LII MIĘDZYNARODOWE COLLOQUIUM BIOMETRYCZNE

SZAMOTUŁY 10-13 WRZEŚNIA 2023

STRESZCZENIA

Ministerstwo Edukacji i Nauki

Projekt dofinansowany ze środków budżetu państwa, przyznanych przez Ministra Edukacji i Nauki w ramach Programu „Doskonała nauka II" – moduł: wsparcie konferencji naukowych

Organizatorzy

Polskie Towarzystwo Biometryczne
Katedra Zastosowań Matematyki i Informatyki, Uniwersytet Przyrodniczy w Lublinie
Katedra Metod Matematycznych i Statystycznych, Uniwersytet Przyrodniczy w Poznaniu

Komitet Naukowy

Andrzej Bochniak
Ivette Gomes
Małgorzata Graczyk
Zofia Hanusz
Marcin Kozak
Maria Kozłowska
Mirosław Krzyśko
Agnieszka Kubik-Komar
Stanisław Mejza
Manuela Neves
Amilcar Oliveira
Teresa Oliveira
Dulce Pereira
Paulo Canas Rodrigues
Leszek Sieczko
Mirosława Wesołowska-Janczarek
Waldemar Wołyński
Dariusz Załuski

Komitet Organizacyjny

Przewodnicząca	Urszula Bronowicka-Mielniczuk
Sekretarz	Monika Różańska-Boczula
Skarbnik	Anna Szczepańska-Álvarez
	Magdalena Ćwiklińska
	Dorota Domagała
	Elżbieta Kubera
	Małgorzata Szczepanik

Edycja

Monika Różańska-Boczula
Dorota Domagała
Jacek Robert Mielniczuk

SPIS TREŚCI

Downside risk aversion and input use of crop farmers in Estonia

OLHA ALEKSANDROVA[1*], ANTS-HANNES VIIRA[1,2]

[1]Estonian University of Life Sciences, Tartu, Estonia

[2]The Centre of Estonian Rural Research and Knowledge, Agricultural Research Department, Tartu, Estonia

[*]olha.aleksandrova@emu.ee

This paper studies the linkages between agricultural land use diversity, farm profitability and risk management in Estonian crop farms. A flexible moment-based approach (Antle, 1987) is used to analyse the impact of agricultural land use diversity on the mean, variance and skewness of farm profitability. Further, we analyse profitability with and without agricultural subsidies. We account for downside risks in farmers' decision making processes, which extents currently used bio-economic modeling approaches that address critical inputs in the crop production use (Di Falco and Chavas, 2006; Finger, 2013).

Using farm-level data from Estonia, econometric evidence shows that agricultural land use diversity can decrease farm profitability and reduce risk exposure. The empirical results for the expected profit indicate that agricultural land use diversity reduces farm profitability, but both critical inputs for crop production (pesticides and fertilizers) increase profitability. We furthermore find that fertlizers and pesticides increase the variance of profit and decreases its skewness, i.e. increases downside risk. In contrast, crop diversity is found to reduce both the variance and the downside risk of profitability.

Furthermore, higher agricultural land use diversity levels can reduce downside risk exposure (e.g. the risk of crop failure). The results of this paper provide useful insights about the linkages between farm profitability, risk exposure and agricultural land use diversity.

Keywords: agricultural land use diversity, profitability, fertilizers, pesticides

References

Antle, J. 1987. *Econometric estimation of producers' risk attitudes*. Am. J. Agric. Econ. 69, 509e522.

Di Falco, S., Chavas, J.-P. 2006. *Crop genetic diversity, farm productivity and the management of environmental risk in rainfed agriculture*. Eur. Rev. Agric. Econ. 33, 289e314.

Finger, R. 2013. *Expanding risk consideration in integrated models - the role of downside risk aversion in irrigation decisions*. Environmental Modelling & Software 43: 169-172 .

Acknowledgements

We thank Ms. Marju Aamisepp from the Estonian Agricultural Research Centre for providing the FADN data.

Incomplete split-block-plot design generating by orthogonally supplemented block designs

KATARZYNA AMBROŻY

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

katarzyna.ambrozy@up.poznan.pl

The construction of incomplete split-block-plot (SBP) design for three factor experiments is considered. Orthogonally supplemented PEB block designs with at most (m+1) – classes of efficiency are used in the construction procedure. The split-block-plot design is incomplete in terms of both row treatments and column treatments, while is complete in relation to the subplot treatments only. There is considered experiment with an orthogonal block structure. Stratum efficiency factors are given when the split-block-plot design is generally balanced. The proposed constructing method is illustrated with a numerical example.

Keywords: general balance, supplemented split-block-plot design, stratum efficiency factors

References

Ambroży K., Mejza I. 2006. Doświadczenia trójczynnikowe z krzyżową i zagnieżdżoną strukturą poziomów czynników. Wyd. Polskie Towarzystwo Biometryczne and PRODRUK, Poznań (in Polish).

Ambroży K., Mejza I. 2011. *Statistical properties of some supplemented split-split-plot design*. Colloquium Biometricum 41, 165-174.

Ambroży K., Mejza I. 2013. *A method of constructing incomplete split-split-plot designs supplemented by whole plot and subplot standards and their analysis*. Colloquium Biometricum, 43, 59-72.

Caliński, T., Kageyama, S. 2000. Block Designs: A Randomization Approach, Volume I: Analysis. Lecture Notes in Statistics 150, Springer-Verlag, New York.

Caliński, T., Kageyama, S. 2003. Block Designs: A Randomization Approach, Volume II: Design. Lecture Notes in Statistics 170, Springer-Verlag, New York.

Gomez K.A, Gomez A.A. 1984. Statistical procedures for agricultural research. Wiley, New York.

Houtman A.M., Speed T.P. 1983. *Balance in designed experiments with orthogonal block structure*. Ann. Statist., 11, 1069-1085.

Puri P.D., Nigam A.K., Narain P. 1977. *Supplemented block designs*. Sankhyā, Ser. B, 39, 189-195.

Resampling-based inference for multivariate data with missing values

LUBNA AMRO*, MARKUS PAULY

Department of Statistics
TU Dortmund University, Germany

*lubna.amro@tu-dortmund.de

Repeated measure designs and split plot plans are widely employed in scientific and medical research. The analysis of such designs is typically based on MANOVA models, requiring complete data, and certain assumptions on the underlying parametric distribution, such as normality or covariance matrix homogeneity. Several nonparametric multivariate methods have been proposed in the literature. They overcome the distributional assumptions, but the issue of missing data remains. The aim is to develop asymptotic correct procedures that are capable of handling missing values without assuming normality and allowing for covariance matrices that are heterogeneous between groups. This is achieved by applying a proper resampling method in combination with quadratic form-type test statistics. An extensive simulation study is conducted, exemplifying the tests for finite sample sizes under different missingness mechanisms.

Keywords: multivariate data, bootstrap, non-normal data, heteroscedasticity

References

Amro, L., Konietschke, F., Pauly, M. 2023. *Incompletely observed nonparametric factorial designs with repeated measurements: A wild bootstrap approach*. Submitted to Biometrical Journal. Preprint:arXiv:2102.02871.

Friedrich, S., Pauly, M. 2018. *MATS: Inference for potentially singular and heteroscedastic MANOVA.* Journal of Multivariate Analysis, 165, 166–179.

Amro, L. 2022. *Resampling-based inference methods for repeated measures data with missing values.* http://dx.doi.org/10.17877/DE290R-22827.

Munzel, U., & Brunner, E. 2000. *Nonparametric methods in multivariate factorial designs.* Journal of Statistical Planning and Inference, 88(1), 117-132.

Acknowledgements

The authors acknowledge DFG's support (No. PA 2409/3-2) and funding from the "From Prediction to Agile Interventions in the Social Sciences (FAIR)" project under the "Profilbildung 2020" program by the Ministry of Culture and Science of Northrhine Westphalia.

Assessment of early, mid-early and mid-late soybean (Glycine max) varieties in Northern Poland

KONRAD BANAŚ[1*], AGNIESZKA OSIECKA[2], TOMASZ LENARTOWICZ[2],
AGNIESZKA ŁACKA[1], HENRYK BUJAK[2,3], MARCIN PRZYSTALSKI[2]

[1]Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Poland

[2]Research Centre for Cultivar Testing (COBORU), Poland

[3]Department of Genetics, Plant Breeding and Seed Production,
Wrocław University of Enviromental and Life Sciences, Poland

[*]konrad.banas@up.poznan.pl

Soybean is a short-day crop with high protein and oil content. It is primarily used to feed animals but is also used for human consumption. Over recent years, soybean production has increased. According to the United Nations Food and Agriculture Organization (FAO), in 2018, the global area of soybean cultivation was 124,921,956 ha, and in the EU, it was 955,670 ha.

In Poland, new varieties of important crops prior to registration are assessed in value-for-cultivation-and-use (VCU) trials and next in post-registration variety trials. Based on the results from the post-registration trials, a recommendation for farmers is issued. In Polish soybean trials the compulsory characteristics are yield, thousand grain weight (TGW), fat and protein content.

It is believed that the cultivation of soybean in Poland will increase in the upcoming years. For this reason, it is important to introduce stable varieties. Usually, stability of agronomic traits is assessed in multi-environment trials. A common approach is to analyze these trials using two-stage approach (see, e.g., Caliński et al., 2017; Damesa et al., 2017; Piepho et al., 2012), where each combination of year and site is treated as environment.

