

**The 48th International
Biometrical Colloquium
in Honour of the 90th Birthday
of Professor Tadeusz Caliński
and
VI Polish – Portuguese Workshop
on Biometry**

Abstracts



Szamotuły, Poland, September 9-13, 2018

**48th International Biometrical Colloquium
and
VI Polish – Portuguese Workshop on Biometry
Szamotuły, Poland**

September 9-13, 2018

Organizatorzy konferencji
Organizers

Polskie Towarzystwo Biometryczne
Polish Biometric Society

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu
Department of Mathematical and Statistical Methods
Poznań University of Life Sciences

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu
Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań

ISBN 978-83-64246-95-1

Published by Polish Biometric Society and Wydawnictwo Drukarnia PRODRUK

The Scientific Committee

Janusz Gołaszewski (Uniwersytet
Warmińsko Mazurski w Olsztynie, Polska)

Ivette Gomes (Faculdade de
Ciências, Universidade de Lisboa,
Portugalia)

Małgorzata Graczyk (Uniwersytet
Przyrodniczy w Poznaniu, Polska)

Zofia Hanusz (Uniwersytet Przyrodniczy
w Lublinie, Polska)

Zygmunt Kaczmarek (Instytut Genetyki Roślin
PAN w Poznaniu, Polska)

Marcin Kozak (Wyższa Szkoła Informatyki
i Zarządzania w Rzeszowie, Polska)

Maria Kozłowska (Uniwersytet
Przyrodniczy w Poznaniu, Polska)

Mirosław Krzyśko (Uniwersytet im.
A. Mickiewicza w Poznaniu, Polska)

Izabela Kuna-Broniowska
(Uniwersytet Przyrodniczy w Lublinie,
Polska)

Wiesław Mądry (Szkoła Główna
Gospodarstwa Wiejskiego w Warszawie,
Polska)

Iwona Mejza (Uniwersytet Przyrodniczy
w Poznaniu, Polska)

Stanisław Mejza (Uniwersytet Przyrodniczy
w Poznaniu, Polska)

Karl Moder (University of Natural Resources
and Life Sciences, Vienna, Austria)

Manuela Neves (ISA, Universidade Técnica
de Lisboa, Portugalia)

Teresa Oliveira (Universidade Aberta,
Lizbona, Portugalia)

Amilcar Oliveira (Universidade Aberta,
Lizbona, Portugalia)

Dulce Pereira (Universidade de Évora,
Portugalia)

Paulo Canas Rodrigues (FCT, Universidade
Nova de Lisboa, Portugalia)

Lisete Sousa (FCUL and CEAUL, Portugal)

Mirosława Wesołowska-Janczarek
(Uniwersytet Przyrodniczy w Lublinie,
Polska)

Waldemar Wołyński (Uniwersytet im.
A. Mickiewicza w Poznaniu, Polska)

Bogna Zawieja (Uniwersytet
Przyrodniczy w Poznaniu, Polska)

Komitet organizacyjny Organizing Committee

Maria Kozłowska-Chair

Małgorzata Graczyk- Co-Chair

Jan Bocianowski

Bogna Zawieja

Katarzyna Ambroży-Deręgowska

Anna Budka

Ewa Skotarczak

Abstracts

Application of bioconcentration factors as a useful tool for evaluation of compost effect in the stabilization process

EWA BAKINOWSKA^{1*}, MONIKA JAKUBUS², NATALIA TATUŚKO²

¹Institute of Mathematics
Poznań University of Technology, Poland

²Department of Soil Science and Land Protection
Poznań University of Life Sciences, Poland

*ewa.bakinowska@put.poznan.pl

The paper concerns the study of contamination soil and plant by heavy metals. The study investigated the effect of compost on heavy metal uptake by two test plants: winter barley and white mustard. The content of metal in the plant was tested using bioconcentration factors (BCF_T , BCF_A) and the pollution level factor (CCL). The investigations were carried out under greenhouse conditions with two soils (light and medium). Our goal was to compare plants (winter barley and white mustard) in terms of average BCF_T (BCF_A and CCL) values on soils with the addition of metal (copper or zinc). The adequate Student's t-test was used for comparison of such two plants with respect to values of BCF_T (BCF_A and CCL). The BCF_T and BCF_A values strongly depend on the plant species. Presented study proved this relationship. Independently of the applied metal and its dose the BCF_T and BCF_A values were always significantly lower for winter barley. Additionally the influence of compost amendment on BCF_T , BCF_A and CCL values was tested. In most cases the values of BCF_T , BCF_A and CCL were significantly lower in the conditions when compost was applied. The violin graphs of bioconcentration factors allowed to determine threshold of toxicity of plants.

Keywords: bioconcentration factors, biowaste compost, heavy metals

Acknowledgements

This study was partially funded by the Ministry of Science and Higher Education (grant number 04/43/DSPB/0101).

Assessing importance of variables in classification problems using collective intelligence methods

ANNA M. BARTKOWIAK^{1*}

¹Institute of Computer Science
University of Wrocław, Poland

*aba@cs.uni.wroc.pl

Say, we have p -dimensional data matrix \mathbf{X} containing n observations coming from two groups of data, for example n patients, some of them suffering from ventilatory disorders, and the others being in normal state, that is not showing symptoms of such disorders. Our goal is to build an algorithm, a discriminant function, which would allow to make a medical diagnosis on the basis of the recorded data. Specifically, the problem to be considered is: Is it necessary to take all the p observed variables as input to the constructed algorithm. May be a subset of size $k < p$ would be sufficient for that purpose?

The problem has been considered successfully in various aspects by applying various statistical methods. The solution depends on the recorded data matrix \mathbf{X} . Taking another sample, the results will differ. To speak about 'significance' one should know the probability distribution of the data. In practice this may cause serious problems.

Quite different approach is offered by more recent Machine Learning methods, especially those working in a supervised mode (see discussion in ref. Breiman 2001). In the following, I will consider an interesting methodology called random forests (RFs), see ref. James et al. (2001). RFs can be used both in regression and classification problems. The methodology uses the concepts of Collective Intelligence and has favorable properties:

- (i) The RFs work directly on original variables (not on new features derived from them).
- (ii) The RFs can work on mixed type variables, that is quantitative (numeric) or qualitative (categorical).
- (iii) The RFs work without assumption on the probability distribution of the variables.
- (iv) The RFs yield an internal unbiased estimate of the generalization error.
- (v) The RFs offer some non-conventional indices of importance of variables in the context of regression and clustering.
- (vi) Moreover, it has been shown that RFs are in principle resistant to outliers.

Generally, the principle of the Collective Intelligence methods is to make inference relying on large amount of diversified data sets and combine the results to such ones that appear 'best' for all the ensemble of the created subsets. The RFs use a large amount of binary decision trees constructed from bootstrapped samples. On the base of the huge number of subsets some indices of importance of variables in contributing to the final decisions is offered (like; MeanDecreaseAccuracy, MeanDecreaseGini) permitting to take for the analysis only those variables, that really matter. The nodes of the trees are established using a randomized subset of available variables. The methodology of decision trees and random forests can be found in ref. James et. al. (2013).

My goal is to study the work of the RF methodology as applied to medical data, where the task is screening healthy state persons from those suffering from ventilatory disorders. The analysis will be performed using the R packages 'Random Forest' (Leo Breiman and Adele Cutler, 2018) and 'Tree' (Brian Ripley, 2018).

Keywords: classification, decision tree, random forest, importance of variables

References

- Breiman L. 2001. *Statistical modeling: the two cultures*. Statistical Science 16 (3): 199-231.
James G., Witten D., Hastie T., Tibshirani R., 2013. *An introduction to statistical learning: with applications in R*. Springer Science+Business Media New York.

Evaluation of deterministic interpolation methods for mapping selected parameters of groundwater quality

URSZULA BRONOWICKA-MIELNICZUK^{1*}, JACEK MIELNICZUK¹, RADOMIR
OBROŚLAK²

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

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

*urszula.bronowicka@up.lublin.pl

The aim of the study was to analyze spatial variability of selected parameters of subsurface waters by using deterministic interpolation methods. In the study we compared five interpolation methods such as inverse distance weighting, modified Shepard's method, triangulation with linear interpolation and radial basis function with two systems of basis functions: multiquadratic and thin plate spline. The cross validation was applied to evaluate the accuracy of interpolation methods through the root mean square error, the mean absolute error, the relative mean absolute error, the relative mean square error and Willmott's index of agreement. The methods which proved to be optimal were used to create spatial variability maps of the analyzed parameters.

Geostatistical interpolation and visualization were performed in Surfer ver. 15. All further data analysis was carried out using R software (ver. 3.4.0).

Keywords: cross-validation, deterministic interpolation methods, groundwater quality parameters, prediction maps

References

- Duchon J. 1977. *Splines minimizing rotation invariant semi-norms in Sobolev spaces. Constructive theory of functions of several variables*, W. Schempp and K. Zeller (eds.), Springer, Berlin, 85-100.
- Franke R., Nielson G. 1980. *Smooth interpolation of large sets of scattered data*. International Journal for Numerical Methods in Engineering 15: 1691-1704.
- Hardy R.L. 1990. *Theory and applications of the multiquadric-biharmonic method*. Computers & Mathematics with Applications 19: 163-208.

