

Application of Multivariate Analysis Methods in Welding Engineering

Abstract: Phenomena and processes taking place during welding are usually very complex and, for this reason, should be described using multivariate methods. The article discusses the methodological basis and selected application areas as regards the solving of welding problems using statistical multivariate methods. In addition, the article presents exemplary applications of the design of experiment, multiple regression analysis, cluster analysis, principal component analysis and logistic regression analysis. The application of multivariate analyses provides the possibility of performing the mathematical description of joining processes, which, after verification, could be used to predict results of such processes, particularly in relation to the properties and fatigue service life of welded structures.

Keywords: welding, weldability of steel, design of experiment, regression analysis, cluster analysis

DOI: [10.17729/ebis.2018.5/15](https://doi.org/10.17729/ebis.2018.5/15)

Introduction

Welding and allied processes include technologies used in numerous sectors of industry. Because of their very complex nature, the above-named processes are rated among special processes and need to be monitored at every stage. Welding and allied processes are defined, among other things, in the PN-EN ISO 9001 standard. Welding processes are characterised by the so-called relevant or principal variables cited in related standards, e.g. of the PN-EN ISO 15614 series of standards concerned with welding procedure qualification. The foregoing facilitates the control of welding processes through monitoring and application

of identified ranges of variables, e.g. heat input (welding linear energy). The remaining welding processes are also governed by related standards. At the same time it should be emphasized that welding engineering incorporates interdisciplinary technologies, the monitoring which requires the knowledge of, e.g. mechanics and strength of materials, thermodynamics, electrical engineering, electronics, materials science, metallurgy, economics and management.

In terms of conditions and parameters, welding processes are complex technological and structural issues, the mathematical description of which is often possible only by means of multivariate analyses of data containing

dr hab. inż. Dariusz Fydrych (PhD (DSc) Habilitated Eng.), Professor at Gdańsk University of Technology; dr inż. Aleksandra Świerczyńska (PhD (DSc) Eng.); dr inż. Grzegorz Rogalski (PhD (DSc) Eng.); dr hab. inż. Jerzy Łabanowski (PhD (DSc) Habilitated Eng.) – Gdańsk University of Technology; Faculty of Mechanical Engineering; Institute of Materials Engineering and Welding;

information about many objects described by many variables [1]. The above-named approach makes it possible to determine the effect of input factors (independent variables) on the properties of an object and identify the relevance of these factors. In many cases, this facilitates the performance of analysis extensively describing a phenomenon subjected to investigation [2-4].

Today's requirements concerning quality systems in welding engineering (e.g. PN-EN ISO 3834-2 or PN-EN 1090-1, 2) require the collection of large amounts of data related to the preparation and course of technological processes as well as the inspection of the quality of welded joints. To satisfy such requirements, welding personnel prepare appropriate documentation, whereas modern welding devices make it possible to monitor and acquisition appropriate welding parameters. The foregoing enables the unequivocal identification of applied materials and fabricated products as well as results in an increase in production flexibility and efficiency and, consequently, a decrease in manufacturing costs. The processing of such large amounts of data poses a significant challenge and requires the use of appropriate methods and resources [3-5]. Increasingly often, the above-named issue is addressed using statistical tools, e.g. multivariate analyses including regression analysis, logistic regression, the design of experiment, Data Mining techniques (e.g. cluster analysis, factor analysis and the principal component analysis) and graphic methods. The aforesaid tools enable the predicting of the properties and fatigue service life of welded structures minimising the risk of failure and are part of a very dynamically developing trend utilising methods enabling the scheduling of inspections and preventive maintenance of objects and systems referred to as Risk Based Inspection (RBI) [6].

Multivariate methods are widely used in humanities, economics, natural sciences, exact sciences, sociology, technical sciences and

medicine [1-4]. Many multivariate methods have been implemented in statistical systems, e.g. Statistica, Statgraphics, SAS, R, SPSS and STATA [1,2,7-9].

