

Original Research

Big Data Analytics of a Waste Recycling Simulation Logistics System

**Martin Straka^{1*}, Marcela Taušová², Andrea Rosová¹, Michal Cehlár²,
Peter Kačmár¹, Martin Sisol², Peter Ignáč¹, Csaba Farkas¹**

¹Institute of Logistics and Transport, BERG Faculty, Technical University of Kosice, Kosice, Slovak Republic

²Institute of Earth Resources, BERG Faculty, Technical University of Kosice, Kosice, Slovak Republic

Received: 9 February 2019

Accepted: 22 April 2019

Abstract

Our paper is focused on data evaluation about the full recycling of waste by special statistical software and by using the principles of logistics. The paper goes further than the paper entitled “Environmental assessment of waste recycling based on principles of logistics and computer simulation design,” which outputs a number of data that need to be reviewed and evaluated separately. Data, representing 15 types of waste for 5 years, enter the analysis. There were the types of waste that make up the most important part of the total waste production by means of descriptive statistics. Thanks to this, they were identified as the most important (from the production point of view) plastic granules with an average of 755.05 t/month, glass with an average of 672.233 t/month and paper with the average of 645.25 t/month. The persistence of particular waste type generation was examined by the variation coefficient in order to reduce the risk of supply of these secondary raw materials in the downstream supply chain. Selected waste elements can be considered relatively stable with a variation coefficient in the range 2.4-4.1%; the least stable type is electronic dust with a coefficient of variation of up to almost 23%.

Keywords: evaluation, waste recycling, statistical analysis software, environment, logistics

Introduction

Application of the current logistics options for the overall waste sorting gives an assumption of reducing the overall environmental load. On the basis of the information about the operation of such systems, it is possible to create accurate computer simulation models [1] where the outcomes of these models are several numbers of other data and information that can be

used to backward investigate systems with the aim to improving our environment [2]. But the problem is that the amount of data collected represents so much information that it is problematic to evaluate effectively in a standard way. For this reason, a statistical analysis program is used to further understand the operation of the overall waste separation system and to analyse data.

The aim of this paper is to point to an effective analysis and evaluation of the data obtained from a system that deals with the complete classification of waste and to understand the behaviour of such a system in order to make it more efficient and to determine the possibilities for further development of the system.

*e-mail: martin.straka@tuke.sk

Theoretical Base

Logistics Approach and Computer Statistical Analysis of Data

Using classical spreadsheet editors to analyse acquired data is popular, but has significant limitations that affect the quality of the data analysis itself. Quality is related to the possibilities of table spreadsheet editors as well as to the skills and knowledge of spreadsheet editor users. For these reasons, the professional statistical program has been used to comprehensively analyse data of complete waste separation, and some outputs will be shown in the other chapters of the paper.

Computer statistical analysis of data (CSAD) allows for significant progress in data analysis obtained for a wide research area. This also helps professionals who do not specifically deal with statistical evaluation and analysis. The use of such program resources is widespread. Authors Tang and Zhang [3] use a comprehensive but simple-to-use software package called data processing system (DPS), which has been developed to execute a range of standard numerical analyses and operations used in experimental design, statistics and data mining. DPS software provides experimenters with the scientific and statistical procedures needed to maximize the knowledge gained from research data. These procedures include ANOVA on sums of squares for balanced data and a GLM approach to analyse any type of experimental design, including unbalanced designs and experiments with missing values. With DPS it is possible to fit statistical models containing factors whether the data are experimental or observational [3]. It is also important to use CSAD for the analysis and evaluation of logistics systems and their environmental aspects. Such an application is included in an analytical and statistical JMP program.

According to the authors Jones and Sall [4], JMP is a statistical software environment that enables scientists, engineers, and business analysts to make discoveries through data exploration. One powerful method for beginning the process of discovery employs statistically designed experiments. A well-designed experiment ensures that the resulting data have large information content. We support this method with custom design, an innovative approach to the statistical design of experiments. But whether your results come from designed experiments or from an observational study, we provide analytical tools that put graphs up front. JMP's graphical user interface (GUI) makes these plots interactive and dynamically linked to each other and to the data [4]. According to the author Smith et al. [5], the use of different statistical techniques within the component parts of an ecosystem services assessment framework are discussed, including data availability and sampling strategies, statistical data analysis, geography and spatial models, meta-analysis, environmental models, societal models, feedbacks and