In the present study, the varietal means for four traits were analyzed using the additive main effects and multiplicative interaction model (AMMI, Gauch). The AMMIκ model was fitted using the statgenGxE package in the statistical program R. The data set consists of aforementioned variety means from soybean trials performed in the years 2020-2022 in Northern Poland. The trials were conducted in nine sites, there were thirteen varieties. The values of means and the weighted averages of the absolute scores (WAAS) were calculated,

which allowed the genotype selection indexes to be evaluated and ranked. The WAAS scores were obtained to assess stability. Moreover, the W_iTOP3 indexes were quantified, and the mega-environments were created. The W_iTOP3 indexes were obtained to assess adaptability.

Keywords: soybean; northern Poland; stability, AMMI model

References

Caliński, T.; Czajka, S.; Kaczmarek, Z.; Krajewski, P.; Pilarczyk, W.; Siatkowski, I.; Siatkowski, M. 2017. *On mixed model analysis of multi-environment variety trials: a reconsideration of the one-stage and the two-stage models and analyses.* Stat. Pap. 58, 433–465.

Damesa, T. M.; Möhring, J.; Worku, M.; Piepho, H.P. 2017. *One step at a time: stage-wise analysis of a series of experiments.* Agron. J. 109, 845–857.

Gauch, H.G. 1988. *Model selection and validation for yield trials with interaction.* Biometrics 44, 705–715.

Gauch, H.G. 1992. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Elsevier, Amsterdam.

Księżak, J.; Bojarczuk, J. 2022. *The seed yield of soybean cultivars and their quantity depending on sowing term.* Agronomy, em 12, 1066.

Piepho, H.P.; Möhring, J.; Schulz-Streeck, T.; Ogutu, J.O. 2012. *A stage-wise approach for analysis of multienvironment trials.* Biom. J. 54, 844–860.

van Rossum B-J, van Eeuwijk F, Boer M, Malosetti M. 2021. Package 'statgenGxE.' In: CRAN. https://cran.r-project.org/ 455 web/packages/ statgenGxE/index.html.

Acknowledgements

This research was stimulated by a co-operation with the Research Centre for Cultivar Testing (Słupia Wielka, Poland), which kindly provided the data.

Glutathione and copper ions as critical factors of green plant regeneration efficiency of triticale in vitro anther culture

PIOTR T. BEDNAREK [1], RENATA ORŁOWSKA [1], DARIUSZ R. MAŃKOWSKI [1*],
JANUSZ ZIMNY [1], KRZYSZTOF KOWALCZYK [2], MICHAŁ NOWAK [2],
JACEK ZEBROWSKI [3]

[1]Plant Breeding and Acclimatization Institute-National Research Institute, Radzików, Poland

[2]Institute of Plant Genetics, Breeding and Biotechnology, University of Life Sciences in Lublin, Poland

[3]Institute of Biology and Biotechnology, University of Rzeszow, Poland

*d.mankowski@ihar.edu.pl

Plant tissue culture techniques are handy tools for obtaining unique plant materials that are difficult to propagate or important for agriculture. Homozygous materials derived through in vitro cultures are invaluable and significantly accelerate the evaluation of new varieties, e.g., cereals. The induction of somatic embryogenesis/androgenesis and the regeneration and its efficiency can be influenced by the external conditions of tissue culture, such as the ingredients present in the induction or regeneration media. We have developed an approach based on biological system, molecular markers, Fourier Transform Infrared spectroscopy, and structural equation modeling technique to establish links between changes in sequence and DNA methylation at specific symmetric (CG, CHG) and asymmetric (CHH) sequences, glutathione, and green plant regeneration efficiency in the presence of variable supplementation of induction medium with copper ions. The methylation-sensitive Amplified Fragment Length Polymorphism was used to assess tissue culture-induced variation, Fourier Transform Infrared spectroscopy to describe the glutathione spectrum, and a structural equation model to develop the relationship between sequence variation, de novo DNA methylation within asymmetric sequence contexts, and copper ions in the induction medium, as well as, glutathione, and green plant efficiency. An essential aspect of the study is demonstrating the contribution of glutathione to green plant regeneration efficiency and indicating the critical role of copper ions in influencing tissue culture-induced variation, glutathione, and obtaining green

regenerants. The model presented here also has practical implications, showing that manipulating the concentration of copper ions in the induction medium may influence cell function and increases green plant regeneration efficiency.

Keywords: androgenesis, copper, glutathione, regeneration efficiency, structure equation modeling, triticale

References

Arbuckle, J. L. 2014. "Amos". Version 27.0 Edn. (Chicago, IL: IBM SPSS).

Bentler, P.M. 2000. Rites, Wrongs, and Gold in Model Testing. Structural Equation Modeling: A Multidisciplinary Journal 7, 82–91. https://doi.org/10.1207/S15328007SEM0701_04

Bollen, K. 1986. Sample size and bentler and Bonett's nonnormed fit index. Psychometrika 51, 375–377. https://doi.org/10.1007/BF02294061

IBMCorp. 2021. "IBM SPSS Statistics for Windows". 28.0.0 Edn. IBMCorp, Amronk, NY.

Kenny, D. A., and McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Struct. Equ. Model. Multidiscip. J. 10, 333–351. doi: https://doi.org/10.1207/S15328007SEM1003_1

Kozak, M., Kang, M.S. 2006. Note on modern path analysis in application to crop science. Communications i Biometry and Crop Science 1, 32–34.

Mac Callum, R. C., Browne, M. W., and Sugawara, H. M. 1996. Power analysis and determination of sample size for covariance structure modeling. Psychol. Methods 1, 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Machczyńska, J., Orłowska, R., Zimny, J., and Bednarek, P. T. 2014. Extended met AFLP approach in studies of the tissue culture induced variation (TCIV) in case of triticale. Mol. Breed. 34:845. https://doi.org/10.1007/s11032-014-0079-2

Mańkowski, D.R. 2013. Modele równań strukturalnych SEM w badaniach rolniczych, Monografie i Rozprawy Naukowe IHAR-PIB. IHAR-PIB, Radzików.

Metakovsky, E. V., Novoselskaya, A. Y., and Sozinov, A. A. 1987. Problems of interpreting results obtained in studies of somaclonal variation in gliadin proteins in wheat. Theor. Appl. Genet. 73, 764–766. doi: 10.1007/bf00260788Suchomel C., Spinelli R., Magagnotti N. 2012. *Productivity of processing hardwood from coppice forests*. Croatian Journal of Forest Engineering 33 (1) 39-47.

Orłowska, R., Bednarek, P. T. 2020. Precise evaluation of tissue culture-induced variation during optimisation of in vitro regeneration regime in barley. Plant Mol. Biol. 103, 33–50. https://doi.org/10.1007/s11103-020-00973-5

Shimizu, S., Kano, Y. 2008. Use of non-normality in structural equation modeling: Application to direction of causation. Journal of Statistical Planning and Inference 138, 3483–3491. https://doi.org/10.1016/j.jspi.2006.01.017.

Acknowledgements

This research was funded by the Ministry of Agriculture and Rural Development, Poland (grant no. HORhn-801-PB-22/15–18).

Data visualisation in R - from simple graphs to animations

URSZULA BRONOWICKA-MIELNICZUK

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

urszula.bronowicka@up.lublin.pl

Data visualization is essential for effective statistical analysis and decision making based on experimental results. In the words of John Tukey, considered by some as one of the fathers of exploratory data analysis:

"*The greatest value of a picture is when it forces us to notice what we never expected to see*".

Data visualization skills are an intrinsic part of the modern researcher's workshop. The variety of statistical methods and programming environments available poses the dilemma of choosing an appropriate method to support statistical analysis.

The R programming environment is a universal and increasingly powerful tool in this area. The *ggplot2* package is the perfect starting point for our journey to learn about key statistical and plotting tools. Its relatively simple syntax makes it easy to implement basic graphical solutions and advanced add-ons that include animation elements.

Keywords: R environment, ggplot2, data visualization, animation

References

Pedersen T., Robinson D. 2022. gganimate: A Grammar of Animated Graphics. https://gganimate.com, https://github.com/thomasp85/gganimate

Tukey J.W. 1977. Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts: Addison-Wesley.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org/

Zrozumienie zależności między zmiennymi jakości ziarna a plonem pszenicy zwyczajnej: implikacje dla zaleceń odmianowych i środowiskowych

ADRIANA DEREJKO

Instytut Rolnictwa, Katedra Biometrii

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

adriana_derejko@sggw.edu.pl

Ocena i zrozumienie wpływu głównych i interakcyjnych efektów na wybrane zmienne jakościowe pszenicy zwyczajnej pozwoli na optymalizację rekomendacji odmian pszenicy zwyczajnej do konkretnych celów wykorzystania ziarna. Jednakże, jeśli wybrana zmienna jakościowa jest pożądana dla konkretnego zastosowania pszenicy, może to być ważniejsze niż stabilny plon odmiany.

Przeprowadzono analizy CART w celu określenia różnic i podobieństw w ocenie wpływu efektów głównych i interakcyjnych na kształtowanie się zmiennych jakościowych ziarna i plonu oraz oceny znaczenia wpływu środowiska lub odmiany na określone cechy ziarna pszenicy. Ponadto wyjaśniono związek między zmiennymi jakościowymi, plonem, glebą i warunkami klimatycznymi, a także efektami genetycznymi.

Uzyskane wyniki umożliwiają wyjaśnienie związku przyczynowo-skutkowego między cechami jakościowymi ziarna pszenicy zwyczajnej, a zmiennymi warunkami pogodowymi i glebowymi oraz znacząco przyczynią się do poprawy zaleceń odmianowych, nie tylko w klimacie umiarkowanym.