Multivariate regression analysis

The most commonly used and widely discussed statistical technique is the multivariate regression analysis aimed to create, verify and, finally, use a model in the predicting of a dependent variable [2, 3]. The multivariate regression analysis is the expansion of simple regression to include situations, where an object is described by many independent variables. Regression analyses constitute a versatile tool enabling the creation of mathematic models having various forms, i.e. linear, non-linear (e.g. square, polynomial, power, exponential and hyperbolic [2,4,9,10]). The adjustment of a model usually involves the use of the least squares method aimed to minimise the sum of squares of differences between the experimental (actual) value of an independent variable and its value predicted by the model [2, 4, 10].

Typical applications of the regression analysis in welding engineering include, among other things, the mathematical description of the effect of welding parameters on parameters characterising the shape of the weld and the operational properties of joints or the effect of the chemical composition of filler materials and parameters characterising the welding thermal cycle on the volume fraction of structures in welds [11,12]. For instance, a commonly used model developed using the regression analysis is the carbon equivalent equation [13]. Author's own experience includes, among other things, the determination of models describing the effect of welding process (136) conditions and parameters on the amount of hydrogen diffusing in the weld deposit [14].

Logistic regression

In cases of phenomena presentable in the form of dichotomic (dual-state) values it is possible

to use logistic regression, for instance, during tests concerning resistance to the formation of various types of technological and fatigue cracks, assuming that all results of tests ending up in the rupture of a specimen can be designated as 1, whereas results of tests lasting longer than previously assumed critical time can be designated as 0 [15,16]. The use of logistic regression enables the development of the dependence having the following general form [2, 4]:

$$f(x) = \frac{e^x}{1 + e^x}$$

where

- x - independent variable of the function,
- e - base of the natural logarithm.

The above-presented equation is characterised by relatively simple graphic interpretation (sigmoidal curve) and can be used to calculate the probability of, e.g. the cracking of a joint in relation to selected test factors [15,16] or the formation of welding imperfections [17].

Design of experiment

The design of experiment constitutes a set of statistical techniques involving the setting of specific values of independent variables and the determination of values of dependent variables [18]. The use of the design of experiment in welding-related issues creates difficulties

in relation to the feasibility of test designs. It sometimes happens that associations of input parameters imposed by designs based on properly selected (in terms of technology) maximum ranges of parameters preclude the performance of an experiment. However, some reference publications present examples of tests connected with the determination of models describing the effect of welding process conditions and parameters on diffusible hydrogen content in deposited metal and the effect of process parameters on the quality of overlay welds. Preliminary tests were based on the Plackett-Burman design verifying the significance of individual independent variables on a dependent variable and the general form of a model [19-24]. In turn, tests proper were performed in accordance with two-level complete designs or fractional designs [24-26].

Cluster analysis

Complex multivariate problems can be solved using Data Mining techniques, enabling the identification of correlations within the large amount of data [4]. One of the Data Mining techniques is the cluster analysis constituting a separate branch of multivariate statistical analysis. Clustering algorithms involve the division of objects into possibly homogenous clusters. In doing so, the priority is to minimise the variability inside clusters and maximise the