loop analysis, and graphical models including Bayesian belief networks. Statistics has an underpinning role by providing tools to link together the component elements along with their uncertainties for a thorough ecosystem services assessment, and should be an integral part of this developing inter-disciplinary research area [5]. According to the author Wang Y. et al. [6] and Bohács et al. [7], a data quality parameter is a qualitative or subjective dimension by which a user evaluates data quality. Source credibility and timeliness are examples. The value is directly or indirectly based on underlying quality indicator values. User defined functions may be used to map quality indicator values to quality parameter values. Author Wang G. et al. [8] wrote that the amount of data produced is significantly increasing, thereby creating challenges for the organizations that would like to reap the benefits from analysing this massive influx of big data. This is because big data can provide unique insights into, inter alia, market trends, customer buying patterns, and maintenance cycles, as well as into ways of lowering costs and enabling more targeted business decisions. Big data analytics (BDA) in logistics and supply chain management (LSCM) has received increasing attention because of its complexity and the prominent role of LSCM in improving overall business performance. LSCM faces the most significant challenges that can potentially result in inefficiencies and wastage in supply chains, such as delayed shipments, rising fuel costs, inconsistent suppliers, and ever-increasing customer expectations, among others [9]. Companies highly expect to capitalize on BDA in logistics and supply chain operations to improve the visibility, flexibility, and integration of global supply chains and logistics processes, effectively manage demand volatility, and handle cost fluctuations [10]. In the strategic phase of supply chain planning, BDA plays a vital role. It has been applied to help companies make strategic decisions on sourcing and supply chain network design, as well as on product design and development. In the operational planning phase, BDBA has been used to assist management in making supply chain operation decisions, which often include demand planning, procurement, production, inventory, and logistics. According to the authors Little and Rubin [11], most statistical software packages allow for the identification of nonrespondents by creating one or more special codes for those entries of the data matrix that are not observed. More than one code might be used to identify particular types of nonresponse, such as "don't know" or "refuse to answer" or "out of legitimate range." Some statistical packages typically exclude units that have missing value codes for any of the variables involved in an analysis. This strategy is generally inappropriate, since the investigator is usually interested in making inferences about the entire target population rather than the portion of the target population that would provide responses on all relevant variables in the analysis. Straka et al. [12] write that logistics systems consist of a finite number of active elements that create

the networks in which the logistics flows occur. These systems have a stochastic and probabilistic character. Before the design of the simulation model, it is necessary to determine whether it is possible to design the model as one system or a collection of several individual parts. Many types of simulation software allow for the design of models according to a hierarchical structure in which the whole large-scale logistics system is represented by one hierarchical block. This gives the possibility of simplifying the entire examined logistics system (including a complete waste recycling system) for simulation and modelling purposes, as well as making it possible to obtain valuable data that is suitable for detailed software analysis.

Possibilities and Outputs of Statistical Software Analysis

The rapid development of information technologies and the considerable increase of their performance gives space for the development and application of specialized software tools, and for the needs of expert activities in the field of statistics. According to Beller et al. and Cheba et al. [13, 14], the use of automatic statistical analysis has been a software engineering best practice for decades. However, we still do not know a lot about its use in real-world software projects. Many authors [15-18] write that very often, models are only valid for one particular point in space. This type of model describes a process such as nitrogen leaching in a detailed way for this single point. For such a model, many parameters must be determined (measured or estimated) because the model reacts very sensitively to

small changes in these parameters, in landscape studies. Most elementary general purpose data analysis makes the fundamental assumption that the observational units analysed represent independent pieces of evidence about the relationship under study [19, 20]. According to Bazeley and Jackson [21], there is a widely held perception that use of a computer helps to ensure rigour in the analysis process. Insofar as computer software will find and include a query procedure, for example, every recorded use of a term or every coded instance of a concept, it ensures a more complete set of data for interpretation than might occur when working manually. There are procedures that can be used, too, to check for completeness, and use of a computer makes it possible to test for negative cases (where concepts are not related). Perhaps using a computer simply ensures that the user is working more methodically, more thoroughly, and more attentively. In this sense, then, it can be claimed that the use of a computer for qualitative analysis can contribute to a more rigorous analysis. Computer simulation is also based on the same principles and its quality depends on the qualitative processing of input data and the ability of authors to create a formalized scheme with its parameters for the creation of a certain computer simulation model [22-26].

The need for continuity of production is gaining ground for activity of individual departments to ensure an undisturbed operation of the company at a minimum cost. The increase in the production capacity is one of the possibilities for achieving a higher profit by fixed cost reduction. The process simulation models are very efficient tools for detecting the bottlenecks in the process course and for improving the process parameters. To

Table 1. Initial analysis of numerical variables: mean, standard deviation, minimum, maximum, sum within 5 years per group of waste.

| Type of waste | N | DF | Mean | Std Dev | Sum | Minimum | Maximum | Coefficient of variation [%] |
|---------------------------|----|----|---------|---------|----------|---------|---------|------------------------------|
| Waste t/month | 60 | 59 | 6707,21 | 48,8033 | 402432 | 6529 | 6942 | 0,73% |
| Plastics granules t/month | 60 | 59 | 755,05 | 26,2778 | 45303 | 696 | 823 | 3,48% |
| Biogas m3/month | 60 | 59 | 352383 | 7001,3 | 2,11E+07 | 341999 | 369000 | 1,99% |
| Digestate m3/month | 60 | 59 | 3523,83 | 95,2455 | 211430 | 3210 | 3900 | 2,70% |
| Paper t/month | 60 | 59 | 645,25 | 15,7605 | 38715 | 607 | 678 | 2,44% |
| Glass t/month | 60 | 59 | 672,233 | 26,9327 | 40334 | 601 | 724 | 4,01% |
| Textile t/month | 60 | 59 | 271 | 17,0145 | 16260 | 219 | 302 | 6,28% |
| Iron t/month | 60 | 59 | 293,633 | 15,1869 | 17618 | 256 | 331 | 5,17% |
| Nonferrous t/month | 60 | 59 | 124,8 | 11,2216 | 7488 | 103 | 152 | 8,99% |
| Wood pelets t/month | 60 | 59 | 230,967 | 16,2366 | 13858 | 184 | 274 | 7,03% |
| Rubber granules t/month | 60 | 59 | 100,433 | 9,3235 | 6026 | 81 | 126 | 9,28% |
| Dangerous waste t/month | 60 | 59 | 67,65 | 8,4829 | 4059 | 48 | 91 | 12,54% |
| Electronic dust t/month | 60 | 59 | 20,8833 | 4,7838 | 1253 | 12 | 33 | 22,91% |
| Gravel t/month | 60 | 59 | 184,717 | 13,4934 | 11083 | 155 | 217 | 7,30% |
| Sand t/month | 60 | 59 | 77,9167 | 9,6679 | 4675 | 53 | 100 | 12,41% |
| Nonrecyclable t/month | 60 | 59 | 268,2 | 14,9947 | 16092 | 239 | 318 | 5,59% |