Słowa kluczowe: pszenica zwyczajna, odmiana, środowisko, rekomendacja

Estimation of unknown parameters in the spring balance weighing designs with positive correlated errors

MAŁGORZATA GRACZYK*, BRONISŁAW CERANKA

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

*malgorzata.graczyk@up.poznan.pl

We are examining the issue of determining the spring balance weighing designs under requirement of D-optimality. In addition, we are investigating these designs assuming that the measurements are positively correlated and have equal variances. In this context, we outline a method for augmenting regular D-optimal spring balance weighing designs with a maximum of three additional runs to achieve a highly D-efficient spring balance weighing design.

Keywords: D-optimality, plan of experiment, spring balance weighing design

References

Bulutoglu D.A., Ryan K.J. 2009. *D-optimal and near D-optimal 2_k fractional factorial designs of resolution V*. Journal of Statistical Planning and Inference 139, 16–22.

Ceranka B., Graczyk M. 2014. *The problem of D-optimality in some experimental designs*. International Journal of Mathematics and Computer Application Research *4, 11-18.*

Ceranka B., Graczyk M. 2018. *Highly D-efficient weighing designs for an even number of objects*. Revstat Statistical Journal 16(4), 475–486.

Harville D.A. 1997. Matrix Algebra from Statistician's Perspective. Springer-Verlag, New York Inc.

Table-eggs production in the light of animal welfare

DAVID HAMPEL[1*], MARKÉTA MATULOVÁ[1]
MARTINA LICHOVNÍKOVÁ[2], JITKA JANOVÁ[1]

[1] Department of Statistics and Operation Analysis
Mendel University in Brno

[2]Department of Animal Breeding
Mendel University in Brno

[*]david.hampel.uso@mendelu.cz

This presentation investigates how animal welfare requirements affect economic performance of table-eggs producers in EU, with a specific focus on the Czech Republic and other Central European countries. The original micro data on table-eggs producers were collected using Orbis database, characteristics of laying hens farming were sourced from Eurostat and internal sources. Data envelopment analysis was used for efficiency assessment considering animal welfare as an additional aspect of table-eggs production. The results indicate a weak negative relationship between the efficiency of table-egg producing enterprises and the share of free-range hen keeping at the country level. Detailed analysis at the farm level in the Czech Republic did not show such a dependence. It can therefore be concluded that the switch from enriched cages to free-range laying hen keeping is not problematic for companies, the costs will probably be paid by the final consumers.

Keywords: Animal welfare, DEA, Efficiency, Poultry industry, Table-eggs producers

References

Abin R., Laca A., Laca A., Diaz M., 2018. Environmental assessment of intensive egg production: A Spanish case study. Journal of Cleaner Production 179: 160–168.

Allendorf J. J., Wettemann P. J. C., 2015. Does animal welfare influence dairy farm efficiency? A two-stage approach. Journal of Dairy Science 98 (11): 7730–7740.

Simar L., Wilson P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136 (1): 31–64.

Tremetsberger L., Winckler C., Kantelhardt J., 2019. Animal health and welfare state and technical efficiency of dairy farms: possible synergies. Animal Welfare 28 (3): 345–352.

Modelling long-term stability of winter wheet varieties in the Czech Republic

JIŘÍ HARTMANN, MARTIN TLÁSKAL*

Central Institute for Supervising and Testing in Agriculture
National Plant Variety Office, Brno, Czech Republic
* martin.tlaskal@ukzuz.cz

The presented work deals with the possibility of verifying the long-term stability of varieties. A model is constructed which includes the factor of variety "aging", genetic and non-genetic trend. The properties of the model and the problems associated with estimation are discussed. An example is given of the application of the model to yield data for winter wheat varieties in the Central Institute of Supervising and Testing in Agriculture experiments from 1998 to 2022.

Classification of alive and dead tree species using multi-temporal ALS data and CIR imagery in the mixed old-growth forest in Poland

AGNIESZKA KAMIŃSKA[1], MACIEJ LISIEWICZ[1], BARTŁOMIEJ KRASZEWSKI[1], ŁUKASZ KUBERSKI[2], KAMIL PILCH[2], KRZYSZTOF STEREŃCZAK[1]

[1]Department of Geomatics
Forest Research Institute, Poland

[2]Department of Natural Forests
Forest Research Institute, Poland

[*]a.kaminska@ibles.waw.pl

Tree species classification is important for a variety of environmental applications, including biodiversity monitoring, wildfire risk assessment, ecosystem services assessment, and sustainable forest management. Moreover, the mapping of standing dead trees after pest infestation is becoming increasingly important. In our study we used a fusion of three remote sensing (RM) datasets including ALS (leaf-on and leaf-off) and colour-infrared (CIR) imagery (leaf-on), to classify different coniferous and deciduous tree species, including dead class, in a mixed temperate forest in Poland. We used intensity and structural variables from the ALS data and spectral information derived from aerial imagery for the classification procedure carried out using Random Forest algorithm. Additionally, we tested the differences in classification accuracy of all the variants included in the data integration. The highest levels of accuracies were obtained for classification based on point clouds from both seasons and including image information. Mean values of overall accuracy and kappa were equal to 82% and 0.80, respectively. Analysis of the leaf-on and leaf-off alone is not sufficient to identify individual tree species due to their different discriminatory power. Classification based on both point clouds was found to give satisfactory and comparable results to classification based on combined information from all three sources.

Keywords: tree species classification; airborne laser scanning (ALS); colour-infrared (CIR) aerial images; multi-temporal data; individual tree; random forest (RF)

References

Breiman, L. Random Forests. *Mach. Learn.* 2001, *45*, 5–32, doi:10.1023/A:1010933404324.

Kamińska, A.; Lisiewicz, M.; Stereńczak, K.; Kraszewski, B.; Sadkowski, R. Species-related single dead tree detection using multi-temporal ALS data and CIR imagery. *Remote Sens. Environ.* 2018, *219*, 31–43, doi:10.1016/j.rse.2018.10.005.

Kamińska A., Lisiewicz M., Stereńczak K. 2021.*Single Tree Classification Using Multi-Temporal ALS Data and CIR Imagery in Mixed Old-Growth Forest in Poland*, Remote Sensing, *13*(24), 5101. https://doi.org/10.3390/rs13245101

Laslier, M.; Ba, A.; Hubert-Moy, L.; Dufour, S. Comparison of leaf-on and leaf-off ALS data for mapping riparian tree species. In Proceedings of the SPIE 2017, Warsaw, Poland, 2017.

Ørka, H.O.; Næsset, E.; Bollandsås, O.M. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* 2009, *113*, 1163–1174.

Ørka, H.O.; Gobakken, T.; Næsset, E.; Ene, L.; Lien, V. Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Can. J. Remote Sens.* 2012, *38*, 125–138, doi:10.5589/m12-021.

Lin, Y.; Hyyppä, J. A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *Int. J. Appl. Earth Obs. Geoinf.* 2016, *46*, 45–55, doi:10.1016/j.jag.2015.11.010.

You, H.T.; Lei, P.; Li, M.S.; Ruan, F.Q. Forest pecies classification based on three-dimensional coordinate and intensity information of airborne LiDAR data with Random Forest method. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Guilin, China, 2020; Volume 42, pp. 117–123.

Acknowledgements

The analyses performed in this manuscript were funded as part of a project entitled "Comprehensive spatial analysis of the dieback of dominant tree species using multi-temporal ALS data and CIR imagery in the Białowieża Primeval Forest", carried out in the Forest Research Institute during 2021–2023. Data was funded by the project "LIFE+ ForBioSensing PL Comprehensive monitoring of stand dynamics in Białowieża Forest supported with remote sensing techniques" which is co-funded by the EU Life Plus programme (contract number LIFE13 ENV/PL/000048) and The National Fund for Environmental Protection and Water Management in Poland (contract number 485/2014/WN10/OP-NM-LF/D).

Spatial analysis of methane emissions from agriculture at the municipality level (LAU-2) in Poland in 2010 and 2020

KATARZYNA KOZICKA*, MARCIN OLLIK, ELŻBIETA WÓJCIK-GRONT

Department of Biometry, Institute of Agriculture
Warsaw University of Life Sciences, Poland

*katarzyna_kozicka@sggw.edu.pl

Methane (CH_4) emissions are influenced mainly by anthropogenic sources, particularly by agriculture - about 40% [Ritchie et al., 2020]. The most important source is livestock farming, where CH_4 emissions are produced during enteric fermentation in rumen and from manure management [Albrektsen et al., 2021; Yusuf et al., 2012]. CH_4 emissions are calculated according to IPCC 2006 methodology based on the population size of the animal category and the CH_4 emission factor (EF). Analysis of changes in livestock populations and greenhouse gases (GHG) emissions at the national level is important for climate change mitigation efforts. Therefore, the aim of the study was to estimate CH_4 emissions from enteric fermentation and manure management for Poland in 2010 and 2020 at the municipal level (LAU-2). The study used livestock population data (dairy and non-dairy cattle, swine, and poultry) from the National Agricultural Census and CH_4 EF for each animal category from the UNFCCC National Inventory Report (NIR) for 2010 and 2020. The results obtained for 2010 and 2020 were compared to analyse changes in the CH_4 emissions from livestock for municipalities in Poland. The research summarizes CH_4 emissions from enteric fermentation and manure management in total and at the municipality level for Poland in 2010 and 2020.