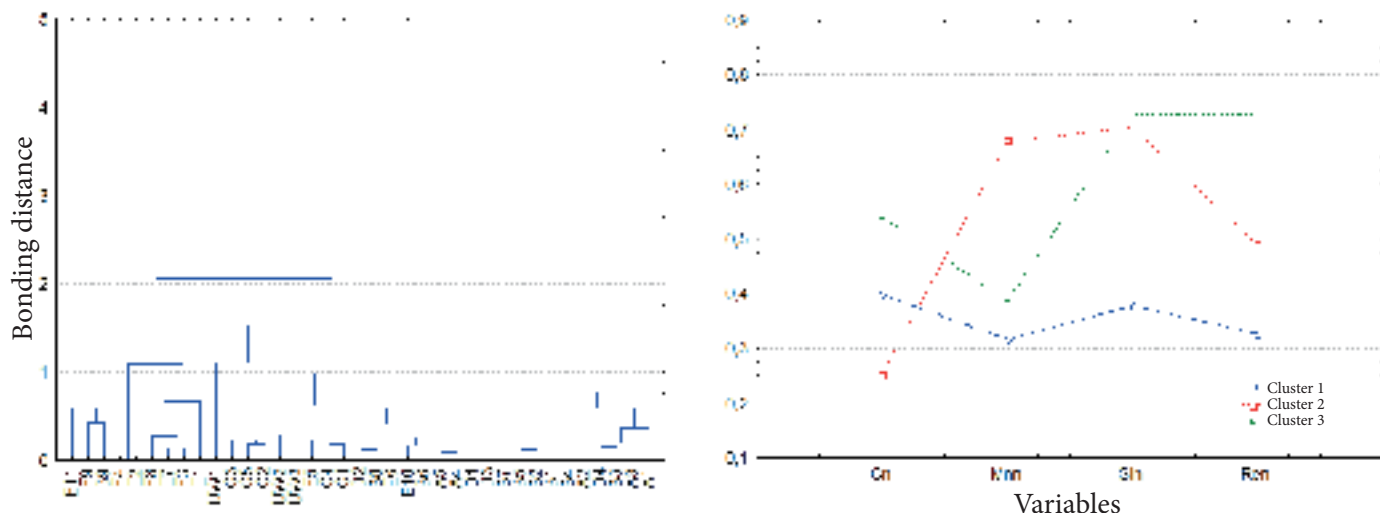


Fig. 1. Results of clustering (cluster analysis) of rutile covered electrodes: a) Euclidean distance dendrogram (Ward method), b) K-means plot [30]

variability between them. Depending on the method of the division of elements into clusters, cluster analysis techniques can be divided into hierarchical and non-hierarchical ones. In welding engineering, the cluster analysis was used when clustering MAG and TIG welding machines, resistance welding machines, rutilé electrodes, covered electrodes, fluxes for automatic submerged arc welding and welding imperfections [27-32]. In the above-named cases, good results were obtained using the hierarchical Ward method followed by the verification and supplementation of its results using the non-hierarchical k-means method. Figure 1 presents the Ward method dendrogram and the diagram of k-means in relation to the clustering of rutilé electrodes. The results of the aforesaid activity enabled the selection of electrodes (from among all analysed electrode grades) most similar in terms of the chemical composition and weld deposit properties (substitutes) to electrodes used in underwater works [30].

A factor which, to the greatest extent, reduces the wider suitability of the method is the necessity of obtaining accurate, objective, homogeneous and complete data describing welding power sources and the mastering of statistical analyses.

Dimensionality reduction methods

Popular methods used in the reduction of dimensionality include the factor analysis and the principal component analysis. The above-named methods make it possible to reduce the number of variables describing an issue under consideration. In addition, the use of these methods can lead to the determination of structure and general patterns in correlations between variables, the verification of previously detected patterns as well as the description and classification of tested objects in new spaces defined by new variables [3]. Both methods lead to the formulation of mathematical models in the form of linear equations. The factor analysis involves the transformation consisting in the replacement of correlated experimental independent variables with new non-correlated variables [3, 10]. As can be seen, the primary objective is the transformation of variables. In turn, the primary objective of the principal component analysis is the reduction of the number of variables, resulting in the creation of diagrams containing factor coordinates of cases and variables. Results of the above-presented analyses can be used as input data for further analytical tests involving the use of other multivariate methods [3, 4, 10]. In spite of the high potential usability of these methods, their use