develop the proper simulation model, both theoretical knowledge (technique of simulation, specific simulation systems) and practical experience (description of the system, its elements and their mutual interactions and links) are necessary. The course of the simulation is to be monitored in every phase. It is possible to determine impacts on total function of the system from the changes that occur at the output of the simulation model [27].

From everything mentioned above, the specialized software can be considered as an important element in the solution of a project, scientific research and theoretical tasks for the needs of practice and academic environment.

Results and Discussion

Initial Analysis of Numerical Variables

The high value of the coefficient of variation indicates that the type of waste reaches significantly different values during the months of the year so that there may be a seasonal impact. The low value of the variation coefficient indicates that the type of waste reaches relatively stable values for individual months of the year, which can be positive when estimating the production of recycled material components.

Interesting may be the type of waste that has the highest and lowest average values and low or high variability (standard deviation) (Table 1). The results of

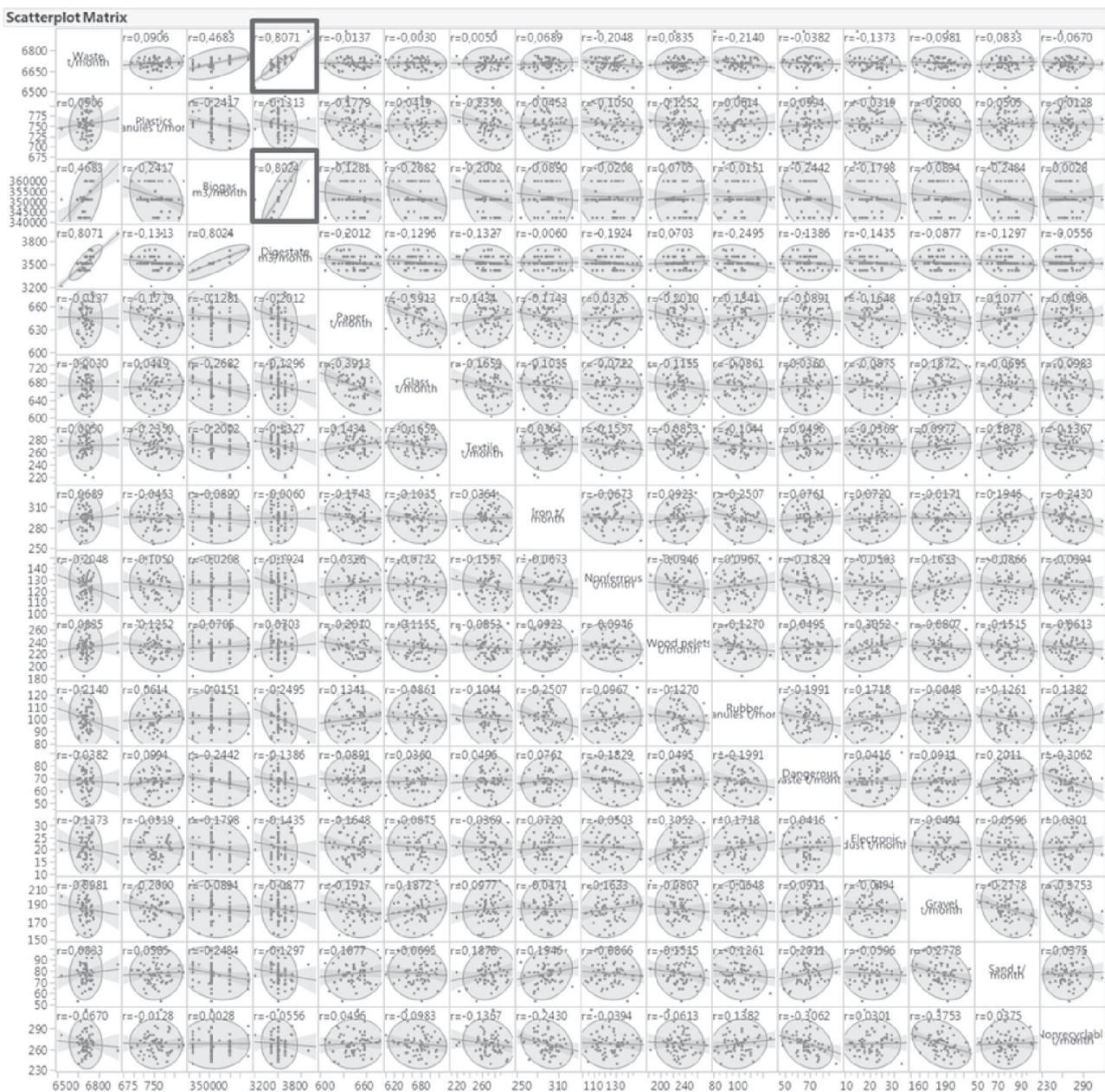


Fig. 1. Multidimensional analysis of all numerical data.