Total CH_4 emissions from livestock farming in Poland increased slightly over the decade. The increase in CH_4 emissions between 2010 and 2020 is mainly due to the increase in the non-dairy cattle population, partially offset by a decrease in swine population. Despite the highest share of total CH_4 emissions from dairy cattle (57.2% in 2020), the increase in emissions by 2020 was not significant (1.5%), due to the decline in dairy cattle population. The non-dairy cattle total CH_4 emissions increased by 18.9% in 2020, whereas emissions from swine decreased by 51.5% in 2020, which is consistent with the population changes. The share of poultry was negligible (below 1.1%), however, due to the increasing population, the emissions in 2020 were higher by 22.3% than in 2010. The highest emissions were observed in particular

voivodships where livestock farming was predominant – Mazowieckie, Podlaskie, and Wielkopolskie Voivodships. In contrast, the lowest emissions were found mainly in southeastern and southwestern Poland, where the decrease in livestock population was the highest among all voivodships. In addition, in particular regions, noticeable changes have been observed. The increased animal population (dairy and non-dairy cattle) caused high increases in CH_4 emissions, especially visible in Podlaskie, Lubuskie, and Wielkopolskie Voivodship.

Changes in livestock population and EF in Poland in past decade are influenced mainly by EU standards and Common Agricultural Policy (milk limits and fluctuations in milk prices, the unprofitability of piglets and fattening swine production, and the ASF disease [National Agricultural Census 2020]. Therefore, it is necessary to provide studies on livestock farming and its changes in Poland based on available reports and databases. In addition, spatial distribution of CH_4 emissions from agriculture enables monitoring GHG emissions by voivodships and municipality.

Keywords: methane emissions, livestock, IPCC, spatial distribution

References

Albrektsen, R., Mikkelsen, M.H., Gyldenk, S., 2021. *Danish emission inventories for agriculture*. Inventories 1985 – 2018.

Central Statistical Office. National Agricultural Census 2020. Report on the results.

Ritchie, H., Roser, M., Rosado, P., 2020. *CO_2 and Greenhouse Gas Emissions*. Our World in Data.

Yusuf, R.O., Noor, Z.Z., Abba, A.H., Hassan, M.A.A., Din, M.F.M., 2012. *Methane emission by sectors: A comprehensive review of emission sources and mitigation methods*. Renewable and Sustainable Energy Reviews 16, 5059–5070.

Porównanie jakości rozpoznawania ziaren pyłku przez detektory jedno i dwustopniowe

ELŻBIETA KUBERA[1*], AGNIESZKA KUBIK-KOMAR[1], KRYSTYNA PIOTROWSKA-WERYSZKO[2], AGATA KONARSKA[2]

[1]Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

[2]Department of Botany and Plant Physiology
University of Life Sciences in Lublin, Poland

[*]elzbieta.kubera@up.lublin.pl

Monitoring pyłków jest w zdecydowanej większości stacji prowadzony za pomocą pułapek wolumetrycznych. Tworzenie preparatów i ich analiza pod mikroskopem jest zadaniem złożonym i czasochłonnym. Celem naszych badań jest stworzenie systemu automatycznego rozpoznawania i zliczania ziaren pyłku poszczególnych taksonów na podstawie zdjęć mikroskopowych.

W niniejszej pracy porównujemy trzy typy detektorów obiektów pod względem poprawności rozpoznawania ziaren pyłku Alnus, Betula, Corylus i Carpinus – dwa detektory jednostopniowe: YOLOv5 (w dwóch wersjach różniących się miarą dopasowania modelu) oraz RetinaNet i dwustopniowy detektor – Faster RCNN. Nasz zbiór danych składa się ze zdjęć mikroskopowych materiału referencyjnego. Każdy model detektora zbudowano trzykrotnie, dzięki czemu otrzymano 12 modeli. Szkolenie każdego detektora trwało 500 epok i polegało na dostrojeniu wstępnie wytrenowanych modeli na naszym zbiorze danych.

Wyniki detekcji wyrażone miarami jakości klasyfikacji: precision, recall i F-score porównano przy użyciu nieparametrycznej alternatywy opartej na rangach dla ANOVA z powtarzanymi pomiarami. Zastosowaliśmy układ F1-LD-F2 z taksonem jako czynnikiem stałym, a także dwoma czynnikami powtarzanego pomiaru – rodzajem detektora i jego powtórzeniem. Dodatkowo, w celu porównania jakości zbudowanych detektorów, wykonano porównania wielokrotne z korektą Holma-Bonferroniego.

Uzyskane wyniki pozwoliły wskazać detektor YOLO jako lepszy od RetinaNet pod względem wszystkich miar klasyfikacyjnych oraz Faster RCNN pod względem recall i F-score.

Nie stwierdzono różnic w rozkładach tych miar pomiędzy dwiema wersjami YOLO. Ponadto żaden z badanych detektorów nie miał preferencji w zakresie precyzji klasyfikacji konkretnych taksonów. Okazuje się jednak, że rozkład recall dla badanych taksonów różni się znacznie w zależności od rodzaju detektora. RetinaNet i Faster RCNN częściej pomijały częściowo widoczne ziarna pyłku znajdujące się w pobliżu krawędzi obrazu niż detektory YOLO, co prawdopodobnie skutkowało różnicą w wynikach recall.

Keywords: object detection, pollen monitoring

References

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P. 2014. COCO - Common Objects in Context. https://arxiv.org/abs/1405.0312

Data visualisation in R - from simple graphs to animations

AGNIESZKA KUBIK-KOMAR

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

agnieszka.kubik@up.lublin.pl

dplyr is a popular R package that provides many functions for data manipulation in data frames. It is often used for efficient data processing and analysis in R. The basic operations that can be performed using the functions from this package are filtering and sorting data, calculating statistics and measures for individual groups or the entire data set. In addition, dplyr provides functionality to handle missing cases and create new variables from existing data. These operations enable data preparation in a manner that streamlines the analysis, making it more centred on crucial aspects of the study, helping to eliminate errors, and making the data structure easier to understand. With dplyr, one can fine-tune the data to the specific analysis needs, which is a critical step in data mining and analysis.

Keywords: R environment, dplyr, data manipulation

References

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

Cotton R., Data Manipulation with dplyr in R Cheat Sheet | DataCamp. https://www.datacamp.com/cheatSheet/data-manipulation-with-dplyr-in-r-cheat-sheet

Functional repeated measures analysis of variance and its application

KATARZYNA KURYŁO*, ŁUKASZ SMAGA

Faculty of Mathematics and Computer Science
Adam Mickiewicz University, Poznań, Poland

*katkur7@st.amu.edu.pl

This paper is motivated by medical studies in which the same patients with multiple sclerosis are examined at several successive visits and described by fractional anisotropy tract profiles, which can be represented as functions. Since the observations for each patient are dependent random processes, they follow a repeated measures design for functional data. To compare the results for different visits, we thus consider functional repeated measures analysis of variance. For this purpose, a pointwise test statistic is constructed by adapting the classical test statistic for one-way repeated measures analysis of variance to the functional data framework. By aggregating the pointwise test statistic, we create two global test statistics. Apart from verifying the general null hypothesis on the equality of mean functions corresponding to different objects, we also propose a simple method for post hoc analysis. We illustrate the finite sample properties of permutation and bootstrap testing procedures in the simulation study. Finally, we analyze a motivating real data example in detail.

Keywords: analysis of variance, bootstrap, functional data analysis, permutation method, post hoc analysis, repeated measures

References

Horváth L., Kokoszka P. 2012. *Inference for functional fata with applications*. Springer.

Martinez-Camblor P., Corral N. 2011. *Repeated measures analysis for functional data*. Computational Statistics & Data Analysis 55, 3244-3256.

Kuryło K., Smaga Ł. 2023a. *Functional repeated measures analysis of variance and its application*. Preprint https://arxiv.org/abs/2306.03883.

Kuryło K., Smaga Ł. 2023b. rmfanova: *Repeated Measures Functional Analysis of Variance*. R package version 0.1.0, https://CRAN.R-project.org/package=rmfanova.

Ramsay J.O., Silverman B.W. 2005. *Functional data analysis*. Second Edition. Springer.

Smaga Ł. 2020. A note on repeated measures analysis for functional data. AStA Advances-139.

Smaga Ł. 2020. *A note on repeated measures analysis for functional data*. AStA Advances in Statistical Analysis 104, 117-139.

Wykorzystanie analizy składowych głównych do badania wpływu nawożenia i sposobu uprawy na skład chemiczny nasion konopi

GRAŻYNA NIEDZIELA

Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

grazyna.niedziela@up.poznan.pl

Uprawa konopi ma duży potencjał gospodarczy, jest bardzo popularna w przemyśle i zainteresowanie nią cały czas rośnie. W ramach badań poddano analizie nasiona odmiany Henola, jednej z najnowszych odmian konopi. Doświadczenie dotyczyło rośliny uprawianej w celu pozyskania olejków eterycznych, a następnie na nasiona (Henola 2) oraz uprawianej wyłącznie na nasiona (Henola 3).

Celem eksperymentu było zoptymalizowanie metod uprawy konopi oraz porównanie zawartości związków bioaktywnych w nasionach roślin Henola 2 i Henola 3. W uprawie ważną rolę odgrywa nawożenie, ponieważ może wpływać na właściwości nasion i oleju.

Wykonując porównania wykorzystano analizę składowych głównych, popularną technikę umożliwiającą redukcję wymiarowości danych i pozwalającą wykryć prawidłowości między zmiennymi. Podczas prezentacji zostanie przedstawione wykorzystanie tej analizy do określenia wpływu nawożenia na skład chemiczny nasion oraz do porównania właściwości nasion Henoli 2 i Henoli 3.

Na podstawie wykonanych analiz można stwierdzić, że rodzaj nawożenia nie wpływa znacząco na skład chemiczny nasion, a uprawa konopi wyłącznie na nasiona i uprawa na dwa cele: na olejki eteryczne, a następnie nasiona nie zmienia zawartości rozważanych związków.

Keywords: Analiza składowych głównych, konopie, nasiona

References

Frankowski, J., Przybylska-Balcerek, A., Graczyk, M., Niedziela, G., Sieracka, D., i Stuper-Szablewska, K. 2023. *The effect of mineral fertilization on the content of bioactive compounds in hemp seeds and oil.* Molecules, 28(12).