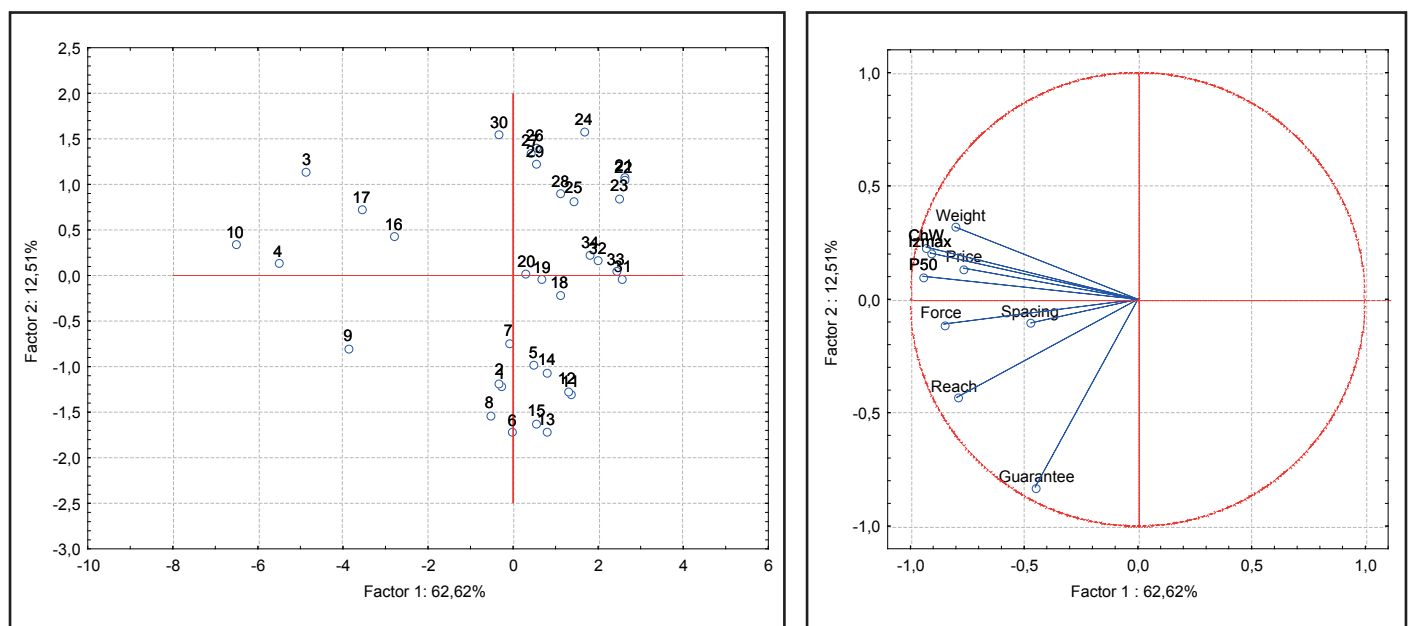


Fig. 2. Diagram of factor coordinates a) cases; b) variables

in cases of welding-related issues is very rare. Figure 2 presents the results of the principal components analysis in relation to stationary spot resistance welding machines based on data presented in publication [29]. In turn J.S. Shih used the principal component analysis to assess the quality of MIG welded joints made of aluminium foam [33].

Linear ordering

A relatively simple technique, in terms of idea and calculations, is linear ordering, consisting in the projection of multidimensional space onto a straight line. The reduction of large dimensionality is performed by calculating the mean value of standardised and simulated values of all diagnostic features [10]. The multiplication of the above-named value by 100 is followed by the obtainment of a synthetic variable, i.e. the index of standardised values, which can be used as input data to create a scatter plot. The results of the linear ordering of stationary spot resistance welding machines based in data presented in publication [29] are presented in Figure 3. The worst device, in terms of assumed diagnostic features, is the welding machine designated as D2, whereas the best device is designated as A8.

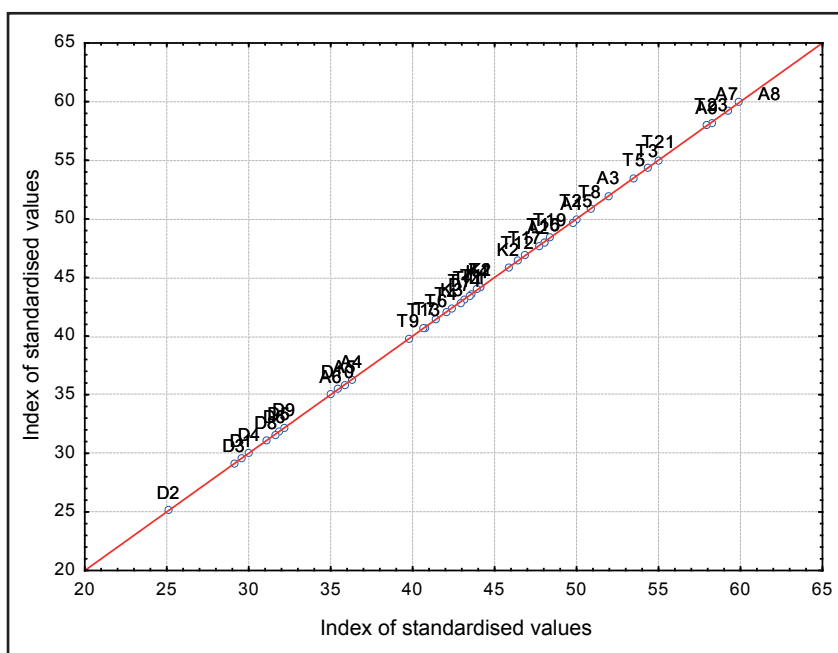


Fig. 3. Results of the linear ordering of stationary spot resistance welding machines; based on publication [29]