Table 2. Pair correlation analysis evaluates the relationship among all pairs of numerical variables.

| Pairwise Correlations | | | | | | | | | | | | | | | |
|-------------------------|-------------------------|-------------|-------|-----------|-----------|-------------|------|------|------|------|---|-----|-----|-----|-----|
| Variable | by Variable | Correlation | Count | Lower 95% | Upper 95% | Signif Prob | -0,8 | -0,6 | -0,4 | -0,2 | 0 | 0,2 | 0,4 | 0,6 | 0,8 |
| Glass t/month | Biogas m3/month | -0,2682 | 60 | -0,4888 | -0,0153 | 0,0383* | | | | | | | | | |
| Sand t/month | Gravel t/month | -0,2778 | 60 | -0,4967 | -0,0257 | 0,0316* | | | | | | | | | |
| Electronic dust t/month | Wood pelets t/month | 0,3052 | 60 | 0,0556 | 0,5189 | 0,0177* | | | | | | | | | |
| Nonrecyclable t/month | Dangerous waste t/month | -0,3062 | 60 | -0,5197 | -0,0567 | 0,0173* | | | | | | | | | |
| Nonrecyclable t/month | Gravel t/month | -0,3753 | 60 | -0,5744 | -0,1341 | 0,0031* | | | | | | | | | |
| Glass t/month | Paper t/month | -0,3913 | 60 | -0,5869 | -0,1526 | 0,0020* | | | | | | | | | |
| Biogas m3/month | Waste t/month | 0,4683 | 60 | 0,2433 | 0,6454 | 0,0002* | | | | | | | | | |
| Digestate m3/month | Biogas m3/month | 0,8024 | 60 | 0,6888 | 0,8775 | <,0001* | | | | | | | | | |
| Digestate m3/month | Waste t/month | 0,8071 | 60 | 0,6958 | 0,8806 | <,0001* | | | | | | | | | |

the analysis ranked plastics granules with an average of 755.05 t/month as being among the most productive types of waste, followed by glass with the average of 672.233 t/month and paper with the average of 645.25 t/month. These types of waste can be considered relatively stable, from the point of their production, with a coefficient of variation in the range 2.4-4.1%.

The multidimensional analysis of all numerical data points to a significant positive correlation between the month indicators of WASTE/DIGESTATE with the correlation coefficient $r = + 0.8071$ and BIOGAS/DIGESTATE with the correlation coefficient $r = + 0.8024$ (Fig. 1).

The relationship among all pairs of numerical variables was judged through pair correlation analysis. The results of the analysis point to pairs where the correlation is statistically significant. In addition to the aforementioned types of waste in the previous analysis, other pair indicators are included (Table 2):

- GLASS/BIOGAS correlation NEGATIVE
- SAND/GRAVEL correlation NEGATIVE
- NONRECYCLABLET/DANGEROUS WASTE correlation NEGATIVE

- NONRECYCLABLET/GRAVEL correlation NEGATIVE
- GLASS/PAPER correlation NEGATIVE
- ELECTRONIC DUST/WOOD PELETS correlation NEGATIVE
- BIOGAS/WASTE correlation NEGATIVE
- WASTE/DIGESTATE correlation NEGATIVE
- BIOGAS/DIGESTATE correlation NEGATIVE

There emerged much higher dependence among variables by repeating the multidimensional analysis of the dependence among the types of waste and the filtering according to the seasons: spring / summer / autumn / winter. The impact of seasonality is evident for some types of waste, which is worth further research for reducing the risk of using individual components of waste as a secondary raw material (Table 3).

Regression and Cluster Analysis of Selected Types of Sorted Waste

Although regression analysis was performed in all pairs (Fig. 2), the result is a statistically significant

