*
- Ewelina Paluch, Janusz Tuchowski, Katarzyna Wójcik, Marcela Zembura, *Analysis of the importance of IT issues in scientific publications in the field of medicine in the years 2000–2020*;
- Jan Kubacki, Alina Jędrzejczak, *Determinants and estimation of private farm income using methods of small area estimation and multivariate statistical analysis*;
- Alina Szkop, *Estimation of variables from the DG-1 survey using time series analysis*;
- Paweł Lańduch, *Estimation of selected variables in enterprise statistics using mass imputation*;
- Ewa Kowalka, *'A picture is worth more than a thousand words' – graphical presentation of numerical data*;
- Barbara Batóg, Jacek Batóg, *Classification of local administrative units in the period 2006–2018: a spatial approach*;
- Agnieszka Stanimir, *Perception of climate change – differences between the Y and BB generation*;
- Marcin Pełka, Andrzej Dudek, *Symbolic data analysis as a tool for credit fraud detection*;
- Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski, *Employment-related commuting to provincial capital cities and their functional areas in 2016*;
- Mirosława Kaczmarek, *Cash and non-cash payments for in-store purchases during the COVID-19 pandemic – similarities and differences between Generation X and Generation Y consumers*;
- Aleksandra Łuczak, Sławomir Kalinowski, *Measuring subjective poverty: methodological and application aspects*;
- Iwona Markowicz, Paweł Baran, *The use of the duration analysis in the study of the time of running the activity of Polish enterprises on the intra-Community market*;
- Wioletta Grzenda, *Survival trees and direct adjusted survival curve – prediction in duration analysis*;

Marcin Salamaga, *Application of survival analysis methods in the study of foreign divestment in Poland*;

Dorota Żebrowska-Suchodolska, *Similarity of investment funds during the pandemic – stock and bond funds*;

Wojciech Roszka, *Statistical data integration in scientific productivity research*;

Maciej Beręsewicz, Dagmara Nikulin, Marcin Szymkowiak, Kamil Wilak, *COVID-19 and the gig economy in Poland*;

Ewa Łaźniewska, Tomasz Górecki, Klaudia Plac, *Regional labour markets as a result of the impact of the COVID-19 pandemic – a Polish-German borderland case study*;

Aleksandra Matuszewska-Janica, *The impact of the COVID-19 pandemic on the labour market in European Union countries*;

Marcin Szymkowiak, Kamil Wilak, Tomasz Józefowski, *Estimation of selected variables associated with the labour market in the light of non-response*;

Olga Kutera, *Fine wine in risk minimising portfolios*.

Members of the Classification and Data Analysis Section of the Polish Statistical Association (SKAD PTS) held their annual meeting on the first day of the conference. The meeting was chaired by Krzysztof Jajuga, PhD, DSc, ProfTit, and its agenda included the following items:

- report on SKAD activities;
- information about planned domestic and international conferences;
- organisation of SKAD conferences in 2022 and 2023;
- election of the representative of SKAD PTS in the IFCS Council for the 2022–2025 term;
- other issues.

The meeting was opened by Prof. Krzysztof Jajuga.

The report describing the activities undertaken by SKAD was presented by the Secretary of the SKAD Council, Barbara Pawełek, PhD, DSc, Assoc. Prof. at Cracow University of Economics. As reported, SKAD had at that moment 232 members (the section's by-laws and membership applications are available on the SKAD website).

Subsequently, Prof. Pawełek presented information about a book containing papers delivered during the previous SKAD conference (which was held in Sopot on 7–9 September 2020), adding that the report from that conference could be found in issue 3/2020 of the Statistical Review.

Prof. Barbara Pawełek introduced the details concerning conferences held in 2021: DSSV (Data Science, Statistics & Visualisation) – ECDA (the European Conference on Data Analysis), held online on 7–9 June; an IFCS online event – 7 September, featuring Grażyna Trzpiot, PhD, DSc, ProfTit from the University of Economics in Katowice as the keynote speaker; 13th Scientific Meeting of the

Classification and Data Analysis Group (ClaDAG), held online on 9–11 September (with two papers delivered by Polish authors).

The following conferences are scheduled to take place in late 2021 and in 2022: the 39th International Conference on Multivariate Statistical Analysis (MSA 2021), 8–10 November 2021 in Łódź; the 15th International Conference in memory of Professor Aleksander Zeliaś on Modelling and Forecasting Socio-economic Phenomena, 9–12 May 2022 in Zakopane; the 17th conference of the International Federation of Classification Societies (IFCS 2022), Porto, Portugal, 19–23 July 2022 (<https://ifcs2022.fep.up.pt/>).

The next point on the agenda concerned the organisation of future SKAD conferences. After a discussion, the participants agreed that a decision on this matter would be made later.

In the next part of the meeting the participants elected a representative of SKAD PTS in the IFCS Council for the 2022–2025 term. The election was presided over by Prof. Andrzej Dudek. Dr. Kamil Wilak was appointed the scrutineer. After Prof. Andrzej Dudek asked the participants of the videoconference to put forward their candidates and Prof. Marek Walesiak proposed Krzysztof Jajuga, who was the only candidate and agreed to stand as one. The scrutineer conducted a secret vote by means of the Doodle application. 23 members of SKAD PTS cast 23 valid votes, all in favour.

In the absence of any further points on the agenda, Prof. Krzysztof Jajuga closed the meeting.

On the last day of the conference, at the closing session, Prof. Marek Walesiak informed all the participants about the election of Krzysztof Jajuga as the representative of SKAD PTS in the IFCS Council and the possibility of having conference papers published in the Springer monograph. Prof. Krzysztof Jajuga thanked the conference organisers and all the participants and invited everyone to the next SKAD conference.