Graphic presentation of multivariate data

Statistical programmes also include graphic functionalities usable when presenting multivariate data, e.g. in the classification of objects. The most popular graphic presentation methods include columnar, star, radar and profile charts as well as Chernoff faces. Figure 4 presents exemplary charts created in relation to data presented in article [29], concerned with the clustering of spot welding machines. A disadvantage of the above-named technique is the necessity of performing the subjective visual assessment of similarities between graphically presented objects (which could be difficult in cases of many slightly varying objects). The Chernoff face analysis is based on a typically human capability of facial recognition. Each of independent variables (diagnostic features) is attributed to one geometrical feature of a face (e.g. face width, ear level, nose length, lip curvature).

It should be emphasized that the use of various methods can lead to different analysis results. For this reason, reference publications recommend using a few potentially suitable variants and the selection of those leading to factually justified results. The above-named situation is exemplified in Figures 2-4, where it is difficult to compare results directly.

Graphic methods are highly effective as they enable the analysis of tens of variables. In spite of this, their use in welding-related publications is extremely rare. For instance, the authors of works [34, 35] used Chernoff faces and radar diagrams to assess the quality of spot welded joints.

Summary

Because of the complex nature of welding processes and quality system monitoring-related tendency to collect increasingly large amounts

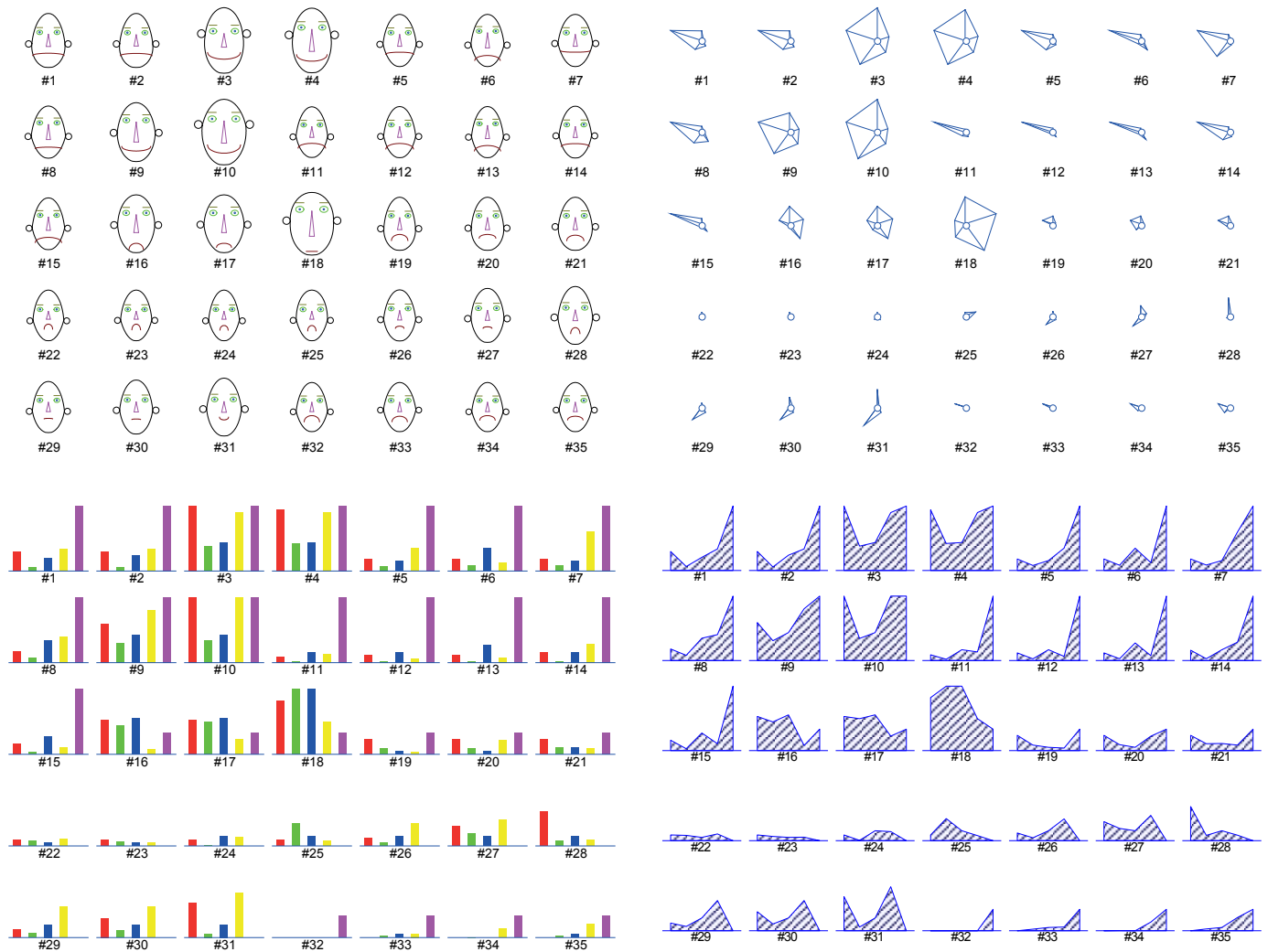


Fig. 4. Image diagrams for 35 stationary spot resistance welding machines: a) Chernoff faces, b) stars, c) columns, d) profiles; based on publication [29]

of data to be processed, it is justified to extend the methodological workshop of welding personnel and of research welding engineers by including multivariate methods. The above-named methods enable, among other things, the prediction of values of independent variables, the clustering and classification of objects, the reduction of the number of variables, the investigation of correlations between variables as well as the visualisation of results. The foregoing could be used, for instance, to verify research hypotheses, making purchase decisions and search for substitutes as well as prediction the service life of joints and welding equipment.

The availability of software programmes for statistical analysis extends the applicability of the above-presented methods, thus verifying the reliability of ongoing research works and facilitating the thorough documentation and

attractive presentation of research-related test results.