- Li J. 2016. *Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained*. *Environmental Modelling & Software* 80: 1-8.
- Li J., Heap A.D. 2011. *A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors*. *Ecological Informatics* 6: 228–241.
- Rocha H. 2009. *On the selection of the most adequate radial basis function*. *Applied Mathematical Modelling* 33: 1573-1583.
- Shepard D. 1968. *A two dimensional interpolation function for irregularly spaced data*. *Proc. 23rd Nat. Conf. ACM*: 517-523.
- Wackernagel H. 1998. *Multivariate geostatistics. An Introduction with applications*. 2nd completely revised edition. Springer.
- Webster R., Oliver M. 2001. *Geostatistics for environmental scientists*. John Wiley & Sons, Ltd, Chichester.
- Willmott C.J. 1981. *On the validation of models*. *Physical Geography* 2: 184-194.

How many indicator species are required to assess the ecological status of a river?

ANNA BUDKA^{1*}, KRZYSZTOF MOLINIŃSKI¹, KRZYSZTOF SZOSZKIEWICZ²

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

²Department of Ecology and Environment Protection,
Poznań University of Life Sciences, Poland

*anna.budka@up.poznan.pl

The presented research was focused on macrophytes, which constitute a primary element in the assessment of the ecological status of surface waters following the guidelines of the Water Framework Directive. In Poland, such assessments are conducted using the Macrophyte Index for Rivers (MIR). The aim of this study was to characterize macrophyte species in rivers in terms of their information value for the evaluation of the ecological status of rivers. The macrophyte survey was carried out at 100 river sites in the lowland area of Poland. Botanical data were used to verify the completeness of the samples (number of species). In the research the information provided by each species is controlled. Entropy was used as a fundamental part of the analysis of information. This analysis showed that the application of a standard approach to study macrophytes in the case of rivers is likely to provide sample underestimation (with missing species). This may potentially lead to incorrect determination of MIR and thus result in a defective environmental decision. On this basis, a sample completeness criterion was constructed, using in which the average value of information for macrophyte species in medium-sized lowland rivers is sufficient for the site to be considered as representative.

Keywords: entropy, indicator taxa, information vector, macrophytes, Macrophyte Index for Rivers (MIR)

References

- Eryomin A.L. 1998. *Information ecology – a viewpoint*. Int. J. Environ. Stud. Sections A&B 3/4: 241–253. DOI.org/10.1080/00207239808711157.
- Dudley S.A., Murphy G.P., File A.L. 2013. *Kin recognition and competition in plants*. Func. Ecol. 27: 898–906. DOI: 10.3732/ajb.0900006.
- Groen C.L.G., Stevers R.A.M., Van Gool C.R., Broekmeyer M.E.A. 1993. *Elaboration of the ecotope system*. Phase III, CML report, 49.

Regular D-optimal weighing designs with non-negative correlations of errors constructed from some block designs

BRONISŁAW CERANKA, MAŁGORZATA GRACZYK*

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

*malgorzata.graczyk@up.poznan.pl

In this paper, the issues related to the regular D-optimal chemical balance weighing design are considered. We study these designs under assumption: the measurement errors are equally non-negative correlated they have the same variances. Here we present the existence conditions of the regular D-optimal design and construction methods.

Keywords: balanced bipartite weighing design, balanced incomplete block design, chemical balance weighing design, D-optimality

References

- Ceranka B., Graczyk M. 2015. *On D-optimal chemical balance weighing designs*. Acta Universitatis Lodzianis, Folia Oeconomica 311: 71-84.
- Ceranka B., Graczyk M. 2016. *New construction of D-optimal weighing designs with non-negative correlations of errors*. Colloquium Biometricum 46: 31-45.
- Katulska K., Smaga Ł. 2013. *A note on D-optimal chemical balance weighing designs and their applications*. Colloquium Biometricum 43: 37-45.

Effect of cultivar, crop management, location and growing seasons for grain yield of Triticale

ADRIANA DEREJKO^{1*}, MARCIN STUDNICKI¹

¹Department of Experimental Design and Bioinformatics
Warsaw University of Life Science – SGGW, Poland

*adriana_derejko@sggw.pl

Triticale (*Triticosecale* Wittmack) is obtained through the crossing of wheat (*Triticum* ssp.) and rye (*Secale cereale* L) and is characterized by high yield potential, good health and grain value, and high tolerance to biotic and abiotic stress. Poland is very important for breeding progress in triticale since it is home to most cultivars; numerous genetic studies on triticale have been carried out. Despite the tremendous interest of triticale, both by breeders and researchers, there are no studies assessing the adaptation of cultivars to environmental conditions across growing seasons. This study was conducted to investigate the influence of cultivar, management, location and growing season on grain yield. At the same time, this approach provides a new way to determine whether there is any dependency between the 8 seasons and to find the cause of the yield response to environmental conditions in a given growing season.

Keywords: cultivar and environmental interaction, growing seasons, linear mixed models

References

- Ammar K, Mergoum M, Rajaram S. 2004. The history and evolution of triticale. In: Mergoum M, Gómez-Macpherson H, editors. *Triticale improvement and production*. Rome: Food and Agriculture Organization of the United Nations: 1–10.
- FAO Plant Production and Protection 2004. *Triticale improvement and production*. Paper 179. Roma.
- FAO 2015. *Statistical Yearbook 2015. Word Food and Agriculture*, Roma.
- Hammer G. L., McLean G., Chapman G., Zheng B., Doherty A., Harrison M. T., van Oosterom E., Jordan D. 2014. *Crop design for specific adaptation in variable dryland production environments*. *Crop & Pasture Science* 65: 614-626.
- Ito D., Afshar R. K., Chen C., Miller P., Kephart K., McVay K., Lamb P., Miller J., Bohannon B., Knox M. 2016. *Multienvironmental evaluation of dry pea and lentil cultivars in Montana using the AMMI Model*. *Crop Science* 56: 520-529.
- Liu S. M., Constable G. A., Reid P. E., Stiller W. N. Cullis B. R. 2013. *The interactions between breeding and crop managements in improved cotton yield*. *Field Crop Research* 148: 49-60.

Acknowledgements

This research is based on the partial results of the Postregistration Variety Testing System (PVTS), across growing seasons 2007/2008-2014/2015 contains 11 modern triticales cultivars came from Polish breeding programmes evaluated across 57 locations.

The problems of multivariate analyses (PCA and FA) of categorical - ordinal variables in sociology. Part B.

Education/grading evaluation.

PAVEL FL'AK^{1*}, SILVIA BACHEROVÁ²

¹ Biometry Commission of the Slovak Academy of Agricultural Sciences in Nitra, Slovak Republic

² Ambulance of Clinical Psychology, Ltd. in Hlohovec, Slovak Republic

*flakpavell@gmail.com

The Principal Component Analysis (PCA) and Factor Analysis (FA) are widely used statistical techniques in the psychological, social and educational sciences. In these scientific areas are analysed not only continuous variables but mainly categorical ones (i.e. binary, dichotomous, mainly ordered categories - ordinal variables). In practice most researchers treat ordinal variables with 5 or more categories as continuous. But ordinal variables (e.g. measured by Likert scales) with the few categories are analysed as categorical variables. The ordinal variables with many categories are often nonnormal and also are characterized by nonhomogeneous in variability. There are various specific statistical approaches of analysis of these data, e.g. by Ordinal Factor Analysis (OFA). In our previous paper (Part A) was presented applications of OFA on the samples of averages of tests of psychology (analyses of various degrees of depression, anxiety, experiences of close relationships and emotional schemas), on the healthy and non-healthy patients from clinic in Slovakia. In the present paper are analysed by univariate and multivariate statistical methods (PCA, FA and OFA) of the knowledge of students by exam tests-marks in secondary (gymnasiums) education.

Keywords: biometrics, categorical variables, education, multivariate analysis, sociology

The variability of the chlorophyll fluorescence parameters and their relationship with the seed yield and its components in the narrow leafed lupin (*Lupinus angustifolius* L.)

BARBARA GÓRYNOWICZ^{1*}, WOJCIECH ŚWIĘCICKI¹, WIESŁAW PILARCZYK²,
WOJCIECH MIKULSKI¹

¹Legume Comparative Genomics Team
Institute of Plant Genetics of the Polish Academy of Sciences in Poznań, Poland

²Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

*bgom@igr.poznan.pl

Narrow leafed lupin (*Lupinus angustifolius* L.) is one of the most important legume species and it plays a significant role in plant and animal production due to high protein content. The aim of this study was to assess the variability of the selected chlorophyll fluorescence parameters and their correlation with the seed yield and its components in narrow leafed lupin cultivars at the flowering and green leaf phases. In 2011-2016 field experiments with selected narrow leafed lupin cultivars were conducted (at 2 locations and 4 replicates in 2011, at 4 locations and 6 replicates in 2012-2014 and at 2 locations and 3 replicates in 2015-2016). In 2012-2016 chlorophyll fluorescence measurements in vivo plants in field conditions were made using the mobile fluorimeter HANDY PEA (*Handy Plant Efficiency Analyser*, Hansatech Instruments Ltd., King's Lynn, Norfolk, UK). On the basis of the measured values, the parameters characterizing the energy conversion in photosystem II (PSII) were calculated using the fluorescent JIP test (Lazár 1999). Seed yield, number of pods, number of seeds per plant, number of seeds per pod and thousand seed weight were assessed on the basis of representative sample of plants from every plot. The variability of chlorophyll fluorescence parameters for individual cultivars of narrow leafed lupin over measurement terms was illustrated graphically. The results of the measurements were analyzed statistically independently for each year using the analysis of variance (ANOVA) under a linear model for randomized complete block design (Elandt 1964) and for a series of experiments under a mixed model usually used in such analysis (Cochran and Cox 1950). The analysis of correlations, the analysis of multiple regression (Draper and Smith 1973) and the

accumulated analysis of variance were applied in order to identify relationships between seed yield and its components and chlorophyll fluorescence parameters, using the MS EXCEL 2016 program and the GENSTAT package (Payne et al. 1987). Based on statistical analysis, the five chlorophyll fluorescence parameters (Fv/Fm, psi0, P.I.csm, ETO/CS, ETO/RC) were selected for further research. Parameters defining the energy flows in terms of cross section and reaction center (ETO/CS and ETO/RC) were positively correlated with seed yield, number of pods and number of seeds per pod, and negatively correlated with thousand seed weight. Other parameters (Fv/Fm, psi0, P.I.csm) were negatively correlated with seed yield and thousand seed weight and generally positively correlated with number of pods and number of seeds per pod. The significant correlation between seed yield and ETO/CS and ETO/RC at the beginning of the green leaf phase suggests the direction of further research aimed at the elaboration of breeding method for seed yield using the mobile equipment. Measurements of photochemical activity of PSII showed a general decrease in all analyzed parameters, indicating poorer condition of leaves of narrow leafed lupin at the end of the growing season. Seed yield appeared to be a trait depending meaningfully on chlorophyll fluorescence parameters.