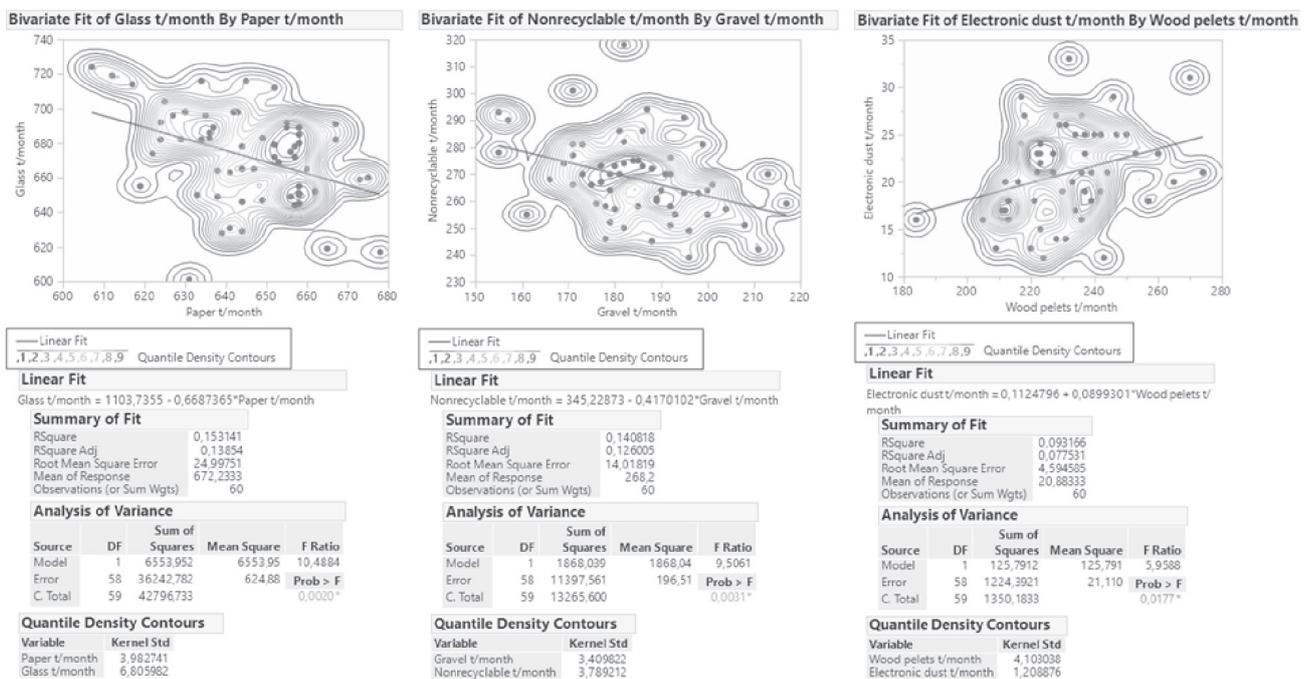


Fig. 2. Regression analysis of selected pairs of sorted waste.

Table 3. Pair correlation analysis evaluates the relationship among all pairs of numerical variables by season.

| ▼ Multivariate season=spring | | | | | | |
|------------------------------|---------------------------|-------------|-------|-----------|-----------|-------------|
| ▲ Pairwise Correlations | | | | | | |
| Variable | by Variable | Correlation | Count | Lower 95% | Upper 95% | Signif Prob |
| Textile t/month | Paper t/month | -0,4281 | 15 | -0,7715 | 0,1077 | 0,1114 |
| Paper t/month | Digestate m3/month | -0,4327 | 15 | -0,7735 | 0,1022 | 0,1072 |
| Paper t/month | Biogas m3/month | -0,4327 | 15 | -0,7735 | 0,1022 | 0,1072 |
| Nonrecyclable t/month | Plastics granules t/month | -0,4337 | 15 | -0,7740 | 0,1010 | 0,1063 |
| Electronic dust t/month | Glass t/month | -0,4428 | 15 | -0,7785 | 0,0899 | 0,0984 |
| Gravel t/month | Dangerous waste t/month | 0,4603 | 15 | -0,0680 | 0,7870 | 0,0842 |
| Nonrecyclable t/month | Dangerous waste t/month | -0,4778 | 15 | -0,7954 | 0,0456 | 0,0717 |
| Biogas m3/month | Waste t/month | 0,4886 | 15 | -0,0316 | 0,8005 | 0,0646 |
| Digestate m3/month | Waste t/month | 0,4886 | 15 | -0,0316 | 0,8005 | 0,0646 |
| Sand t/month | Electronic dust t/month | -0,4911 | 15 | -0,8017 | 0,0283 | 0,0630 |
| Gravel t/month | Wood pelets t/month | -0,5408 | 15 | -0,8246 | -0,0395 | 0,0374* |
| Wood pelets t/month | Glass t/month | -0,5448 | 15 | -0,8264 | -0,0451 | 0,0357* |
| Electronic dust t/month | Wood pelets t/month | 0,5466 | 15 | 0,0476 | 0,8272 | 0,0350* |
| Nonrecyclable t/month | Nonferrous t/month | 0,5777 | 15 | 0,0930 | 0,8411 | 0,0241* |
| Sand t/month | Rubber granules t/month | -0,6310 | 15 | -0,8640 | -0,1755 | 0,0117* |
| Digestate m3/month | Biogas m3/month | 1,0000 | 15 | 1,0000 | 1,0000 | <,0001* |

| ▼ Multivariate season=summer | | | | | | |
|------------------------------|-------------------------|-------------|-------|-----------|-----------|-------------|
| ▲ Pairwise Correlations | | | | | | |
| Variable | by Variable | Correlation | Count | Lower 95% | Upper 95% | Signif Prob |
| Nonferrous t/month | Glass t/month | -0,4533 | 15 | -0,7836 | 0,0768 | 0,0897 |
| Dangerous waste t/month | Nonferrous t/month | -0,4549 | 15 | -0,7844 | 0,0748 | 0,0884 |
| Electronic dust t/month | Iron t/month | -0,4630 | 15 | -0,7883 | 0,0646 | 0,0822 |
| Nonrecyclable t/month | Textile t/month | -0,4714 | 15 | -0,7923 | 0,0539 | 0,0761 |
| Paper t/month | Waste t/month | -0,4806 | 15 | -0,7967 | 0,0420 | 0,0698 |
| Paper t/month | Digestate m3/month | -0,5035 | 15 | -0,8075 | 0,0118 | 0,0557 |
| Rubber granules t/month | Digestate m3/month | -0,5092 | 15 | -0,8101 | 0,0042 | 0,0526 |
| Nonrecyclable t/month | Rubber granules t/month | 0,5163 | 15 | 0,0055 | 0,8134 | 0,0488* |
| Sand t/month | Gravel t/month | -0,5594 | 15 | -0,8330 | -0,0661 | 0,0301* |
| Rubber granules t/month | Waste t/month | -0,5634 | 15 | -0,8347 | -0,0718 | 0,0288* |
| Textile t/month | Biogas m3/month | -0,5846 | 15 | -0,8441 | -0,1032 | 0,0221* |
| Digestate m3/month | Biogas m3/month | 0,6187 | 15 | 0,1558 | 0,8588 | 0,0139* |
| Digestate m3/month | Waste t/month | 0,9468 | 15 | 0,8438 | 0,9825 | <,0001* |