However, the aforesaid techniques are limited by the necessity of mastering the fundamentals of research methodology as well as the objective and application range of individual methods and (in some cases) the use of specialist software programmes. Luckily, in most situations analytical tests can be performed using popular spreadsheets [9].

A very important stage when performing analyses is the preparation of data. Calculations should be preceded by the selection of diagnostic features, the verification of their correctness and completeness, the selection of a method used for the replacement of missing data and the determination of variability. In cases of diagnostic features of significantly varying values it is necessary to perform the normalisation

and, if need be, the stimulation of the former. Sometimes, data are entered to a software programme in the form of the matrix of correlations or covariation. Most methods require the exclusion of collinearity, i.e. the correlation of independent variables, as it could lead to incoherent results [4]. Certain analyses (e.g. cluster analysis) are highly sensitive to the presence of outlying points [3].

Following the completion of analyses, their results should be subjected to verification. Depending on the applied technique, the verification could involve checking the agreement of the distribution of raw residuals with the normal distribution (regression model diagnostics) or checking the correct attribution of objects to clusters (cluster analysis). It is absolutely necessary to apply the principle of confronting obtained results with the researcher's knowledge and experience. The foregoing is of particular importance in relation to statistical techniques, the results of which strongly depend on the applied variant of the method.

It should be emphasized that there is a factually justified possibility of the complimentary application of various analyses enabling the more precise investigation of issues. Examples of the above-named approach include the application of various cluster analysis techniques [27-30], cluster analysis and regression analysis [1], principal component analysis and the design of experiment (Taguchi method) [33], principal component analysis and neural network analysis [36] and other variants [36,37].

The solutions presented in the article do not exhaust the above-presented issues. For instance, the study did not include the method making it possible to solve research problems described by qualitative variables. Depending on the amount of data, types of independent variables (quantitative, qualitative) and research problems (prediction, clustering, classification), welding-related problems can be solved using other multivariate methods, e.g. decision tree, correspondence analysis, discriminatory

analysis or reliability analysis [2-4, 16, 38].