Keywords: chlorophyll fluorescence parameters, narrow leafed lupin, seed yield

References

- Elandt R. 1964. Statystyka matematyczna w zastosowaniu do doświadczeń rolniczego. Państwowe Wydawnictwo Naukowe, Warszawa: 360-361.
- Cochran W.G., Cox G.M. 1950. Analysis of the results of a series experiments. Experimental Designs, Wiley & Sons, New York: 545-568.
- Draper N.R., Smith H. 1973. Analiza regresji stosowana. Państwowe Wydawnictwo Naukowe, Warszawa.
- Lazár D. 1999. *Chlorophyll a fluorescence induction*. Biochemica et Biophysica Acta 1412: 1-28.
- Payne R.W., Lane P.W., Ainsley A.E., Bricknell K.E., Digby P.G.N., Harding S.A., Leech P.K., Sampson H.R., Todd A.D., Verrier P.J., White R.B. 1987. GENSTAT 5 Reference Manual. Clarendon Press, Oxford, England.

Acknowledgements

This research is based on the partial results of the National Multi-Year Project (Resolution no. 149/2011, 9 August 2011 for 2011-2015 and Resolution no. 222/2015, 15 December 2015 for 2016-2020) financed by the Ministry of Agriculture and Rural Development, Poland.

Parallel Coordinates and Multivariate Functional Data in the Aspect of Correlation

JOLANTA GRALA-MICHALAK

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

grala@amu.edu.pl

Parallel coordinates is a widely used visualization technique for multivariate data and high-dimensional geometry. Parallel coordinates are constructed by placing axes in parallel with respect to 2D Cartesian coordinate system in the plane. This formulation provides a mapping of points to lines and vice-versa, what is named as the point-line duality (Inselberg 1991). More generally, there is a curve-curve duality between Cartesian and parallel coordinates (Inselberg 2009).

It is possible to represent the spatio-temporal data in parallel coordinates but the clutters produced by a large number of lines hide the patterns and correlations contained in the data. The remedy for this problem is to analyze the data visualized in parallel coordinates as the multivariate functional data. The two aspects: Kendall's dependence for ordered data and canonical correlation for continuous data in the context of parallel coordinates will be presented.

Keywords: canonical correlation, functional data, Kendall's correlation, parallel coordinates

References

- Górecki T., Krzyśko M., Ratajczak W., Wołyński W. 2016. *An Extension of the Classical Distance Correlation Coefficient for Multivariate Functional Data with Applications*. *Statistics in Transition new series*, 17 (3): 449-466.
- Inselberg A. 2009. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer Verlag, New York.
- Inselberg A., Dimsdale B. 1991. *Parallel Coordinates. A Tool for Visualizing Multivariate Relations*. In: *Human-Machine Interactive Systems*, (ed. Allen Klinger) serie: Languages and Information Systems, Plenum Press, New York.
- Valencia D., Lillo R.E., Romo J. 2013. *A Kendall Correlation Coefficient for Functional Dependence*. *Statistics and Econometrics Series 28, Working Papers*: 13-32.

PLS-SEM to assess the work-family conflict

LUÍS M. GRILO^{1*}, HELENA L. GRILO²

¹Department of Mathematics, Polytechnic Institute of Tomar (PIT), Portugal; Mathematics and Applications Center (CMA), FCT / UNL; Research Center in Smart Cities (Ci2), PIT; CIICESI / ESTG - P. Porto, Portugal

²Center for Applied Research in Economics and Land Management (CIAEGT) Polytechnic Institute of Tomar, Portugal

*lgrilo@ipt.pt

In almost every human activity sectors worldwide, many workers are exposed to the psychosocial risks. Although these risks are not easy to identify (once they depend on many and varied sources as well as they may occur at every level of the organisations), their effects lead to higher costs in terms of workers' health and safety, for the organization and for the society in general. The exposure to a work environment subjected to psychosocial risks have negative impacts in the social sphere, namely causing conflicts in work and poor family relationships. In order to assess the effects of psychosocial risks in the work-family conflicts of workers from a services Portuguese company, a survey was conducted through an adapted version of the Copenhagen Psychosocial Questionnaire (COPSOQ), with variables expressed in a Likert-type scale. Then, data were analysed using the Structural Equation Model (SEM) and the estimated model was obtained through the Partial Least Squares (PLS) approach (where a nonparametric bootstrap procedure was used to evaluate the statistical significance of the estimated path coefficients with PLS-SEM). The latent variables "recognition" and "quantitative demands" appear as predictors of "stress" and the last two are both direct predictors of "work-family conflict". The occupational experts consider these results of great importance since they try to prevent workers' health, avoiding bad work atmosphere, poor social relationships and absenteeism.

Keywords: Health problems, Job stress, Psychosocial risks, Survey, SmartPLS

References

- Directorate-General for the Humanisation of Labour 2013. Guide to the prevention of psychosocial risks at work. Université de Namur (Flohimont V., Lambert C., Berrewaerts J., Zaghdane S., Füzfa A.). www.employment.belgium.be.
- Grilo L. M., Vairinhos V. 2018. *PLS-PM to evaluate workers' health perception*. V Workshop on Computational Data Analysis and Numerical Methods, Instituto Politécnico do Porto, Felgueiras, Portugal, May 11-12 (p. 61, Book of Abstracts).

- Grilo L.M., Grilo H.L., Gonçalves S.P., Junça A. (2017). *Multinomial Logistic Regression in Worker's Health*. ICCMSE 2017, AIP Conf. Proc. 1906, 110001.
- Hair J.F., Hult G.T.M., Ringle C.M., Sarstedt M. (2007). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd Ed., Thousand Oakes, CA: Sage).
- Kristensen T.S., Borritz M., Villadsen E., Christensen K.B. 2005. *The Copenhagen Burnout Inventory: A new tool for the assessment of burnout*. *Work & Stress* 19: 192-207.
- Maslach C., Leiter M.P. 2016. *Understanding the burnout experience: recent research and its implications for psychiatry*. *World Psychiatry* 15(2): 103-111.
- Maslach C., Leiter M.P. 2008. *Early predictors of job burnout and engagement*. *Journal of Applied Psychology* 93: 498-512.
- Montero-Marin J., Zubiaga F., Cereceda M., Piva Demarzo M.M., Trenc P., Garcia-Campayo J. 2016. *Burnout Subtypes and Absence of Self-Compassion in Primary Healthcare Professionals: A Cross-Sectional Study*. *PLoS ONE* 11(6): e0157499.
- Ringle C. M., Wende S., Becker J.-M. 2015. *SmartPLS 3*. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>.
- Silva C. F. 2016. *Copenhagen Psychosocial Questionnaire (COPSOQ)*. Portuguese version. FCT-MEC, Universidade de Aveiro, *Análise Exacta*.
- Stehlík M., Helpersdorfer Ch., Hermann P., Šupina J., Grilo L.M., Maidana J.P., Fuders F., Stehlíková S. 2017. *Financial and risk modelling with semicontinuous covariances*. *Information Sciences* 394-395, 246-272.

Acknowledgements

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

Use of classification and regression trees (CART) for analyzing determinants of winter wheat yield variation among fields in Poland

MARZENA IWAŃSKA^{1*}, ANDRZEJ OLEKSY³, MARIUSZ DACKO⁴,
TADEUSZ OLEKSIAK², BARBARA SKOWERA⁵, ELŻBIETA WÓJCIK-GRONT¹

¹Department of Experimental Design and Bioinformatics
Warsaw University of Life Science – SGGW, Poland

²Plant Breeding and Acclimatization Institute – National Research Institute in Radzików,
Poland

³Institute of Plant Production, University of Agriculture in Krakow, Poland

⁴Institute of Economic and Social Sciences, University of Agriculture in Krakow, Poland

⁵Department of Ecology, Climatology, and Air Protection
University of Agriculture in Krakow, Poland

*m.iwanska123@gmail.com

Nowadays, one of staple food sources is wheat. Its production requires good environmental conditions which are not always provided. Then, agricultural practices might mitigate the effects of unfavorable weather or poor quality soils. The environmental and crop management variables influence on yield can be evaluated only based on representative long term data collected on farms through well prepared surveys. In this work authors analyzed winter wheat yield variability, among 3868 fields, across the West and East side of Poland for 12 years, dependence on both soil-weather and crop management factors using classification and regression tree (CART) method. The most important crop management variables which misuse can cause low yield of wheat are: the insufficient use of fungicides, phosphorus deficiency, not optimal date of sowing, poor quality of seeds, failure of the herbicides protection, no crop rotation and cultivars with unknown origin not suitable for the region. The environmental variables very important for obtaining high yields are: farms with large agricultural area (10 ha or larger), good quality soils with stable pH. This work allows to propose the strategies helping in more effective winter wheat production based on the identification of a specific region's characteristics crucial for its cultivation.