| ▼ Multivariate season=autumn | | | | | | |
|------------------------------|---------------------------|-------------|-------|-----------|-----------|-------------|
| ▲ Pairwise Correlations | | | | | | |
| Variable | by Variable | Correlation | Count | Lower 95% | Upper 95% | Signif Prob |
| Dangerous waste t/month | Plastics granules t/month | -0,4610 | 15 | -0,7873 | 0,0671 | 0,0837 |
| Sand t/month | Digestate m3/month | -0,4630 | 15 | -0,7883 | 0,0646 | 0,0822 |
| Nonferrous t/month | Digestate m3/month | -0,4645 | 15 | -0,7890 | 0,0626 | 0,0811 |
| Wood pelets t/month | Paper t/month | -0,4673 | 15 | -0,7904 | 0,0591 | 0,0790 |
| Sand t/month | Biogas m3/month | -0,4677 | 15 | -0,7906 | 0,0586 | 0,0787 |
| Gravel t/month | Glass t/month | 0,4769 | 15 | -0,0468 | 0,7949 | 0,0723 |
| Sand t/month | Textile t/month | 0,4785 | 15 | -0,0447 | 0,7957 | 0,0712 |
| Glass t/month | Biogas m3/month | -0,4931 | 15 | -0,8026 | 0,0257 | 0,0618 |
| Rubber granules t/month | Paper t/month | 0,5415 | 15 | 0,0404 | 0,8249 | 0,0371* |
| Glass t/month | Waste t/month | -0,5513 | 15 | -0,8293 | -0,0544 | 0,0332* |
| Electronic dust t/month | Paper t/month | -0,5628 | 15 | -0,8345 | -0,0710 | 0,0290* |
| Rubber granules t/month | Glass t/month | -0,6098 | 15 | -0,8550 | -0,1418 | 0,0158* |
| Gravel t/month | Paper t/month | -0,6791 | 15 | -0,8839 | -0,2559 | 0,0054* |
| Electronic dust t/month | Wood pelets t/month | 0,6940 | 15 | 0,2820 | 0,8899 | 0,0041* |
| Glass t/month | Paper t/month | -0,7050 | 15 | -0,8943 | -0,3017 | 0,0033* |
| Biogas m3/month | Waste t/month | 0,7135 | 15 | 0,3171 | 0,8977 | 0,0028* |
| Digestate m3/month | Waste t/month | 0,7179 | 15 | 0,3253 | 0,8994 | 0,0026* |
| Digestate m3/month | Biogas m3/month | 0,9659 | 15 | 0,8980 | 0,9889 | <,0001* |

| ▼ Multivariate season=winter | | | | | | |
|------------------------------|---------------------------|-------------|-------|-----------|-----------|-------------|
| ▲ Pairwise Correlations | | | | | | |
| Variable | by Variable | Correlation | Count | Lower 95% | Upper 95% | Signif Prob |
| Rubber granules t/month | Textile t/month | -0,4007 | 15 | -0,7373 | 0,1405 | 0,1588 |
| Iron t/month | Glass t/month | -0,4023 | 15 | -0,7583 | 0,1385 | 0,1371 |
| Sand t/month | Electronic dust t/month | 0,4036 | 15 | -0,1370 | 0,7589 | 0,1358 |
| Rubber granules t/month | Iron t/month | -0,4063 | 15 | -0,7603 | 0,1338 | 0,1329 |
| Iron t/month | Digestate m3/month | -0,4090 | 15 | -0,7617 | 0,1306 | 0,1301 |
| Paper t/month | Plastics granules t/month | -0,4112 | 15 | -0,7628 | 0,1280 | 0,1278 |
| Nonferrous t/month | Plastics granules t/month | 0,4191 | 15 | -0,1186 | 0,7668 | 0,1199 |
| Nonrecyclable t/month | Iron t/month | -0,4286 | 15 | -0,7715 | 0,1073 | 0,1110 |
| Dangerous waste t/month | Biogas m3/month | -0,4602 | 15 | -0,7869 | 0,0681 | 0,0843 |
| Nonrecyclable t/month | Paper t/month | 0,4622 | 15 | -0,0656 | 0,7879 | 0,0828 |
| Electronic dust t/month | Waste t/month | -0,4696 | 15 | -0,7915 | 0,0562 | 0,0774 |
| Dangerous waste t/month | Digestate m3/month | -0,4727 | 15 | -0,7929 | 0,0522 | 0,0752 |
| Nonferrous t/month | Textile t/month | -0,4981 | 15 | -0,8050 | 0,0190 | 0,0588 |
| Dangerous waste t/month | Paper t/month | -0,5049 | 15 | -0,8081 | 0,0100 | 0,0549 |
| Electronic dust t/month | Biogas m3/month | -0,5930 | 15 | -0,8477 | -0,1160 | 0,0198* |
| Electronic dust t/month | Digestate m3/month | -0,5972 | 15 | -0,8496 | -0,1224 | 0,0187* |
| Glass t/month | Paper t/month | -0,6036 | 15 | -0,8523 | -0,1322 | 0,0172* |
| Nonrecyclable t/month | Gravel t/month | -0,6097 | 15 | -0,8549 | -0,1416 | 0,0158* |
| Textile t/month | Plastics granules t/month | -0,6112 | 15 | -0,8556 | -0,1440 | 0,0155* |
| Digestate m3/month | Waste t/month | 0,6873 | 15 | 0,2701 | 0,8872 | 0,0046* |
| Biogas m3/month | Waste t/month | 0,6917 | 15 | 0,2779 | 0,8890 | 0,0043* |
| Electronic dust t/month | Dangerous waste t/month | 0,7042 | 15 | 0,3002 | 0,8940 | 0,0034* |
| Digestate m3/month | Biogas m3/month | 0,9986 | 15 | 0,9956 | 0,9995 | <,0001* |