References

- [1] Frątczak E., Gołata E., Klimanek T., Ptak-Chmielewska A., Pęczkowski M.: Wielowymiarowa analiza statystyczna: teoria-przykłady zastosowań z systemem SAS. Szkoła Główna Handlowa - Oficyna Wydawnicza, Warszawa, 2016.
- [2] Stanisław A.: Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 2. Modele liniowe i nieliniowe. Wydawnictwo Statsoft Polska, Kraków, 2007.
- [3] Stanisław A.: Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe. Wydawnictwo Statsoft Polska, Kraków, 2007.
- [4] Larose D.T.: Metody i modele eksploracji danych. Wydawnictwo Naukowe PWN, Warszawa, 2008.
- [5] Pei F.Q., Li D.B., Tong Y.F.: Double-layered big data analytics architecture for solar cells series welding machine. *Computers in Industry*, 2018, 97, pp. 17-23.
- [6] Dong Y., Frangopol D. M.: Incorporation of risk and updating in inspection of fatigue-sensitive details of ship structures. *International Journal of Fatigue*, 2016, 82(3), pp. 676-688.
- [7] Gatnar E., Walesiak M.: Statystyczna analiza danych z wykorzystaniem programu R. PWN, Warszawa, 2009.
- [8] Młodak A.: Analiza taksonomiczna w statystyce regionalnej. Centrum Doradztwa i Informacji Difin, Warszawa, 2006.
- [9] Rabiej M.: Analizy statystyczne z programami Statistica i Excel. Wydawnictwo Helion, Warszawa, 2017.
- [10] Kisielińska J., Stańko S.: Wielowymiarowa analiza danych w ekonomice rolnictwa. *Roczniki Nauk Rolniczych, seria G*, 2009, 96(2).
- [11] Karaoğlu S., Secgin A.: Sensitivity

- analysis of submerged arc welding process parameters. *Journal of Materials Processing Technology*, 2008, 202(1-3), pp. 500-507.
- [12] Ganjigatti J.P., Pratiha, D.K., Choudhury A.R.: Global versus cluster-wise regression analyses for prediction of bead geometry in MIG welding process. *Journal of Materials Processing Technology*, 2007, 189(1-3), pp. 352-366.
- [13] Den Uijl N. J., Nishibata H., Smith S., Okada T., Van Der Veldt T., Uchihara M., Fukui K.: Prediction of post weld hardness of advanced high strength steels for automotive application using a dedicated carbon equivalent number. *Welding in the World*, 2008, vol. 52, no. 11-12, pp. 18-29.
- [14] Świerczyńska A.: Effect of technological factors on diffusing hydrogen content in the weld deposit of rutile flux-cored wires. *Institute of Welding Bulletin*, 2013, no. 5, pp. 66-71.
- [15] Świerczyńska A.: Niszczenie wodorowe złączy spawanych ze stali ferrytyczno-austenicznych typu superdupleks. *Rozprawa doktorska, Politechnika Gdańska, Gdańsk*, 2015.
- [16] Fydrych D., Rogalski G., Świerczyńska A.: Zastosowanie próby implant do oceny spawalności stali. *Przegląd Spawalnictwa*, 2016, vol. 88, no. 8, pp. 28-32.
- [17] Gaja H., Liou F.: Defect classification of laser metal deposition using logistic regression and artificial neural networks for pattern recognition. *The International Journal of Advanced Manufacturing Technology*, 2018, 94(1-4), pp. 315-326.
- [18] Dobosz M.: Wspomagana komputerowo statystyczna analiza wyników badań. *Akademicka Oficyna Wydawnicza EXIT, Warszawa*, 2004.
- [19] Świerczyńska A., Fydrych D., Rogalski G.: Diffusible hydrogen management in underwater wet self-shielded flux cored arc welding. *International Journal of Hydrogen Energy*, 2017, 42(38), pp. 24532-24540.
- [20] Fydrych D., Świerczyńska A., Tomków J.: Diffusible hydrogen control in flux cored arc welding process. *Key Engineering Materials*, 2014, 597, pp. 171-178.
- [21] Bober M.: Badania wpływu głównych parametrów napawania plazmowego na geometrię napoin w oparciu o metody planowania eksperymentu. *Przegląd Spawalnictwa*, 2017, vol. 89, no. 4, pp. 5-9.
- [22] Bęczkowski R.: Influence of Cladding Parameters with FCAW on Bead Geometry with Plackett-Burman Experiment Used. *Archives of Metallurgy and Materials*, 2018, 63(2), pp. 681-687.
- [23] Bober M., Tobota K.: Badania istotności wpływu podstawowych parametrów napawania plazmowego na geometrię napoin. *Przegląd Spawalnictwa*, 2015, vol. 87, no. 9, pp. 24-28.
- [24] Fydrych D., Rogalski G., Świerczyńska A., Łabanowski J.: Możliwości sterowania ilością wodoru dyfundującego w złączach spawanych. *Przegląd Spawalnictwa*, 2017, t. 89, nr 10, s. 21-26.
- [25] García-García V., Reyes-Calderón F., Camacho-Arriaga J. C.: Optimization of experimental temperature measurement in GTAW process by means of DoE technique and computational modeling. *Measurement*, 2016, 88, pp. 297-309.
- [26] Fydrych D., Świerczyńska A., Rogalski G.: Effect of underwater wet welding conditions on the diffusible hydrogen content in deposited metal. *Metallurgia Italiana*, 2015, no. 11/12, pp. 47-52.
- [27] Fydrych D., Sommer S., Rogalski G.: Wspomaganie decyzji zakupu urządzeń do spawania metodą MIG/MAG z wykorzystaniem analizy skupień. *Przegląd Spawalnictwa*, 2014, no. 12, pp. 26-33.
- [28] Fydrych D., Sommer S., Rogalski G.: Analiza skupień jako metoda wspomaganie decyzji zakupu urządzeń prądu stałego