Keywords: classification and regression trees (CART), determinants of yield, farm data, winter wheat, yield modeling,

References

- Breiman L., Friedman J. H., Olshen R. A., Stone C. J. 1984. *Classification and Regression Trees*. Chapman and Hall (Wadsworth, Inc.), US New York (p. 254).
- Dacko M., Zając T., Synowiec A., Oleksy A., Klimek-Kopyra A., Kulig B. 2016. *New approach to determine biological and environmental factors influencing mass of a single pea (*Pisum sativum* L.) seed in Silesia region in Poland using a CART model*. *European Journal of Agronomy* 74: 29-37.
- Delmotte S., Tittonell P., Mouret J. C., Hammond R., Lopez-Ridaura S. 2011. *On farm assessment of rice yield variability and productivity gaps between organic and conventional cropping systems under Mediterranean climate*. *European Journal of Agronomy* 35 (4): 223-236.
- Ferraro D. O., Rivero D. E., Ghersa C. M. 2009. *An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees*. *Field Crops Research* 112 (2-3): 149-157.
- Krupnik T. J., Ahmed Z. U., Timsina J., Yasmin S., Hossain F., Al Mamun A., Mridha A.I., McDonald A. J. 2015. *Untangling crop management and environmental influences on wheat yield variability in Bangladesh: an application of non-parametric approaches*. *Agricultural Systems* 139: 166-179.
- Lobell D. B., Ortiz-Monasterio J. I., Asner G. P., Naylor R. L., Falcon W. P. 2005. *Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape*. *Agronomy Journal* 97 (1): 241-249.
- Lobell D., Ortiz-Monasterio J., Addams C., Asner G. 2002. *Soil, climate, and management impacts on regional wheat productivity in Mexico from remote sensing*. *Agric. For. Meteorol.* 114: 31–43.
- Loyce C., Meynard J. M., Bouchard C., Rolland B., Lonnet P., Bataillon P., Bernicot M. H., Bonnefoy M., Charrier X., Debote B., Demarquet, T., Duperrier B., Félix I., Heddadj D., Leblanc O., Leleu M., Mangin P., Méausoone M., Doussinault G. 2008. *Interaction between cultivar and crop management effects on winter wheat diseases, lodging, and yield*. *Crop Protection* 27 (7): 1131-1142.
- Macholdt J., Honermeier B. 2017. *Yield Stability in Winter Wheat Production: A Survey on German Farmers' and Advisors' Views*. *Agronomy* 7 (3): 45.
- Montesino-San Martín M., Olesen J. E., Porter J. R. 2014. *A genotype, environment and management (GxExM) analysis of adaptation in winter wheat to climate change in Denmark*. *Agric. For. Meteorol.* 187: 1–13.
- Reidsma P., Ewert F., Lansink A. O., Leemans R. 2010. *Adaptation to climate change and climate variability in European agriculture: The importance of farm level responses*. *Eur. J. Agron.* 32: 91–102.
- Roel A., Firpo H., Plant R. E. 2007. *Why do some farmers get higher yields? Multivariate analysis of a group of Uruguayan rice farmers*. *Computers and Electronics in Agriculture* 58 (1): 78-92.
- Sileshi G., Akinnifesi F. K., Debusho L. K., Beedy T., Ajayi O. C., Mong'omba S. 2010. *Variation in maize yield gaps with plant nutrient inputs, soil type and climate across sub-Saharan Africa*. *Field Crops Research* 116 (1-2): 1-13.

- Tittonell P., Shepherd K. D., Vanlauwe B., Giller K. E. 2008. *Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—an application of classification and regression tree analysis*. *Agric. Ecosyst. Environ.* 123: 137–150.
- Zhang J., Liu Q., Xu M., Zhao B. 2012. *Effects of soil properties and agronomic practices on wheat yield variability in Fengqiu County of North China Plain*. *African Journal of Agricultural Research* 7 (11): 1650-1658.
- Zheng H., Chen L., Han X., Ma Y., Zhao X. 2010. *Effectiveness of phosphorus application in improving regional soybean yields under drought stress: A multivariate regression tree analysis*. *African Journal of Agricultural Research* 5 (23): 3251-3258.
- Zheng H., Chen L., Han X., Zhao X., Ma Y. 2009. *Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: the importance of phosphorus application rates under drought conditions*. *Agric. Ecosyst. Environ.* 132: 98–105.

Multivariate coefficient of variation for functional data

MIROSLAW KRZYŚKO^{1*}, ŁUKASZ SMAGA²

¹Interfaculty Institute of Mathematics and Statistics
The President Stanisław Wojciechowski State University of Applied Sciences in Kalisz,
Poland

²Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań, Poland

*mkrzysko@amu.edu.pl

In this paper, the adaptation of the multivariate coefficient of variation to functional data is considered. Similarly to the coefficient of variation and its multivariate generalizations, the functional multivariate coefficient of variation (FMCV) may be helpful for comparing the relative variation in different populations or the performance of different equipments characterized by univariate or multivariate functional data. Some theoretical properties of the new functional data analysis method are discussed. By using the basis functions representation of the data, it is shown that the FMCV reduces to the multivariate coefficient of variation of a vector of coefficients of that representation, which results in effective computing of the FMCV. The performance of the classical and robust estimators of the FMCV is compared in finite sample setting by simulation studies. The new methods are illustrated on the ECG data.

Keywords: dispersion measure, functional data analysis, multivariate coefficient of variation, robust estimation, variability measure

References

- Albert A., Zhang L. 2010. *A Novel Definition of the Multivariate Coefficient of Variation*. Biometrical Journal 52: 667–675.
- Reyment R.A. 1960. *Studies on Nigerian Upper Cretaceous and Lower Tertiary Ostracoda: part 1. Senonian and Maastrichtian Ostracoda*. Stockholm Contributions in Geology 7: 1–238.
- Van Valen L. 1974. *Multivariate Structural Statistics in Natural History*. Journal of Theoretical Biology 45: 235–247.
- Voinov V.G., Nikulin M.S. 1996. *Unbiased estimators and their applications, Vol. 2, Multivariate Case*. Kluwer, Dordrecht.

Deep Learning in Biometrical Studies

ELŻBIETA KUBERA^{1*}, AGNIESZKA KUBIK-KOMAR¹

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

*elzbieta.kubera@up.lublin.pl

Big companies, such as Google, Facebook, nVidia, Apple and Toyota allocated recently huge amounts of money to Artificial Intelligence (AI) projects development using the promising novel technology called deep learning. For example, created by Google project called DeepMind is a self-learning system aimed at understanding the human's brain.

Deep learning technology, which is also called deep neural networks, is closely related to the AI's original purpose – to mime human brain activities and teach machines to see, hear and understand like humans do. Deep network consists of layers of neurons. In traditional – shallow neural network, such as multilayer perceptron, there are usually only three or four cascaded fully connected layers – one input, one output and no more than two hidden layers of neurons, while deep networks are built of many layers of different types. The set and order of layers is called network's model, and is closely related to specific network purpose. Image recognition is a case where Deep-Belief Network (DBN) (Hinton, 2009; Hinton et al, 2006) is most often used. Multi-object recognition from images is commonly performed by Convolutional Neural Network (CNN) (LeCun et al, 2015). Both network architectures are capable to perform classification directly from the image, so no feature extraction is needed before model building. Sound (voice) recognition is usually performed using Recurrent Network, which is used also for time series prediction.

In this work, deep learning is used to recognize pollen taxa based on microscopic images. The results show that the deep learning method gives quite good results in the difficult task of classifying objects directly from the images. In this case, the pollen expert's work during data preparation can be minimized.

Keywords: biometrical data, classification, deep neural networks

References

- Hinton G. 2009. *Deep belief networks*. Scholarpedia 4 (5): 5947.
Hinton G., Osindero S., Teh Y. W. 2006. *A Fast Learning Algorithm for Deep Belief Nets*. Neural Computation 18 (7): 1527–1554.
LeCun Y., Bengio Y., Hinton G. 2015. *Deep learning*. Nature 521 (7553): 436–444.

Application of multivariate statistical analyzes for comparison of Betula pollen seasons 2001-2016 in Lublin

AGNIESZKA KUBIK-KOMAR

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

*agnieszka.kubik@up.lublin.pl

The aim of the study was finding similar groups of birch pollen season in Lublin in 2001-2016 based on the pollen seasons parameters such as start, end, duration, peak date, peak value and annual pollen sum. Cluster analysis was used along with an optimal number of clusters algorithms as well as principal components analysis. Seasons were grouped in three clusters: 1) characterized by the late start and short duration, 2) characterized by the late end and low concentration of pollen and 3) characterized by the early start, long duration and high concentration of pollen.

Keywords: cluster analysis, optimal number of clusters, principal component analysis, pollen seasons

References

- Andersen T. B. 1991. *A model to predict the beginning of the pollen season*. Grana, 30(1):269–275.
- Calinski T., Harabasz J. 1974. *A dendrite method for cluster analysis*. Communications in Statistics, 3 (1): 1–27.
- Mufti G. B., Bertrand P., Moubarki L. E. 2000. *Determining the number of groups from measures of cluster stability*. Proceedings of International Symposium on Applied Stochastic Models and Data Analysis: 404–413.
- Rousseeuw P. J. 1987. *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics 20 (C): 53–65.
- Stach A., Emberlin J., Smith M., Adams-Groom B., Myszkowska D. 2008. *Factors that determine the severity of Betula spp. pollen seasons in Poland (Poznań and Kraków) and the United Kingdom (Worcester and London)*. International Journal of Biometeorology 52: 311-321.