Górecki, T. 2011. *Podstawy statystyki z przykładami w R.* Wydawnictwo BTC.

Krzyśko, M. 2009. *Podstawy wielowymiarowego wnioskowania statystycznego.* Wydawnictwo Naukowe UAM.

Classification methods in the diagnosis of breast cancer

ANNA MAGDALENA OGŁOSZKA*, ŁUKASZ SMAGA

Faculty of Mathematics and Computer Science
Adam Mickiewicz University, Poznań, Poland

*a.m.ogloszka@gmail.com

We consider the problem of diagnosing the type of cancer with the use of machine learning and statistical methods. Nowadays, the problem of neoplasms, in particular breast cancer, is one of the biggest problems of humanity. Identification of cancer and its type is extremely important. In solving this problem, the classification methods can be used as objective tools, which can be helpful for doctors making a diagnosis. For this reason, we discuss many efficient classificators in the context of cancer detection. In addition, we also consider the topic of data set transformations to deal with the problem of data unbalance as well as measures of classification quality. In the experimental part, an attempt will be made to find the best classifier and to improve the quality of the original data set to obtain the highest values of classification quality measures for the particular data set.

Keywords: cancer, breast cancer, classification, classifier, unbalanced data, quality measurements of classification

References

Breast Cancer Wisconsin Data Set. Available at https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+\Wisconsin+%28Diagnostic%29

Górecki T. 2011. *Basics of statistics with examples in R*. BTC (in Polish).

Krzyśko, M., Wołyński, W., Górecki, T., Skorzybut, M. 2008. *Machine learning - pattern recognition, cluster analysis and dimensional reduction*. WNT, Warsaw (in Polish).

Ogłoszka, A.M. 2022. *Classification methods in the diagnosis of breast cancer*. Master's thesis defended in Data Science. Adam Mickiewicz University, Poznań (in Polish).

Wojciechowska U., Didkowska J. 2013 *Illnesses and deaths from malignant neoplasms in Poland*. National Cancer Registry from National Oncology Institute Maria Skłodowskiej-Curie - National Research Institute. Available at http://onkologia.org.pl/raporty/

Wolberg, W.H., Street, W.N., Mangasarian, O.L. 1994. *Machine learning techniques to diagnose breast cancer from fine-needle aspirates*. Cancer Letters 77, 163-171.

Acknowledgements

This research is based on the partial results of the Master's thesis defended in Data Science in Adam Mickiewicz University, Poznań: Ogłoszka, A.M. (2022). Classification methods in the diagnosis of breast cancer (in Polish).

Modeling Water Level Fluctuation in River Basins Using Singular Spectrum Analysis

AMILCAR OLIVEIRA[1,2*], CARLA SARMENTO[1]

[1]Department of Scences and Tecnhology, Universidade Aberta, Lisboa, Portugal

[2]CEAUL – Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal

[*]amilcar.oliveira@uab.pt

Water scarcity affected 29% of the EU territory during at least one season in 2019. In the face of climate change, it is very important to understand the risk of water scarcity. Water scarcity is becoming a growing problem in southern European countries, such as Portugal. In 2019, Portugal, faced one of the most significant water scarcity conditions in the EU-27 on the seasonal scale (seasonal WEI 66%).The main objective of this work is to study the water level fluctuation in river basins, in order to predict the risks of lack of water. The study area is located in 29 reservoirs from different river basins in Portugal. The collected data refer to the period from November 1993 to August 2022, with a total number of records of 9686. We started by improving the quality of the data and built a monthly time series of the volume of water stored. Next, we analyzed the time series using Singular Spectrum Analysis (SSA), which is a non-parametric technique for analyzing time series.

Keywords: risk, water scarcity, time series, singular spectrum analysis.

References

Golyandina N., Nekrutkin V., Zhigljavsky A. 2001. *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC.

Golyandina N., Zhigljavsky A. 2013. *Singular Spectrum Analysis for Time Series.* Springer.

Mahmoudvand R., Najari N., Zokaei M. 2013. *On the Parameters for Reconstruction and Forecasting in the Singular Spectrum Analysis*. Communication in Statistics: Simulations and Computations, Vol. 42, 860-870.

Rodrigues C. P., Mahmoudvand R. 2016. *Correlation Analysis in Contaminated Data by Singular Spectrum Analysis*. Quality and Reliability Engineering International, Forthcoming: doi: 10.1002/qre.2027.

Acknowledgements

This work is partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia under the project UIDB/00006/2020.

C.R. Rao: A Beacon of Excellence in Statistical research and practice

TERESA A. OLIVEIRA

DCeT-Department of Sciences and Technology
Universidade Aberta, Lisbon, Portugal

and

CEAUL-Center of Statistics and Applications of the University of Lisbon

*teresa.oliveira@uab.pt

This talk shines a spotlight on the remarkable achievements of C.R. Rao, an eminent statistician and mathematician, whose contributions have illuminated the field of statistics and related areas. Rao's transformative ideas in estimation theory, sufficiency and completeness, experimental design, biometry and data science, have revolutionized the way researchers approach data analysis and interpretation. Furthermore, his mentorship and educational efforts have fostered the growth of countless statisticians, ensuring his legacy will continue for generations to come. The article underscores Rao's exceptional impact and highlights his numerous prestigious awards, solidifying his place as a beacon of excellence in the field. Some of the Rao's pivotal role in shaping the future of statistical research and practice will be presented.

Keywords: C. R. Rao, statistics, experimental design, biometry.

References

Chaudhuri A., Gupta, S.N., Roychoudhury, R. -Editors. 2021. A Tribute to the Legend of Professor C. R. Rao: The Centenary Volume (Indian Statistical Institute Series) Kindle Edition. Springer.

Arnold, B.C., Balakrishnan N., Coelho C.A. -Editors. 2022. Methodology and Applications of Statistics: A Volume in Honor of C.R. Rao on the Occasion of his 100th Birthday, Springer International Publishing

Rao, C.R., Govindaraju, V. 2013. Handbook of Statistics. Safari, an O'Reilly Media Company. North Holland Ed.

Acknowledgements

This research was partially funded by FCT—Fundação para a Ciência e a Tecnologia, under the project—UIBD/00006/2020.

Evaluation of water quality in the Danube River (Serbia) based on an improvised water quality index and multivariate analyses

MONIKA RÓŻAŃSKA-BOCZULA[1*], ANTONI GRZYWNA[2], JASNA GRABIĆ[3]

[1]Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

[2] Department of Environmental Engineering and Geodesy
University of Life Sciences in Lublin, Poland

[3] Department of Water Management
University of Novi Sad, Serbia

[*]monika.boczula@up.lublin.pl

This paper presents the results of a study conducted in Serbia in 2019 regarding water quality in the Danube River. Multivariate statistical techniques and an improvised water quality index (WQI) were used to describe temporal and spatial changes. The proposed modification of the WQI makes it possible to reduce the number of physicochemical parameters needed to determine the state of the water by 72%, significantly reducing the time and cost of the study. Based on the analyses carried out, water quality was found to be good upstream and excellent downstream. The highest levels of pollution were observed in winter.

Keywords: the Danube river, Serbia, spatial and temporal variability, water quality index, multivariate statistical methods

References

Bărbulescu A., Barbeş L. 2020. *Assessing the water quality of the Danube River (at Chiciu, Romania) by statistical methods*. Environ. Earth Sci., 79: 1–14.

Grzywna A., Bronowicka-Mielniczuk U. 2020. *Spatial and temporal variability of water quality in the Bystrzyca river basin, Poland*. Water, 12, 190, doi: 10.3390/w12010190.

Iticescu C., Georgescu L.P., Murariu G., Topa C.M., Timofti M., Pintilie V., Arseni M. 2019. *Lower Danube water quality quantified through WQI and multivariate analysis*. Water, 11, 1305, doi: 10.3390/w11061305.

Varol M. 2020. *Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study*. Environ. Pollut., 266, 115417, doi: 10.1016/j.envpol.2020.115417.

WHO. 2017. Guidelines for Drinking Water Quality: Fourth Edition Incorporating the First Addendum, World Health Organization.

Estimands in clinical trials. How to address intercurrent events using the composite strategy for longitudinal continuous outcome.

KINGA SAŁAPA

AstraZeneca Pharma Poland

kinga.salapa@astrazeneca.com

In 2019 the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) introduced the addendum to efficacy guidelines E9 (R1). It focuses on statistical principles related to estimands and sensitivity analysis. It is highlighted that a clear trial objectives should be translated into key clinical questions of interest by defining suitable estimands (i.e. "what is to be estimated").

The important part of the defining estimand is clarification of how a potential intercurrent event (ICE) is addressed in the clinical question of interest. The intercurrent events are events occurring after treatment initiation and affect either the interpretation or the existence of the measurement associated with the clinical questions, e.g. use of alternative treatment or rescue medication, discontinuation of treatment or terminal events such as death. It should be noted that the addendum makes it clear that missing values due to e.g. study withdrawal or loss to follow up are not intercurrent events and are not to be reflected in the estimand, but instead represent limitations to the data that need to be addressed in the statistical analysis.

Five different strategies are introduced to address potential intercurrent events: treatment policy, hypothetical strategy, composite strategy, while on treatment and principal stratum. The focus is on composite strategy where an intercurrent event is considered in itself to be informative about the patient's outcome. In other words the composite strategy treat the occurrence of the ICE as representing a failure of treatment.