Table 4. Variable clustering method.

| Cluster Summary | | | | |
|-----------------|-------------------|------------------------------|---|---|
| Cluster | Number of Members | Most Representative Variable | Cluster Proportion of Variation Explained | Total Proportion of Variation Explained |
| 1 | 4 | Digestate m3/month | 0,611 | 0,153 |
| 3 | 2 | Glass t/month | 0,696 | 0,087 |
| 5 | 2 | Nonrecyclable t/month | 0,688 | 0,086 |
| 2 | 3 | Rubber granules t/month | 0,442 | 0,083 |
| 4 | 2 | Plastics granules t/month | 0,617 | 0,077 |
| 7 | 2 | Nonferrous t/month | 0,591 | 0,074 |
| 6 | 1 | Sand t/month | 1 | 0,063 |

Proportion of variation explained by clustering: 0,622

linear regression model in all cases. The descriptive capability of the models is in the range of (7.1; 65)%.

We examined the common features of each type of waste by cluster analysis, and these clusters were formed on the basis of similar behaviour. The principle of cluster analysis is to group variables so that maximum homogeneity of variables is reached within the cluster and maximum variability among clusters. Variable clustering provides a method of grouping similar variables into representative groups (Table 4). Each cluster can be represented by one component or

a variable. Component is a linear combination of all variables in a cluster. Alternatively, a cluster can be represented by an identified variable that is the most representative of a cluster. Cluster variables can be used as a reduction method. Instead of using a large set of variables in modelling, either the cluster components or the most representative cluster variable can be used to explain most variations in the data.

The hierarchical clustering method begins with each observation of its own clustering. In each step, the clustering process calculates the distance among all the