New developments and applications of Birnbaum-Saunders models

VICTOR LEIVA, HELTON SAULO

The Birnbaum-Saunders model is an asymmetrical statistical distribution that is being widely studied to describe data collected in diverse areas. We discuss new developments and applications of Birnbaum-Saunders models. A formal motivation is presented, by means of the proportionate-effect law, to use the Birnbaum-Saunders model as a useful distribution for environmental, engineering, medical and neurological variables. The methodologies discussed in this work include several types of approaches. Applications with real-world data sets are carried out for each discussed methodology.

On modeling and analysis of barley malt data in years

IWONA MEJZA^{1*}, KATARZYNA AMBROŻY-DERĘGOWSKA¹, JAN BOCIANOWSKI¹,
JÓZEF BŁAŻEWICZ², MAREK LISZEWSKI³, KAMILA NOWOSAD⁴, DARIUSZ
ZALEWSKI⁴

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

²Department of Fermentation and Cereals Technology
Wrocław University of Environmental and Life Sciences, Poland

³Institute of Agroecology and Plant Production
Wrocław University of Environmental and Life Sciences, Poland

⁴Department of Genetics, Plant Breeding and Seed Production
Wrocław University of Environmental and Life Sciences, Poland

*iwona.mejza@up.poznan.pl

Main purpose of the paper was model fitting of data deriving from a three-year experiment with barley malt. Two linear models were considered. They are fixed linear model with fixed effects of years and other factors and mixed linear model with random effects of years and fixed effects of other factors. Two cultivars of brewing barley, Sebastian and Mauritia, six methods of nitrogen fertilization and four germination days terms were analysed. Three quantitative traits: practical extractivity of the malt, a malting productivity, and a quality coefficient Q have been observed. The initial point for the statistical analyses was an available experimental material, which were barley grain samples destined for the malt. These analyses were performed in the series of years with respect to fixed or random effects of years. Due to a strong differentiation of the years of research and some significant interactions of factors with years, annual analyses were also carried out.

Keywords: complete randomized design, fixed linear model, mixed linear model, HSD Tukey's test, brewing barley, extractivity, malt, malting productivity, quality coefficient

Regularization of the covariance matrix for metabolomic data

ADAM MIELDZIOC¹, MONIKA MOKRZYCKA^{2*}, ANETA SAWIKOWSKA^{1,2}

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

² Institute of Plant Genetics Polish Academy of Sciences in Poznań, Poland

*mmok@igr.poznan.pl

For data obtained by gas chromatography coupled with mass spectrometry (GC–MS), concerning the drought resistance in barley, the problem of characterization of a covariance structure is considered with the use of two methods. The first one is based on Frobenius norm and the second one on the entropy loss function. For the most commonly used covariance structures: compound symmetry, three-diagonal and penta-diagonal Toeplitz and autoregression of order one, the most relevant structure for the Frobenius norm is the penta-diagonal Toeplitz matrix whilst for the entropy loss function – compound symmetry structure (cf. Mieldzioc et al. 2018). Discrepancies obtained from the regularization are relatively large, therefore, we continue regularization of the covariance matrix with the use of hierarchical clustering and heatmaps. We show that for Frobenius norm regularization method a block portioned structure with compound symmetry diagonal blocks is better choice.

Keywords: autoregression matrix, banded Toeplitz matrix, compound symmetry matrix, covariance structure, entropy loss function, estimation, Frobenius norm, regularization

References

Mieldzioc A., Mokrzycka M., Sawikowska A. 2018. *Estimation of the covariance matrix for metabolomic data*. Submitted.

Acknowledgements

This research is partially supported by the project no. WND-POIG.01.03.01-00-101/08.

Testing interaction in unrepeatd designs

KARL MODER^{1*}

¹Department of Landscape, Spatial and Infrastructure Sciences
University of Natural resources and Life Sciences in Vienna, Austria

*karl.moder@boku.ac.at

Unrepeatd experimental designs like Randomized Complete Block Designs, Incomplete Block Designs, split-plot designs are probably the most widely used experimental designs. Despite many advantages they suffer from one serious drawback: In a common linear model there is no test on interaction effects in ANOVA, as there is only one observation for each combination of a block and factor effects.

Several people tried to overcome this problem by using some additional restrictions to the model. None of these methods are used in practice, especially as most of them are non-linear. A review on such tests is given by Karabatos [2005] and Alin & Kurt [2006].

Here a new method is introduced which permits a test of interactions in non-repeatd designs. The underlying models are linear and identical to those of common factorial designs.

A big advantage of the proposed model is, that one can use any common statistical program packages like SAS, SPSS, R ... for analysis by doing a few additional calculations.

Keywords: unrepeatd designs, block designs, interaction, power

References

- Karabatos G. 2005. *Additivity Test*. Encyclopedia of Statistics in Behavioral Science. Wiley, New York. 25-29.
- Alin A., Kurt S. 2006. *Testing non-additivity (interaction) in two-way anova tables with no replication*. Statistics in Medicine 15: 63-85.

Genomic DNA k-mers: basic notions and applications

MARIA NUC^{*}, PAWEŁ KRAJEWSKI^{**}

¹Institute of Plant Genetics, Polish Academy of Sciences in Poznań, Poland

^{*}mnuc@igr.poznan.pl, ^{**}pkra@igr.poznan.pl

Having a genomic DNA sequence, k-mers are all the words of a given length k contained in this sequence. This simple tool is commonly and widely used in biological big data analysis: in sequence assembly, sequence alignment, checking quality of data, barcoding, genome size estimation, analysis of k-mer spectrum and much more.

This presentation will be a short overview of k-mers, containing basic notions and applications. At first we will explain what k-mers are and we will list some areas in scientific research where k-mers appear. Then we will describe how are they applied in biology and give a few "step by step" examples. We will also talk about how to choose a proper k and how changing k affects the results of calculations. We will give some examples showing this dependence.

At the end we will discuss advantages and disadvantages of k-mers - why do they work well in some areas and in some others not sufficiently, so that they have to be improved by additional parameters.

Keywords: alignment, assembly, barcoding, DNA sequence, genome, k-mer

References

- Chor B., Horn D., Goldman N., et al. 2009. *Genomic DNA k-mer spectra: models and modalities*. *Genome Biology* 10: R108.
- Sievers A., Bosiek K., Bisch M., et al. 2017. *K-mer content, correlation, and position analysis of genome DNA sequences for the identification of function and evolutionary features*. *Genes* 8 (4): E122.

Experimental Design using Confounding Techniques: An experiment on *Aniba Rosaeodora* in the Central Amazônia

TERESA A. OLIVEIRA^{1,2}, ROBERVAL LIMA^{3,4}, AMÍLCAR OLIVEIRA^{1,2}

¹Department of Sciences and Technology
Universidade Aberta, Portugal

²Center of Statistics and Applications
University of Lisbon, Portugal

³Embrapa Amazônia Ocidental

⁴Universidade Paulista, Brazil

*teresa.oliveira@uab.pt

This work is based on a real experiment on *Aniba Rosaeodoro* in the Central Amazônia, under a reforestation program. A Factorial Block Design with Confounding was used to compare the behaviour of three different fertilizers, nitrogen, phosphorus and potassium. For the three fertilizers three different levels were considered and the experiment was located in the region of Maués-AM-Brazil. On computations we used SAS and language R, mainly the package *Agricolae*.

The results analysis indicate that the experimental technique was effective in discriminating between the fertilizers and at the same time it allowed to reduce the experimental area and the cost of deployment.

Keywords: Confounding, Experimental Design, Factorial Block Designs, Reforestation

References

- Montgomery D.C. 2001. *Design and Analysis of Experiments*. 5^o Ed., John Wiley & Sons, New York.
- Neter J., Kutner M.H., Nachtsheim C.J., Wasserman W. 1996. *Applied Linear Statistical Models*. 3rd. Ed. WCB/McGraw-Hill, New York.
- Oliveira T.A. 2000. *Planeamento de Experiências: Novas Perspectivas*, Tese de Doutorado. Universidade de Lisboa.
- R Development Core Team 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Yates F. 1937. The design and analysis of factorial experiments, *Technical Communication*, 35, Imperial Bureau of Soil Science, Harpenden.

Acknowledgements

This research is based on the partial results of the Funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through the project UID/MAT/00006/2013.

Statistical modelling and big data analytics: research and challenges

AMÍLCAR OLIVEIRA^{1,2,*}, TERESA A. OLIVEIRA^{1,2}

¹Department of Sciences and Technology
Universidade Aberta, Portugal

²Center of Statistics and Applications
University of Lisbon, Portugal

*amilcar.oliveira@ uab.pt

Big Data (BD), have potential to evaluate and enhance decision-making process and recently attracted strong interest from academics and practitioners. BD became the new paradigm in the world of information collection and analysis. BD is generally characterized by five dimensions: volume (the number of observations, attributes and relationships), variety (the diversity of data sources, formats, media and content), velocity (production speed and data change), veracity (quality, origin and reliability of data) and value. Big Data Analytics (BDA) is increasingly becoming a trendsetter that many organizations are adopting for the purpose of building valuable BD information. Disciplines such as Computer Science, Engineering and Statistics play a fundamental role in the analysis of BD, each with its own specificity, but all equally important. In this talk we will discuss the essential role of Statistical Modelling in this new world of BD and BDA, illustrating how Statistics and BD denote a crucial union. We will start with an introduction of BD and the several existing definitions and associated concepts. This presentation focus also on the research we are currently trying to implement using "Big Data" solutions in the online teaching system at Universidade Aberta, in order to monitor this system of education and to improve results.