There are several ways of how to handle data after ICE under composite strategy. The responder analysis can be applied, where response corresponds to having a favourable clinical outcome without ICE. A patient who experiences ICE is assigned as non-responder. A drawback of this approach is that the statistical power is often reduced when binary data are used comparing the analysis performed on the original continuous outcome. Other option could be a transformation of the continuous data to the ordinal scale but the issue with loss of power

still applies. Moreover not in all cases the cut-offs used in transformation are widely known and accepted. The most reasonable seems to use a continuous outcome in the analysis and assign an unfavorable value to patients who experienced ICE. A range of possible approaches might be considered for how to choose the unfavourable outcome, eg. one approach might be to use pre-specify single value y* for all visits after ICE but determine a suitable value of y* might be difficult, the other one could be to estimate it based on the observed data. Darken et al. presented the alternative approaches such as non-stochastic estimated single imputation or multiple imputation, where an unfavourable value is taken from the function of the outcome data distribution. The 5th percentile from a treatment group, assuming the low values of the distribution are unfavourable, is used to define the outcome at visit following an ICE.

Keywords: clinical trials, estimand, composite strategy

References

ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials – Training materials. 2019. Retrieved 17th May 2022 from https://www.ich.org/page/efficacy-guidelines

Darken P, Nyberg J, Ballal S, Wright D. 2020. *The attributable estimand: A new approach to account for intercurrent events.* Pharmaceutical Statistics. 19:626–635. https://doi.org/10.1002/pst.2019

Some properties of a two-way elimination of heterogeneity design

IDZI SIATKOWSKI

Department of Mathematical and Statistical Methods

Poznan University of Life Sciences

idzi.siatkowski@up.poznan.pl

We consider the decomposability property, orthogonality and the connectedness of the two-way elimination of heterogeneity design. Further, we characterize the efficiency-balance and the variance-balance of a class of two-way elimination of heterogeneity designs in relation to the corresponding balance properties of subdesigns.

Keywords: Latin square design, row-column design, two-way elimination of heterogeneity designs

References

Baksalary J.K., Siatkowski I. 1993. Bounds for the efficiency factor for a class of two-way elimination of heterogeneity designs. *Gujarat Statistical Review*, 1-9.

Baksalary J.K., Siatkowski I. 1993. Decomposability of the C-matrix of a two-way elimination of heterogeneity design. *Journal of Statistical Planning and Inference* 36, 301-310.

Siatkowski I. 1993. On connectedness of a two-way elimination of heterogeneity design. *Biometrical Journal* 35, 283-289.

Siatkowski I. 1993. Orthogonality in a two-way elimination of heterogeneity design. *Utilitas Mathematica* 43, 225-233.

Siatkowski I. 1994. Some bounds for balanced two-way elimination of heterogeneity designs. *Statistics & Probability Letters* 21, 95-99.

Early detection of phosphorus deficiency stress in cucumber (cucumis sativus L.) plants based on selected physiological parameters

LESZEK SIECZKO[1*], KATARZYNA KOWALCZYK[2], GAJC-WOLSKA JANINA[2] , KOWALCZYK WALDEMAR[3], PIOTR DĄBROWSKI[4], WOJCIECH BORUCKI[5], MONIKA JANASZEK-MAŃKOWSKA[6], NIEDZIŃSKA MONIKA[2], JAROSŁAW LEON PRZYBYŁ[2], ZUZANNA MALWINA JASZCZUK[7], HAZEM KALAJI[8]

[1] Department of Biometry, Institute of Agriculture,
Warsaw University of Life Sciences-SGGW, Poland

[2] Department of Vegetable and Medicinal Plants, Institute of Horticultural Sciences,
Warsaw University of Life Sciences-SGGW, Poland

[3] The National Institute of Horticultural Research,
Skierniewice, Poland

[4] Department of Environmental Management, Institute of Environmental Engineering,
Warsaw University of Life Sciences-SGGW, Poland

[5] Department of Plant Physiology, Institute of Biology,
Warsaw University of Life Sciences-SGGW, Poland

[6] Department of Fundamentals of Engineering and Power Engineering, Institute of Mechanical Engineering, Warsaw University of Life Sciences-SGGW, Poland

[7] Faculty of Agriculture and Ecology,
Warsaw University of Life Sciences WULS-SGGW, Poland

[8] Department of Botany, Institute of Biology,
Warsaw University of Life Sciences-SGGW, Poland

*Leszk_Sieczko@sggw.edu.pl

Identification of phytomonitoring and physiological features that provide information about the health condition of plants can help mitigate the effects of stress and improve plant productivity. In this study, we employed various tools and techniques to indirectly assess the response of cucumber plants to phosphorus deficiency and understand the mechanisms standing behind this response. Analyses of chlorophyll and carotenoids content in leaves revealed a rapid response of these indicators in older leaves. Changes in photosynthetic performance at the molecular level were found to be dependent on leaf age and the timing of measurements. By employing multidimensional statistical methods, we were able to demonstrate the

interdependence of the examined features in relation to the experimental system, including the timing of measurements. This information contributes to understanding the variability of characteristics associated with phosphorus deficiency in cucumber plants during prolonged deficiency periods and allow early identification of this stress.

Keywords: chlorophyll fluorescence; confocal microscopy; greenhouse cucumber; Leaf Area Index; multivariate statistical analyses, photosynthetic pigments.

References

Bayoumi Y., Abd-Alkarim E., El-Ramady H. et al.: Grafting Improves Fruit Yield of Cucumber Plants Grown under Combined Heat and Soil Salinity Stresses. Horticulturae, 7, 61.: 2021. https://doi.org/10.3390/ horticulturae7030061

Borch K.; Bouma T.J.; Lynch J.P.; Brown K.M.: Ethylene: A Regulator of Root Architectural Responses to Soil Phosphorus Availability. Plant Cell Environ. 22: 425–43, 1999.

Cetner M.D., Kalaji H.M., Borucki W., Kowalczyk K.: Phosphorus deficiency affects the I-step of chlorophyll a fluorescence induction curve of radish. Photosynthetica 58: 671-681, 2020.

Ciereszko, I.; Gniazdowska, A.; Mikulska, M.; Rychter, A.M.: Assimilate Translocation in Bean Plants (Phaseolus vulgaris L.) during Phosphate Deficiency. J. Plant Physiol., 149: 343–348, 1996.

Raghothama K.G.: Phosphate Acquisition 1999. Annual Review of Plant Physiology and Plant Molecular Biology Vol. 50: 665-693 (Volume publication date June 1999), 1999. https://doi.org/10.1146/annurev.arplant.50.1.665

Saleque M.A., Abedin M.J., Ahmed Z.U., Hasan M., Panaullah G. M.: Influences of phosphorus deficiency on the uptake of nitrogen, potassium, calcium, magnesium, sulfur, and zinc in lowland rice varieties, Journal of Plant Nutrition, 24(10): 1621-1632, 2001. DOI:10.1081/PLN-1001060252001

Sieczko L., Dąbrowski P., Kowalczyk K., Gajc-Wolska J., Borucki W., Janaszek-Mańkowska M., Kowalczyk W., Farci D., Kalaji, M.H.: Early detection of phosphorus deficiency stress in cucumber at the cellular level using chlorophyll fluorescence signals. Journal of Water and Land Development: 176–187, 2022. https://doi.org/10.24425/jwld.2022.143734

Zambrosi F.C.B., Mesquita G.L., Menino G., Tanaka F.A.O., Mattos D.Jr., Quaggio J.A.: Anatomical and ultrastructural damage to citrus leaves from phosphite spray depends on phosphorus supply to roots. Plant Soil 416: 1–13, 2017.

Zambrosi F.C.B., Mesquita G.L., Tanaka F.A.O.: Assessment of leaf ultrastructure offers insights into mechanisms regulating sugarcane performance under low-phosphorus stress, Acta Physiologiae Plantarum (2020) 42: 54, 2020. https://doi.org/10.1007/s11738-020-03052-w

Behavior and reproductive success in a farm population of red foxes (Vulpes vulpes) from a biometrical perspective

EWA SKOTARCZAK[1*], ANDRZEJ FILISTOWICZ[2]

PIOTR PRZYSIECKI[3], TOMASZ SZWACZKOWSKI[4]

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

[2]Institute of Animal Science
Wrocław University of Life and Environmental Sciences

[3]Higher School of Agriculture and Building,
Leszno, Poland

[4]Department of Genetics and Animal Breeding
Poznań University of Life Sciences, Poland

[*]ewa.skotarczak@up.poznan.pl

Modern livestock breeding involves a wide range of improved traits, from reproduction to welfare. Reproductive success as well as housing conditions and feeding technology translate directly into economic results. In recent years, more and more attention has been paid to animal welfare, which is closely related to the behaviour of animals. By contrast to production measurable traits, the assessment of well-being is largely indirect (Calkins et al., 2013)

The objective of the present study was to examine the relationship between behavioural tests and reproductive characteristics. Data was extracted from a pedigree farm "Batorówka" (1960-2020 years). Detailed description of the data is given by Grzybek et al. (2021) and Przysiecki et al. (2023). Among the numerous recorded traits, the current analysis focused on the following for males: number of matings, length of the reproduction season, number of reproductive seasons, day of first and last mating, as well as for dams: number of mating days, number of mates, pregnancy length, litter size at birth, and litter size at weaning. Both sexes underwent two behavioural tests. The independence between observed behaviour and reproductive characteristics was assessed using the chi-square test for contingency tables.

In the majority of the examined contingency tables, the null hypothesis of independence between the assessed traits was rejected. However, further analyses are necessary to determine the magnitude and direction of these relationships.

Keywords: red fox, behavior, reproduction

References

Calkins, E.S., Fuller, T.K., Asa, C.S., Sievert, P.R., & Coonan, T.J. (2013). *Factors influencing reproductive success and litter size in captive island foxes*. The Journal of Wildlife Management, 77(2), 346-351.