- do spawania metodą TIG. *Logistyka*, 2014, no. 6, pp. 14158-14164.
- [29] Kędziorska M., Świerczyńska A., Rogalski G., Fydrych D.: Dobór urządzeń do zgrzewania oporowego punktowego z wykorzystaniem analizy skupień. *Przegląd Spawalnictwa*, 2018, no. 5, pp. 80-85.
- [30] Fydrych D., Rogalski G., Świerczyńska A., Łabanowski J.: Ocena przydatności komercyjnych elektrod otulonych do spawania mokrego pod wodą z wykorzystaniem analizy skupień. *Przegląd Spawalnictwa*, 2015, vol. 87, no. 10, pp. 27-30.
- [31] Kujawińska A., Rogalewicz M., Diering M.: Application of expectation maximization method for purchase decision-making support in welding branch. *Management and Production Engineering Review*, 2016, 7(2), pp. 29-33.
- [32] Ji W., AbouRizk S.M., Zaïane O.R., Li Y.: Complexity analysis approach for prefabricated construction products using uncertain data clustering. *Journal of Construction Engineering and Management*, 2018, 144(8), pp. 1-11.
- [33] Shih J.S., Tzeng Y.F., Yang J.B.: Principal component analysis for multiple quality characteristics optimization of metal inert gas welding aluminum foam plate. *Materials & Design*, 2011, 32(3), pp. 1253-1261.
- [34] Zhang H., Hou Y., Zhao J., Wang L., Xi T., Li Y.: Automatic welding quality classification for the spot welding based on the Hopfield associative memory neural network and Chernoff face description of the electrode displacement signal features. *Mechanical Systems and Signal Processing*, 2017, 85, pp. 1035-1043.
- [35] Zhang H., Hou Y., Zhang J., Qi X., Wang F.: A new method for nondestructive quality evaluation of the resistance spot welding based on the radar chart method and the decision tree classifier. *The International Journal of Advanced Manufacturing Technology*, 2015, 78(5-8), pp. 841-851.
- [36] Zhang Y., Gao X., Katayama S.: Weld appearance prediction with BP neural network improved by genetic algorithm during disk laser welding. *Journal of Manufacturing Systems*, 2015, no. 34, pp. 53-59.
- [37] Vedrtnam A., Singh G., Kumar A.: Optimizing submerged arc welding using response surface methodology, regression analysis, and genetic algorithm. *Defence Technology*, 2018, 14(3), pp. 204-212.
- [38] Kim K.Y., Ahmed F.: Semantic weldability prediction with RSW quality dataset and knowledge construction. *Advanced Engineering Informatics*, 2018, no. 38, pp. 41-53.