Keywords: Big Data, Big Data Analytics, E-learning, Statistical Modelling

References

- Agarwal R., Dhar V. 2014. *Editorial – big data, data science, and analytics: the opportunity and challenge for is research*. Information Systems Research 25 (3): 443-448.
- Ali A., Qadir J., ur Rasool R., Sathiaselvan A., Zwitter A., Crowcroft J. 2016. *Big data for development: applications and techniques*. Big Data Anal. 1:2.
- Chen C.L.P., Zhang C.Y. 2014. *Data-intensive applications, challenges, techniques and technologies: a survey on big data*. Information Sciences 275: 314-347.

- Costa C., Santos M.Y. 2017. *Big Data: State-of-the-art Concepts, Techniques, Technologies, Modeling Approaches and Research Challenges*. IAENG International Journal of Computer Science 44 (3): 285-301.
- Emani C.K., Cullot N., Nicolle C. 2015. *Understandable big data: a survey*. Comput. Sci. Rev. 17: 70–81.
- Gandomi A., Haider M. 2015. *Beyond the hype: Big data concepts, methods, and analytics*. International Journal of Information Management 35 (2): 137–144.
- Kankanhalli A., Hahn J., Tan S., Gao G. 2016. *Big data and analytics in healthcare: introduction to the special section*. Inf. Syst. Front. 18: 233–235.
- Oussous A., Benjelloun F., Lahcen A., Belfkih S. 2017. *Big Data technologies: A survey*. Journal of King Saud University – Computer and Information Sciences. (In press).
- Rajaraman V. 2016. *Big data analytics*. Resonance 21: 695–716.
- Rudin C., Dunson D., Irizarry R., Ji H., Laber E., Leek J., McCormick T., Rose S., Schafer C., van der Laan M., Wasserman L., Xue L. 2014. *Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society*. A Working Group of the American Statistical Association. (www.amstat.org/asa/files/pdfs/POL-BigDataStatisticsJune2014.pdf).
- Ward J.S., Barker A. 2013. *Undefined by Data: A Survey of Big Data Definitions*. arXiv:1309.5821 [cs.DB].
- Watson H.J. 2014. *Tutorial: big data analytics: Concepts, technologies, and applications*. Communications of the Association for Information Systems 34 (1): 1247-1268.
- Zicari R.V. 2014. *Big Data: Challenges and Opportunities*. Akerkar R. (Ed.), Big data computing, CRC Press, Taylor & Francis Group, Florida, USA, pp. 103-128.

Acknowledgements

This research is based on the partial results of the Funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through the project UID/MAT/00006/2013.

Evaluation of relationships between grain yield of cereals and vegetation indices based on satellite data from MODIS

EWA PANEK, DARIUSZ GOZDOWSKI*

¹Department of Experimental Design and Bioinformatic
Warsaw University of Life Sciences, Poland

*dariusz_gozdowski@sggw.pl

Remote sensing data acquired from satellite sensors can be useful for evaluation of crop monitoring at various spatial scale (Sakamoto et al., 2014). One of the satellite sensor which is used for worldwide monitoring of crop status and yield prediction is MODIS which operates on Terra and Aqua satellites. For evaluation of crop status vegetation indices are used. One of the most important vegetation index is NDVI (normalized difference vegetation index) which is calculated on the basis of red and near infrared spectral bands (Rouse et al. 1973). For the analyses NDVI from MODIS at spatial resolution 250 m was used. Data from years 2012-2016 for period from beginning of March to the end of June at one week interval were used for the analyses. Grain yield of cereals was obtained from Central Statistical Office (GUS 2012-2016). Relationships between NDVI from subsequent measurements and grain yield of cereals were evaluated using analysis of regression on the basis of averaged data for 16 provinces (voivodeships) of Poland. The strongest relationships were observed for NDVI acquired in the end of May and in the beginning of July. Relationships for NDVI acquired in early spring or in the end of June were very weak. It means that forecasting of grain yield of cereals using satellite data for Poland is possible only for quite short period.

Keywords: analysis of regression, cereals, grain yield prediction, remote sensing, vegetation indices

References

- Rouse J.W., Haas R.H., Scheel J.A., Deering D.W. 1973. *Monitoring Vegetation Systems in the Great Plains with ERTS*. Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium 1: 48-62
- Sakamoto T., Gitelson A.A., Arkebauer T.J. 2014. *Near real-time prediction of US corn yields based on time-series MODIS data*. Remote Sensing of Environment 147: 219-231.
- GUS. 2012-2018. Local Data Bank. <https://bdl.stat.gov.pl>

Acknowledgements

This research is based on the partial results of the FERTISAT: Satellite-based Service for Variable Rate Nitrogen Application in Cereal Production – 4000118613/16/NL/EM project financed by the European Space Agency.

Characterization of capillary action of the blood in paper tips using GLMs

J. A. PEREIRA^{1*}, R. BORGES¹, L. MENDES¹, T. A. OLIVEIRA^{2,3}

¹Department of Periodontology, Dental Medicine School
University of Porto, Portugal

²Department of Sciences and Technology, Open University Portugal

³CEAUL, Center of Statistics and Applications, University of Lisbon, Portugal

*lobopereir@mail@gmail.com

The inflammation of gingiva, usually referred as gingivitis or gum disease, is a highly prevalent disease, see Stamm (1986). If left untreated, it may progress and evolve to periodontitis, a more destructive form of periodontal disease. The presence of bleeding, on periodontal probing, is an objective sign of gingivitis and indicates a high risk of periodontitis. The magnitude of bleeding has the potential to be an indicator of gingivitis severity. To assess the magnitude of bleeding we propose a measuring method based on the capillary action of blood by paper tips, used to dry the tooth canal during the endodontic treatment. An experimental study was conducted by using 200 paper tips of two different brands with calibers ranging from 50 to 80. Ten identical paper tips were randomly grouped and immersed in 3 mm deep defibrinated horse blood during 5 and 10 seconds. Each paper tip was weighed with a precision scale Kern ALJ160-4A. An authorisation of the ethic committee was obtained before the experimental procedures. The mean blood absorption of the paper tips was modeled by generalized linear models, see McCullagh and Nelder (1989) according with brand, caliber and immersion time. The homogeneity of each group of ten paper tips was assessed by the post test coefficient of variation. The results showed that the amount of absorbed blood varied with brand, caliber and time of immersion and, moreover, that the paper tip of brand A and caliber 80 was the most homogenous and the most absorbent, at the significance level considered.

Keywords: capillary action, generalized linear models, gingivitis

References

- Stamm J.W. 1986. *Epidemiology of gingivitis*. J Clin Periodontol. 13 (5): 360-366.
McCullagh P., Nelder J. A. 1989. Generalized linear models. 2nd ed., Chapman and Hall, London.

Non-pharmacological motor-cognitive treatment to improve the mental health of elderly adults

JAVIERA PONCE, VICTOR LEIVA

The objective of this paper is to propose a program of physical-cognitive dual task and to measure its impact in Chilean institutionalized elderly adults. We use an experimental design study with pre and post-intervention evaluations, measuring the cognitive and depressive levels by means of the Pfeiffer test and the Yesavage scale, respectively. The program was applied during 12 weeks to adults between 68 and 90 years old. The statistical analysis was based on the nonparametric Wilcoxon test for paired samples and was contrasted with its parametric version. The results were obtained with the statistical software R, which showed statistically significant differences in the cognitive level (p -value < 0.05) and highly significant (p -value < 0.001) in the level of depression with both tests (parametric and nonparametric). Due to the almost null evidence of scientific interventions of programs that integrate physical activity and cognitive tasks together in Chilean elderly adults, a program of physical-cognitive dual task was proposed as a non-pharmacological treatment, easy to apply and of low cost to benefit their integral health, which improves significantly the cognitive and depressive levels of institutionalized elderly adults.

Distribution of the product of two normal variables.

A state of the art.

J. ANTONIO SEIJAS-MACIAS^{1*}, AMÍLCAR OLIVEIRA², TERESA OLIVEIRA²

¹Department of Economics, Universidade da Coruña in A Coruña, Spain
& Centro de Estatística e Aplicações, Universidade Lisboa, Portugal

²Department of Science and Technology, Universidade Aberta in Lisbon, Portugal
& Centro de Estatística e Aplicações, Universidade Lisboa, Portugal

* antonio.smacias@udc.gal

This paper presents a study of the evolution the characterisation of the product of two normal distributed variables. From the first studies in the middle of the XX Century until the most recent references from this decade. Until now, existence of a unify theory of the product was not proved. Partial results present the scope the product using different approaches: Bessel Functions, Numerical Integration, Pearson Type distribution....

Keywords: Gaussian Distribution, Product, Bessel Function, Rohatgi's Theorem

References

- Aroian L. A. 1947. *The probability function of the product of two normally distributed variables*. The Annals of Mathematical Statistics 18 (2): 265-271.
- Craig C. 1936. *On the frequency of the function xy* . The Annals of Mathematical Statistics 7: 1-15.
- Cui G., Yu X., Iommelli S., Kong L. 2016. *Exact Distribution for the Product of Two Correlated Gaussian Random Variables*. IEEE Signal Processing Letters 23 (11): 1662-1666.
- Glen A., Lemmis L. M., Drew J. H. 2004. *Computing the distribution of the product of two continuous random variables*. Computational Statistics & Data Analysis 44: 451-464.
- Nadarajah S., Pogány T. K. 2016. *On the distribution of the product of correlated normal random variables*. Comptes Research de la Academie Sciences Paris, Serie I 354: 201-204.
- Seijas-Macías J. A., Oliveira A. 2012. *An approach to distribution of the product of two normal variables*. Discussiones Mathematicae. Probability and Statistics 32 (1-2): 87-99.