Grzybek, P., Przysiecki, P., Filistowicz, A., Dobrzanski, J., & Szwaczkowski, T. (2021). *Pedigree analysis of a red fox (Vulpes vulpes) population.* Annals of Animal Science, 21, 457-468.

Przysiecki P, Filistowicz A, Skotarczak E, Dobrzynska P, Szwaczkowski T. (2023). *Bayesian analysis of genetic and environmental effects on litter traits in a red fox (Vulpes vulpes) herd under long-term selection*. Animal Science Journal, 94(1):e13807.

Theoretical vs empirical distributions of chicken performance traits

EWA SKOTARCZAK[1*], TOMASZ SZWACZKOWSKI[2], WOJCIECH MUELLER[3]
SEBASTIAN KUJAWA[3], PRZEMYSŁAW NOWAK[3], PRZEMYSŁAW IDZIASZEK[3]
KRZYSZTOF KOSZELA[3], ANNA SWAT[4], KATARZYNA POŁTOWICZ[5]

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

[2]Department of Genetics and Animal Breeding
Poznań University of Life Sciences, Poland

[3]Department of Biosystems Engineering
Poznań University of Life Sciences, Poland

[4]Pedigree Farm of Laying Hen "Rszew" in Konstantynów Łódzki, Poland

[5]Department of Poultry Breeding
National Research Institute of Animal Production, Balice, Poland

[*]ewa.skotarczak@up.poznan.pl

The BLUP method under an animal model is commonly used to predict breeding values for farm animals (Henderson, 1975). This method allows for comparison of individuals across generations. By applying mixed model methodology, specifically the multitrait animal model, it is possible to estimate genetic parameters such as heritability and genetic correlation. However, the accuracy of the predicted parameters depends on certain assumptions which may be more or less fulfilled by the datasets collected in a breeding farm (Gianola, 2000).

The objective of this research was to assess the empirical distributions of productive traits in laying hens and compare them to theoretical assumptions. The data was collected from the Pedigree Farm of Laying Hens "Rszew" in Konstantynów Łódzki, Poland. Six traits were recorded: body weight, average egg weight, sexual maturity, initial egg production rate of initial egg production, and average of egg shell thickness. An analysis includes 824 and 1444 recorded birds in two consecutive years 2022 and 2023, respectively. The BLUP approach was employed to predict combined breeding values, while (co)variance components were estimated using the REML method.

In populations under genetic selection, it is widely recognized that selection can introduce a bias in the normality of data, leading to an overestimation of environmental error (Cervantes at.al., 2016). From a practical perspective, the transformation of data can complicate the interpretation of results from statistical models. As a result, breeders prefer to retain and analyze the raw data. Furthermore, breeders consider it crucial to use well-established and readily available computational methods. However, the reliability of predictors obtained from these models is contingent upon the model aligning with the theoretical assumptions. In addition to data transformation, another approach such as the elimination of outliers can be used to improve the empirical distribution of the data.

Keywords: BLUP assumptions, laying hens, empirical distribution

References

Cervantes, I., Gutiérrez, J.P. & Meuwissen, T.H. Response to selection while maximizing genetic variance in small populations. Genetics Selection Evolution 48, 69 (2016).

Gianola, D. 2000. *Statistics in Animal Breeding*. Journal of the American Statistical Association, 95(449), 296–299.

Henderson, C. R. 1975. *Best Linear Unbiased Estimation and Prediction under a Selection Model*. Biometrics, 31(2), 423–447.

Acknowledgements

The study is financed by the GENDROB research project of „DZIAŁANIE 16 WSPÓŁPRACA” – contract nr 00038.DDD.6509.00101.2019.06.

On the detectability of symmetric distributions that deviate from normality due to small excess kurtosis

PIOTR SULEWSKI

Institut of Exact and Technical Sciences
Pomeranian University in Słupsk, Poland

piotr.sulewski@apsl.edu.pl

The aim of this article is to seamlessly test GoFT's ability to detect deviations from normality. The very specific case is considered namely deviation from normality consists in coincidence of symmetry and small excess kurtosis (EX). The first step in fulfilling the above aim is to collect the set of normality-oriented GoFTs being recommended for use, mainly in recently published source literature. The second step is to create a family of symmetric distributions with non-constant EX, further called alternatives. Formulas for calculating EX values are provided for each distribution. To compare alternatives with the Normal distribution the appropriate similarity measure is applied. The third step is the Monte Carlo simulation. The study is carried out involving twenty GoFTs and thirty one alternatives. The obtained results show that the GoFTs taken into consideration detect deviation from normality distributions of positive EX much more frequently than distributions of negative EX. Appropriate GoFTs are recommended.

Keywords: normal distribution, goodness-of-fit testing, excess kurtosis modeling

References

Desgagne, A., Lafaye de Micheaux, P. 2018. A powerful and interpretable alternative to the Jarque-Bera test of normality based on 2nd-power Skewness and Kurtosis, using the Rao's Score Test on the APD family. *Journal of Applied Statistics* 45(13), 2307–27.

Desgagné, A., Lafaye de Micheaux, P., Ouimet, F. 2022. Goodness-of-fit tests for Laplace, Gaussian and exponential power distributions based on λ-th power skewness and kurtosis. *Statistics*, 1-29.

Sulewski, P. 2019. Modification of Anderson-Darling goodness-of-fit test for normality. *Afinidad*, 76(588).

Sulewski, P. 2020. Modified Lilliefors goodness-of-fit test for normality, *Communications in Statistics - Simulation and Computation*, 51(3), 1199-1219.

Sulewski, P. 2021. Two component modified Lilliefors test for normality. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 16(2), 429-455.

Sulewski P. 2021. DS Normal Distribution: Properties and Applications, *Lobachevskii Journal of Mathematics* 42(12), 2980–2999.

Sulewski, P. 2022. Normal distribution with plasticizing component. *Communications in Statistics-Theory and Methods*, 51(11), 3806-3835.

Sulewski, P. 2022. Easily Changeable Kurtosis Distribution. *Austrian Journal of Statistics*, 52, 1-24.

Distribution of mineral fertilizer granules' crushing strength

MAŁGORZATA SZCZEPANIK[1*], NORBERT LESZCZYŃSKI[2]

[1]Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

[2]Department of Agricultural, Forestry and Transport Machines
University of Life Sciences in Lublin, Poland

[*]malgorzata.szczepanik@up.lublin.pl

Selected granular fertilizers with different content of multicomponent micronutrients of the smart fertilizer type were investigated. For the granules of each fertilizer, measurements of the crushing strength were made, and then the best-matched probability distributions of the crushing strength were indicated among Johnson, Gaussian mixture, generalized extreme value, normal, log-normal, triangular, Weibull, Rayleigh and generalized Pareto distributions.

Keywords: mineral fertilizers, crushing force

References

Biskupski A., Schab S., Myka A., Dawidowicz M. 2012. *Właściwości fizykochemiczne nawozów dostępnych na rynku i możliwości wytwarzania nawozów mieszanych.* Chemik 66 (5), 541–548.

Leszczyński N., Przystupa W., Nowak J., Tatarczak J., Kowalczuk J., Rusek P. 2016. *Studies on compression strength of superphosphate and urea granules*. Przemysł Chemiczny 95 (1), 121–124.

Leszczyński N., Przystupa W., Nowak J. 2017. *Wytrzymałość na ściskanie granul superfosfatu i mocznika. Compression strength of superphosphate and urea granules.* Przemysł Chemiczny 96 (9) 1963–1967.

Portnikov D., Kalman H., Aman S., Tomas J. 2013. *Investigating the testing procedure limits for measuring particle strength distribution.* Powder Technology, 237, 489–496.

Raimondi G., Maucieri C., Toffanin A., Renella G., Borin M. 2021. *Smart fertilizers: what should we mean and where should we go?* Italian Journal of Agronomy, 16(AOP), 1794.

Rozenblat Y., Portnikov D., Levy A., Kalman H., Aman S., Tomas J. 2011. *Strength distribution of particles under compression.* Powder Technology, 208, 215–224.

Salman A.D., Fu J., Gorham D.A., Hounslow M.J. 2003. *Impact breakage of fertilizer granules.* Powder Technology, 130, 359– 366.

Subero-Couroyer C., Ghadiri M., Brunard N., Kolenda F. 2003. *Weibull analysis of quasi-static crushing strength of catalyst particles.* Trans IChemE Journal, 81A, 953–962.

Testing the relationship between two groups of features

ANNA SZCZEPAŃSKA-ÁLVAREZ[1*], MALWINA MROWIŃSKA[1]

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

[*]anna.szczepanska-alvarez@up.poznan.pl

Studying the association between traits allows to understand the nature of variable relationships in different fields of science e.g. in agriculture, medicine, biology, chemistry, meteorology and genetics. In this talk we analyze the relationship between two groups of characteristics: the group of variables observed over time and the features which are unchanged during the experiment. Assuming that the observations have a multivariate normal distribution with block-structure of the covariance matrix, we can use the likelihood ratio test to verify the separability of the top- left block of the covariance matrix and the independence of two groups of characteristics. The results are illustrated by the experimental data.

Keywords: block covariance structure, separable covariance structure, maximum likelihood estimation

References

Lu N., Zimmerman D.L. 2005. *The likelihood ratio test for a separable covariance matrix*. Statistics & Probability Letters 73:449–457.

Manceur A., Dutilleul P. 2013. *Maximum likelihood estimation for the tensor normal distribution: Algorytm, minimum sample size and empirical bias and dispersion*. Journal of Computational and Applied Mathematics 239:37-49.