Acknowledgements

This research is based on the partial results of the Funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through the project UID/MAT/00006/2013.

Breeding value of selected blackcurrant (*Ribes nigrum* L.) genotypes for fruit yield and its quality

ŁUKASZ SELIGA, AGNIESZKA MASNY, STANISŁAW PLUTA

Department of Breeding of Horticultural Crops Research
Research Institute of Horticulture in Skierniewice, Poland

*Lukasz.Seliga@inhort.pl

The study was conducted at the Experimental Orchard in Dąbrowice, belonging to the Research Institute of Horticulture in Skierniewice, Poland. The aim of the research was to assess the breeding value, based on the effects of general and specific combining abilities (GCA and SCA), of 15 parental genotypes of blackcurrant in terms of fruit yield and its quality. The plant material consisted of seedlings of F₁ generation obtained by crossing, in a factorial mating design, twelve maternal ('Bona', 'Big Ben', 'Chereshnieva', 'Kupoliniai', 'Gofert', 'Tines', 'Sofievskaia', 'Tihope', 'Ores', 'Ruben', 'Titania' and D13B/11 clone) and three paternal cultivars ('Ceres', 'Foxendown' and 'Saniuta'). It was found that the cultivars, 'Ruben' and D13B/11 had significantly positive effects of general combining ability (GCA) for fruit yield. 'Saniuta' and D13B/11 had the positive GCA effects for fruit weight. For 'Chereshnieva', 'Kupoliniai', 'Gofert', 'Tines', 'Sofievskaia', 'Tihope', 'Titania', 'Ceres' and 'Saniuta' positive GCA effects were estimated for soluble solids content, whereas 'Kupoliniai', 'Ruben', 'Sofievskaia', 'Titania', 'Ceres', 'Gofert' and 'Ores' had positive GCA effects on the ascorbic acid content in fruit. The significantly positive values of specific combining ability (SCA), estimated for the following crossing combinations: 'Ruben' × 'Foxendown', 'Titania' × 'Ceres', 'Tihope' × 'Foxendown', 'Kupoliniai' × 'Saniuta', 'Gofert' × 'Foxendown', 'Big Ben' × 'Saniuta', 'Gofert' × 'Saniuta' and 'Tines' × 'Ceres' for at least two traits describing fruit yield and its quality, were evidence of the interaction of both these parental genotypes in the creation of new dessert type cultivars of blackcurrant.

Keywords: breeding value, GCA, hybrid family, SCA, trait

Acknowledgements

This work was performed in the frame of multiannual programme "Actions to improve the competitiveness and innovation in the horticultural sector with regard to quality and food safety and environmental protection", financed by the Polish Ministry of Agriculture and Rural Development.

R in the statistical analysis for life scientists

IDZI SIATKOWSKI*, JOANNA ZYPRYCH-WALCZAK

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

*idzi@up.poznan.pl

We present a book intended for all beginning programmers and statisticians who want to learn the basics of computational and graphical capabilities of the R platform used in statistics. The manuscript contains examples of scripts written in R and their implementations, mainly concerning issues from the field of life sciences. The reader has the opportunity to learn the basics of the R language and learn how to create meaningful visualizations. We present statistical issues solved with the help of R. After reading the content and following examples, the reader will be able to create high-quality figures and solve statistical problems, such as testing or regression.

Keywords: R, regression, RStudio, testing

Impact assessment of breed, sex and length of rearing on slaughter characteristics of geese

EWA SKOTARCZAK^{1*}, EWA GORNOWICZ², LIDIA LEWKO², KRZYSZTOF MOLIŃSKI¹

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

²Waterfowl Genetic Resources Station in Dworzyska
National Research Institute of Animal Production in Kraków, Poland

*ewa.skotarczak@up.poznan.pl

About 95% of the total goose meat production in Poland is exported within the product group meat and poultry products, the majority of them are the whole carcass – about 67% based mainly on White Kołuda goose®. Whereas for the Polish consumer important distinguishing features when buying poultry meat are price, carcass size, availability of separate elements as well as visual characteristics and above all fatness (Buzala et al., 2014). It seems that good breeding material for production fattening geese on the national market are regional varieties, which are currently kept in conservative flocks (Kapkowska i in., 2011; Lewko i in., 2017).

The experimental material was meat geese of three breeds (288 birds in group): White Kołuda (W-31), Pomeranian (Po) and Kielecka (Ki) in sex ratio 1:1 for 17 and 21 weeks, under technological conditions of conventional oat fattening.

A total of 19 characteristics were analysed statistically which were divided into two groups: measured and countable. The characteristics measured (g) included 11 parameters: body weight before slaughter and weight of: carcass, breast muscles, leg muscles (thigh and shank), abdominal fat, skin with subcutaneous fat, neck without skin, skeleton with muscles back, wings with skin and total muscle weight (breast and leg muscles) as well as the value of the sum of the neck, skin, skeleton and wings masses (broth elements). The countable features included 8 characteristics (%): dressing percentage – carcass weight/body weight before slaughter, meatiness – muscles in total/carcass weight ratio and weight of the abdominal fat, of the skin with subcutaneous fat, the neck without skin, the skeleton with muscles back, the wings with skin as well the value of the sum of the neck, skin, skeleton and wings masses in relations to the carcass weight.

The results were statistically analysed using the R package (R Core Team, 2013). In the first phase of analysis used Shapiro-Wilk test to verify the assumption of normality of random variables distributions describing the above features. For analysis of characteristics for which the assumption of normality was fulfilled used three-way analysis of variance, although the factors were sex, breed and length of rearing. Traits were analyzed using single-trait events model, separately for each features. The assumption of homogeneity of variances in particular groups were checked using Bratlett test and in no case there were no grounds for rejecting the hypothesis of homogeneity appropriate variances. The significance of sex, breed and term on variability of these traits, for which the hypothesis of normality was rejected, were tested by Kruskal-Wallis test and as a factor used combination of the form: sex_breed_term (12 levels). Furthermore was applied approach using so-called “aligned rank transform” available in the package ARTool for software R (Wobbrock et al., 2011).

It has been shown that the only significant interactions is sex x breed for the weight: before slaughter, abdominal fat, wings and proportion of broth elements in carcass as well as sex x term only for the last-mentioned features. The carcasses of Kielecka and White Kolumbia® geese were characterized by a significantly higher content of broth elements after a shorter period of rearing (17 weeks). W-31 geese receive significantly more favourable values for most slaughter parameters compared to conservative flocks (Ki and Po). The exception is abdominal fat and skin with subcutaneous fat content and meatiness.

Keywords: breed, goose, meatiness, rearing, sex, slaughter quality

References

- Buzala M., Adamski M., Janicki B. 2014. *Characteristics of performance traits and the quality of meat and fat in polish oat geese*. World's Poultry Science Journal 70 (3): 531-542.
- Kapkowska E., Gumulka M., Rabsztyn A., Poltowicz K., Andres K. 2011. *Comparative study on fattening results of Zatorska and White Kolumbia® geese*. Ann. Anim. Sci. 2: 207-217.
- Lewko L., Gornowicz E., Pietrzak M., Korol W. 2017. *The effect of origin, sex and feeding on sensory evaluation and some quality characteristics of goose meat from Polish native flocks*. Ann. Anim. Sci., 4:1185-1196.
- R Core Team 2013. R: A language and environment for statistical computing. R 269 Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Wobbrock J.O., Findlater L., Gergle D., Higgins J.J. 2011. *The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures*. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, pp. 143-146.

*48th International Biometrical Colloquium in Honour of the 90th Birthday of
Professor Tadeusz Caliński and VI Polish – Portuguese Workshop on Biometry,
Szamotuły, September 9-13, 2018*

Acknowledgements

This research is based on the partial results of the research project nr 03-009.2 financed by National Research Institute of Animal Production in Kraków.

Two-Piece Power Normal Distribution

PIOTR SULEWSKI^{1*}

¹Institute of Mathematics
Pomeranian University in Słupsk, Poland

*piotr.sulewski@apsl.edu.pl

This paper puts forward a new flexible distribution, which is called the two-piece power-normal distribution. Cumulative distribution function and probability density functions, quantiles, mean, variance, skewness, kurtosis, order statistics, random number generator in the paper are studied. The problem of obtaining the maximum likelihood estimates is shown. A numerical example describing skewness and kurtosis of distribution is presented. Finally, an appropriate measure to compare the flexibility of various distributions using skewness and kurtosis is proposed.

Keywords: flexibility of distribution; maximum likelihood estimator; random number generator; skewness and kurtosis; two-piece normal distribution

References

- Ardalan A., Sadooghi-Alvandi S. M., Nematollahi A. R. 2012. *The two-piece Normal-Laplace distribution*. Communications in Statistics-Theory and Methods 41 (20): 3759-3785.
- Azzalini A. 1985. *A class of distributions which includes the normal ones*. Scandinavian Journal of Statistics 12 (2): 171-178.
- Holla M. S., Bhattacharya S. K. 1968. *On a compound Gaussian distribution*. Annals of the Institute of Statistical Mathematics Tokyo 20: 331-336.
- John S. 1982. *The three-parameter two-piece normal family of distributions and its fitting*. Communications in Statistics-Theory and Methods 11 (8): 879-885.
- Jones M. C. 2015. *On families of distributions with shape parameters*. Int. Stat. Rev. 83 (2): 175-192.
- Kim H. J. 2005. *On a class of two-piece skew-normal distributions*. Stat. J. Theor. Appl. Stat. 39 (6): 537-553.
- Kimber A. C., Jeynes C. 1987. *An application of the truncated two-piece normal distribution to the measurement of depths of arsenic implants in silicon*. Appl. Stat. 36: 352-357
- Kimber A. C. 1985. *Methods for the two-piece normal distribution*. Communications in Statistics-Theory and Methods 14 (1): 235-245.
- Kumar C. S., Anusree M. R. 2014. *On some properties of a general class of two-piece skew normal distribution*. Journal of the Japan Statistical Society 44 (2): 179-194.
- Sulewski P. 2018. *The normal distribution with plasticizing component*. Communications in Statistics – Theory and Methods, under review.