Seber G.A.F. 1984. Multivariate observations. Wiley series in probability and mathematical statistics. John Wiley $&$ Sons, New Jersey.

Probability distribution of constrained maximum coordinates

MARTIN TLÁSKAL

Central Institute for Supervising and Testing in Agriculture, National Plant Variety Office, Brno, Czech Republic

martin.tlaskal@ukzuz.cz

The paper deals with the problem of finding the probability distribution of the random vector that determines constrained maximum coordinates of a regression function. First the asymptotic distribution of the unconstrained maximum is derived. Then the distribution of constrained maximum coordinates is suggested for data from the food industry regarding the influence of emulsifying salts on the hardness of processed cheese.

Keywords: Delta method, emulsifying salts, processed cheese, regression analysis

References

Buňka F., Doudová L., Weiserová E., Kuchař D., Michálek J., Slavíková Š., Kráčmar S. 2012. *The effect of different ternary mixtures of sodium phosphates on hardness of processed cheese spreads*. International Journal of Food Science & Technology 47: 2063-2071.

Buňka F., Doudová L., Weiserová E., Kuchař D., Ponížil P., Začalová D., Nagyová G.,Pachlová V., Michálek J. 2013. *The effect of emulsifying salt composition and cheese maturity on the textural properties of processed cheese*. International Dairy Journal 29 (1): 1-7.

Greene W. H. 2003. Econometric Analysis (5th ed.). Prentice-Hall.

Manjunath B. G., Wilhelm S. 2012. *Moments calculation for the doubly truncated multivariate normal density*. arXiv:1206.5387.

Preparing high quality speech corpora for speech data analysis

ALICJA WIECZORKOWSKA

Department Multimedia
Polish-Japanese Academy of Information Technology, Poland

alicja@poljap.edu.pl

With ubiquitous presence of smartphones, and with smart devices gaining popularity, speech-based communication is becoming more and more common. Speech corpora are needed for training the algorithms for automatic speech recognition, and to prepare speech synthesis tools. Such corpora are also very useful to analyze speech data for various purposes. Therefore, speech corpora are being created all over the world, for various languages, sometimes with thousands of hours of speech (Chen et al., 2021). Additionally, speech corpora are usually annotated (Bird and Harrington, 2001), as annotated corpora allow connecting speech and language search.

There are many types of speech corpora, and various criteria can be applied to evaluate their quality (Egbert et al, 2022). For instance, database-related quality criteria can be used, but language-related criteria or technical quality of the recordings are more typical here. The planning phase of the project of the speech corpus creation should start with establishing the principles for guiding the entire process of the corpus creation.

It this talk, the components of speech corpus quality evaluation will be presented, and issues related to obtaining high quality corpora will be discussed.

Keywords: speech signal, speech corpora

References

Bird, S., Harrington, J., 2001. Editorial. Speech annotation and corpus tools. Speech Communication 33, 1-4.

Chen, G. Chai, S. Wang, G. Du, J., Zhang, W.-Q., Weng, C., Su, D., Povey, D.,. Trmal, J., Zhang, J., Jin, M., Khudanpur, S., Watanabe, S., Zhao, S., Zou, W., Li, X., Yao, X., Wang, You, Z., Yan, Z., 2021: GigaSpeech: An Evolving, Multi-domain ASR Corpus with 10,000 Hours of Transcribed Audio. INTERSPEECH 2021, http://dx.doi.org/10.21437/Interspeech.2021-1965

Egbert, J., Biber, D., & Gray, B., 2022. Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness. Cambridge University Press.

Variation of tree crown sizes of pine trees in age classes

BOGNA ZAWIEJA[1*], KATARZYNA KAŹMIERCZAK[2]

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

[2]Department of Forest Management
Poznań University of Life Sciences, Poland

[*]bogna.zawieja@up.poznan.pl

The crown of the tree is one of the most important parts of the tree, as it determines the basic life functions. It is also the basis for determining the condition of the tree. Its size is affected by both the age of the tree and its position in the biosocial structure of the stand. Measurements of pine features were made on circular surfaces of successive age subclasses from IIa to Vb and VIIa. Tree trunk features were measured: dbh (cm), total tree height (m), live crown height (m) and the height of the widest part of the crown (m). The width of the crown (m) was also measured and the crown features were assumed as: crown projection area, crown length, crown length of light and shade, share of crown light and shadow, relative crown length expressed as a percentage, crown volume, crown deflection coefficient and Seebach's growth space number.

Then, a nuclear discriminant analysis was performed to check whether trees belonging to different age groups could be distinguished in terms of the measured tree characteristics. This analysis was performed separately for the features of the trunk and the features of the crown.

Keywords: light-crown, shadow-crown, Kraft's classes, Age classes, Scots pine

References

Baudat G, Anouar F. 2000. Generalized discriminant analysis using a kernel approach. *Neural Comput*. Oct;12(10):2385-404. doi: 10.1162/089976600300014980. PMID: 11032039.

Grala-Michalak J., Kaźmierczak K. 2011. Discriminant analysis for Kraft's classes of trees. *Biometrical Letters* 48 (1): 67-81.

Kaźmierczak K. 2009. Wybrane miary przestrzeni wzrostu pojedynczego drzewa w bliskorębnym drzewostanie sosnowym. *Sylwan* 5: 298-303.

Kaźmierczak K. 2010. Kształtowanie się wybranych cech przestrzeni wzrostu pojedynczego drzewa w 50-letnim drzewostanie sosnowym. *Sylwan* 154(4): 267-274.

Kaźmierczak K. 2012. Długość koron sosen w drzewostanach trzech klas wieku. *Forestry Letters* 103: 25-32.

Kaźmierczak K. 2012. Przestrzeń wzrostu sosny w 35-letnim drzewostanie na przykładzie wybranych miar przestrzeni wzrostu pojedynczego drzewa. *Sylwan* 156(4): 280-286.

Kaźmierczak K. 2013. Biosocjalne zróżnicowanie miąższości grubizny sosen drzewostanów trzech klas wieku oraz jej związki z przestrzenią wzrostu drzewa. *Sylwan* 157(2): 122-129.

Kaźmierczak K. 2013. Zróżnicowanie bieżącego przyrostu sosen w trzech klasach wieku. Leśne Prace Badawcze 74(2): 93-100.

Kaźmierczak K. 2014. Length and relative length of crowns in 88-year old pines and its relationships with volume increment. *Colloquium Biometricum* 44: 89-95.

Kaźmierczak K. 2017. Biosocjalne zróżnicowanie względnej i bezwzględnej długości koron dębów w 56-letnim drzewostanie. *Acta Sci. Pol. Silv. Colendar. Ratio Ind. Lignar*., 16 (1): 39-46. DOI: 10.17306/J.AFW.2017.1.4

Kaźmierczak K., Korzeniewicz R., Borzyszkowski W. 2016. Biosocjalne zróżnicowanie wzrostu i przyrostu 85-letnich sosen. *Acta Sci. Pol. Silv. Colendar. Ratio Ind. Lignar*., 15(2), 73–78. DOI: 10.17306/J.AFW.2016.2.9

Kaźmierczak K., Najgrakowski T. 2012. Tree crown dimensions and its relationships with tree volume based on Scots pine (Pinus sylvestris L.). *Colloquium Biometricum* 42: 17-23.

Krzyśko, M., Waszak, L., Wołyński, W. 2014. Comparison of Kernel Discriminant Coordinates Algorithms. *Communications in Statistics - Simulation and Computation* 43: 2138 - 2148.

Turski M., Kwaśna H., Beker C., Jaszczak R., Kaźmierczak K., Najgrakowski T., Borzyszkowski W. 2019. The influence of age and crown position on growth efficiency along a Scots pine chronosequence. *iForest - Biogeosciences and Forestry* 12: 474-479. DOI: 10.3832/ifor2953-012

Zawieja B., Kaźmierczak K. 2015. The method of standing trees allocation to different biosocial classes. *Colloquium Biometricum* 45, s. 79-92.

Applying Neural Networks for Mustard Variety Recognition

BOGNA ZAWIEJA*, LAURA SLEBIODA

Department of Mathematical and Statistical Methods
Poznań University of Live Sciences, Poland

*bogna.zawieja@up.poznan.pl

This paper presents the application of neural networks for mustard variety recognition. Mustard plants exhibit significant variations in morphology across different varieties, making their identification challenging. The proposed method utilizes a convolutional neural network (CNN) architecture to extract discriminative features from input attributes of mustard plants. The dataset used for training and evaluation consists of 57 diverse mustard varieties, with each variety annotated with corresponding 15 features. Data preprocessing involves standardization and one-hot encoding techniques to ensure optimal model performance. The trained CNN achieves high accuracy in classifying mustard varieties, demonstrating its effectiveness in automating the recognition process. Additionally, a threshold-based approach is introduced to identify cases where the predicted variety deviates significantly from known varieties. By setting a threshold for the percentage of different varieties among predictions, the method can detect the presence of new varieties. Experimental results demonstrate the robustness and potential practical application of the proposed method in mustard variety recognition. This research contributes to the field of agricultural automation by providing an efficient and reliable approach for identifying mustard plant varieties using neural networks.

Keywords: Mustard Plant, Neural Networks, Variety Recognition

References

Chollet, F. 2021. Deep Learning with Python (Second Edition). *Manning*.

Thomas, J., Kuruvilla, K. M., Hrideek, T. K. 2014. 21 - Mustard. *Handbook of Herbs and Spices (Second Edition), Woodhead Publishing Series in Food Science, Technology and Nutrition*, 388-398.

Bisen, D. 2021. Deep convolutional neural network based plant species recognition through features of leaf. *Multimed Tools Appl* **80**, 6443–6456. DOI: https://doi.org/10.1007/s11042-020-10038-w