Effectiveness of selection of parental components used for crossbreeding in winter wheat cultivation (*T. aestivum* ssp. *vulgare*) on the basis of the yield assessment of their offspring

TADEUSZ ŚMIAŁOWSKI*, DARIUSZ R. MAŃKOWSKI, MAŁGORZATA R. CYRAN,
WIOLETTA M. DYNKOWSKA

Plant Breeding and Acclimatization Institute, National Research Institute in Radzików,
Poland

*zhsmialo@cyf-kr.edu.pl

An analysis of the impact of the genetic origin of 1966 varieties and winter wheat strains on grain yielding was performed based on the results of field experiments located in 6 to 10 localities from 2004 to 2017. The offspring of 1136 objects obtained as a result of single crosses (AxB) and 603 objects obtained as a result of compound crossbreeds (A*B)*C or (A*B)*(C*D) were studied. On this basis, an attempt was made to assess the effectiveness of the selection of parental components: mothers and fathers in obtaining high yield potential in the offspring. Before the analysis, the yield of the objects was “corrected” by subtracting or adding from the crops of the tested varieties the deviation of the average yields of the reference varieties from the general average. It was found that the most-used parental component was the winter wheat variety - Tonacja, both as a mother (32 times) and father (56 times). Her offspring with other varieties yielded high but not record high. The highest yield in the offspring was obtained using the Legenda - 114.9 dt/ha variant of winter wheat used in 10 combinations, and the most fertile father turned out to be the Zyta winter wheat variety - 119.4 dt/ha used in 8 cross-breeding combinations. However, both "the longest components" have not been crossed. The most valuable object of winter wheat in the analyzed material turned out to be the line: STH 5295/1 (119.4 dt/ha) bred in 2014 as a result of selection of plants obtained from the simple A*B cross-breeding combination i.e. (MIB295*ZYTA). There was no significant impact on the yield of selected winter wheat crops of the type of crossings used.

References

Anoshenko B.Y. 1998. *Estimation of parental value for varieties used in plant breeding*. Plant Breeding 117: 131-137.

- Chrzastek M., Paczos-Grzęda E., Kruk K. 2007. *Efektywność krzyżowań wstecznych mieszańców międzygatunkowych owsa*. ZP. PNR. 517: 818-826.
- Witkowska K., Witkowski E. Węgrzyn S. 2002. Efekty krzyżowania pszenicy ozimej zwyczajnej (*T. aestivum* L.) z pszenicą ozimą twardą (*T durum* Desf.). *VI Międzynarodowe Sympozjum „Genetyka Ilościowa Roślin Uprawnych”*, Duszniki Zdrój. Streszczenia: 132-133.

Some Details About Pediatric High Blood Pressure

M. FILOMENA TEODORO^{1,2*}, CARLA SIMÃO^{3,4}

¹CINAV, Center of Naval Research, Portuguese Naval Academy, Base Naval de Lisboa,
Alfeite, 1910-001 Almada, Portugal

²CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico,
Lisbon University, Avenida Rovisco Pais, 1, 1048-001 Lisboa, Portugal

³Medicine Faculty, Lisbon University, Av. Professor Egas Moniz, 1600-190 Lisboa, Portugal

⁴Pediatric Department of Santa Maria's Hospital, Centro Hospitalar Lisboa Norte, Av.

Professor Egas Moniz, 1600-190 Lisboa, Portugal

*maria.alves.teodoro@marinha.pt

Taking into account the risk factors of pediatric hypertension, it is important to evaluate the pediatric arterial hypertension prevalence and the caregivers Knowledge about this disease. In [Teodoro 2017] was done a first approach implementing and analyzing statistically an experimental and simple questionnaire including 5 questions with dichotomous answers where some social demographic characteristics were related related with the hypertension literacy level. An improved questionnaire applied to children caregivers and filled online was completed later using some multivariate techniques. At present, we obtain estimates about the childhood hypertension prevalence in several regions of Portugal. As preliminary approach [Teodoro 2019], we have performed an analysis of variance. The results evidences significant differences of high blood pressure prevalence between girls and boys; also the children's age is a significant issue to take into consideration. We are still going on with the estimation of some models using generalized linear techniques and mixed models.

Keywords: analysis of variance, caregiver, general linear models, pediatric hypertension, questionnaire, statistical approach

References

- Teodoro M.F. et al. 2019 *Comparing Childhood Hypertension Prevalence in Several Regions in Portugal*. In: Machado J., Soares F., Veiga G. (eds) Innovation, Engineering and Entrepreneurship. Lecture Notes in Electrical Engineering, 505 (29): 206-213.
- Teodoro, M.F., Simão, C. 2017 *Perception about Pediatric Hypertension*. Journal of Computational and Applied Mathematics 312: 209–215.

Acknowledgements

This work was supported by Portuguese funds through the Center for Computational and Stochastic Mathematics (CEMAT), The Portuguese Foundation for Science and Technology (FCT), University of Lisbon, Portugal, project UID/Multi/04621/2013, and Center of Naval Research (CINAV), Naval Academy, Portuguese Navy, Portugal.

Estimating wood volume of forest stands using Airborne Laser Scanning data

KRZYSZTOF UKALSKI^{1*}, STANISŁAW MIŚCICKI², KAROLINA PARKITNA³,
GRZEGORZ KROK³, KRZYSZTOF STEREŃCZAK³

¹Department of Econometrics and Statistics
Warsaw University of Life Sciences, Poland

²Department of Forest Management Planning and Forest Economics
Warsaw University of Life Sciences, Poland

³Laboratory of Geomatics
Forest Research Institute in Raszyn, Poland

*krzysztof_ukalski@sggw.pl

Foresters are interested in the methods of determining the characteristics of forest stands, in particular, relevant in terms of the inventory growing stock volume as the above ground volume of living stems. In order to reduce the costs of forest inventory and increase accuracy, various methods are used, including the laser scanning technique LIDAR (Light Detection and Ranging). Airborne Laser Scanning (ALS) is used for rapid examination of large forest areas. The basic product of the ALS is a cloud of points with known spatial coordinates, which are the places of reflections of laser beams from obstacles encountered. On the basis of spatial coordinates for the separated segments, representing single tree, within a grid cell, the values of traits used for wood volume modeling are determined. In order to validate the model, ground-based growing stock volume measurements are also needed. These measurements are made at sample plots in the examined forest stand using traditional caliper and ultrasonic hypsometer to obtain tree diameter and height parameters.

The aim of the presented research is to estimate the average growing stock volume based on three traits obtained from ALS data: average maximum height of the segments in the grid cell, the sum of the maximum height of the segments in the grid cell and the sum of the segments area in the grid cell.

Combining terrestrial and ALS measurements, linear and non-linear regression models were determined using the GLIMMIX procedure in SAS. The models considering only fixed effects and models with fixed and random effects (grid cells) were taken into account. Of

these, two were selected: the first model with fixed effects with a determination coefficient of 70.12% and the second, including a random effect, with a determination coefficient of 91.86%. Both models can be used to estimate the average growing stock volume of forest stands. However, the advantage of the first model is known methodology for calculating the estimation error. The second model, despite a better fit, is less useful, because the method of determining the estimation error is unknown.

Keywords: ALS data, LIDAR, non-linear regression, wood volume

Acknowledgements

This research is based on the partial results of the REMBIOFOR project “Remote sensing based assessment of woody biomass and carbon storage in forests” supported by the National Centre for Research and Development within the program „Natural environment, agriculture and forestry” BIOSTRATEG, under the agreement no. BIOSTRATEG1/267755/4/NCBR/2015.

Assessment of *Lathyrus* species accession variability using visual and statistical methods

BOGNA ZAWIEJA^{1*}, WOJCIECH RYBIŃSKI², KAMIŁA NOWOSAD³,
JAN BOCIANOWSKI¹

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

²Institute of Plant Genetics Polish Academy of Sciences, Poznań, Poland

³Department of Genetics, Plant Breeding and Seed Production
Wrocław University of Environmental and Life Sciences, Poland

*bogna.zawieja@up.poznan.pl

Underestimated lesser-known species often prove to be a very attractive object of research. For example, they adapt very well to various marginal growing conditions, e.g. highlands, arid areas, salt-affected soils, etc. Some species of the genus *Lathyrus* may be examples of such crops. In this study the species and their accessions were compared as regards to coefficients of variation. A considerable degree of variability is important to breeders of new varieties, since the chance of obtaining new cultivars significantly different from established varieties is thus increased. Thus the coefficients of variation were compared using both visual and statistical methods, with those highest in value being of greatest interest. The highest variability of 100 seeds weight was noticed in species *L. aphaca*, *L. clymenum*, *L. hirsutus*.

Keywords: Andrews curves, coefficient of variation, linear discriminant analysis, nonlinear kernel discriminant analysis.