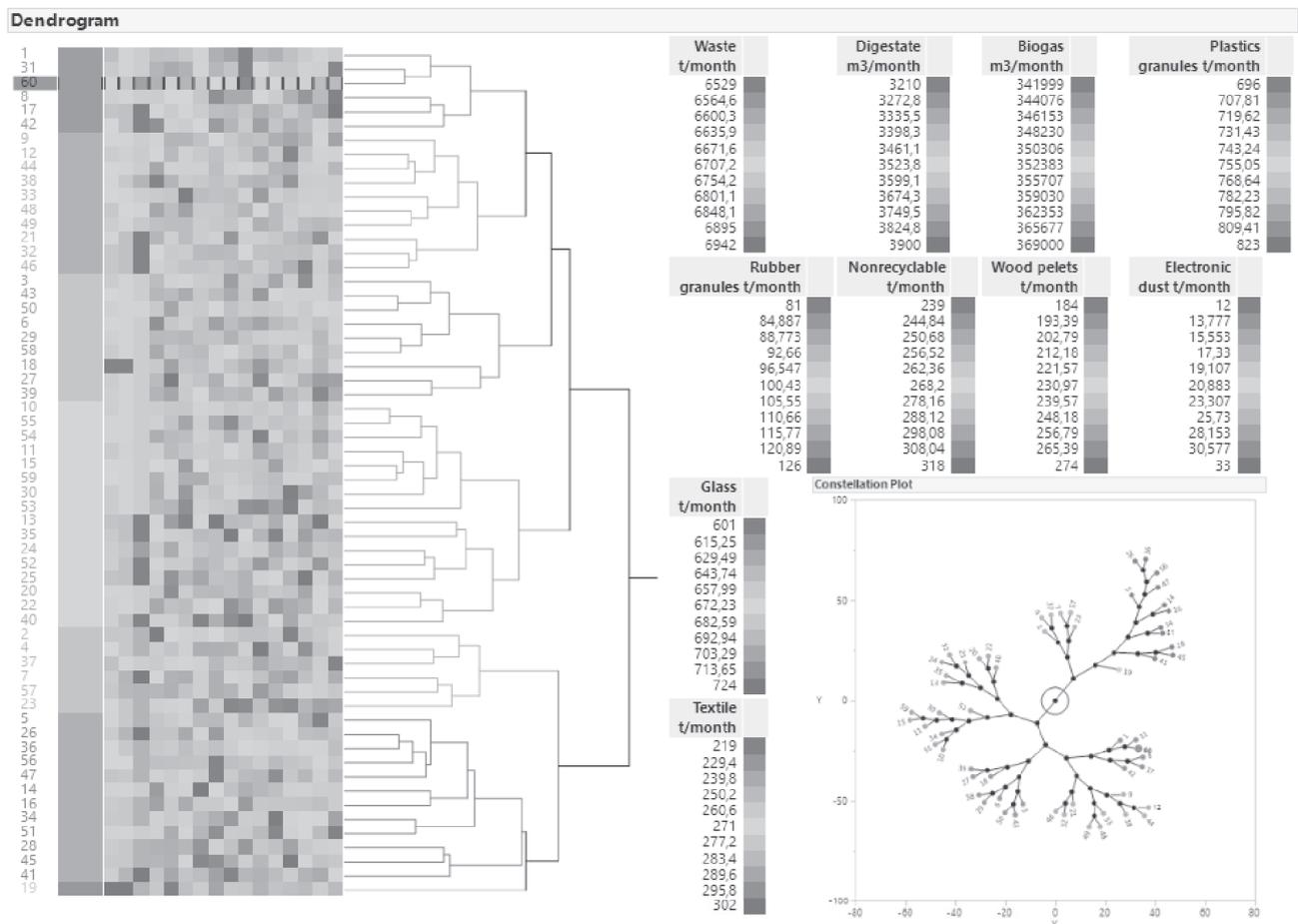


Fig. 3. Hierarchical clustering method.

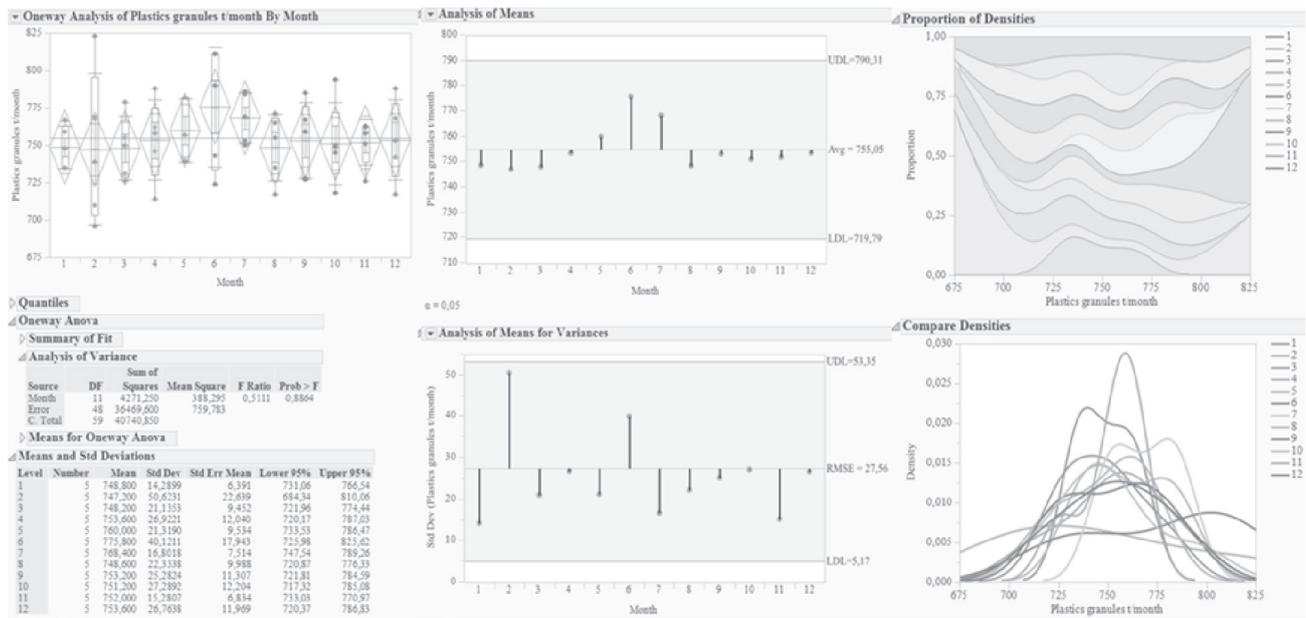


Fig. 4. Analysis of plastics granule variability.

pairs of clusters and connects the two closest clusters. This process continues until all points are contained in one cluster. Hierarchical clustering is also called agglomerative clustering due to the combined approach it uses. The agglomerative process is shown as a tree called the dendrogram (Fig. 3).

Analysis of Variability by Months of Selected Types of Sorted Waste

The variability of the quantity produced of a particular type of waste in relation to each month was analysed by the analysis of ANOVA, ANOM,

ANOM for variances, graphical analysis of proportion densities and compared densities. Our aim was to exclude or confirm the impact of seasonality on waste production. In most cases, the statistically significant variability of the variable in relation to the calendar month was not confirmed. We also examined the variability of the standard deviation in particular months, which represents the variance of the values, in relation to the mean deviation for the type of waste. The results notice the type of waste where statistically significant deviations and certain months have been recorded which are different from the rest of the year (Figs. 4-6).

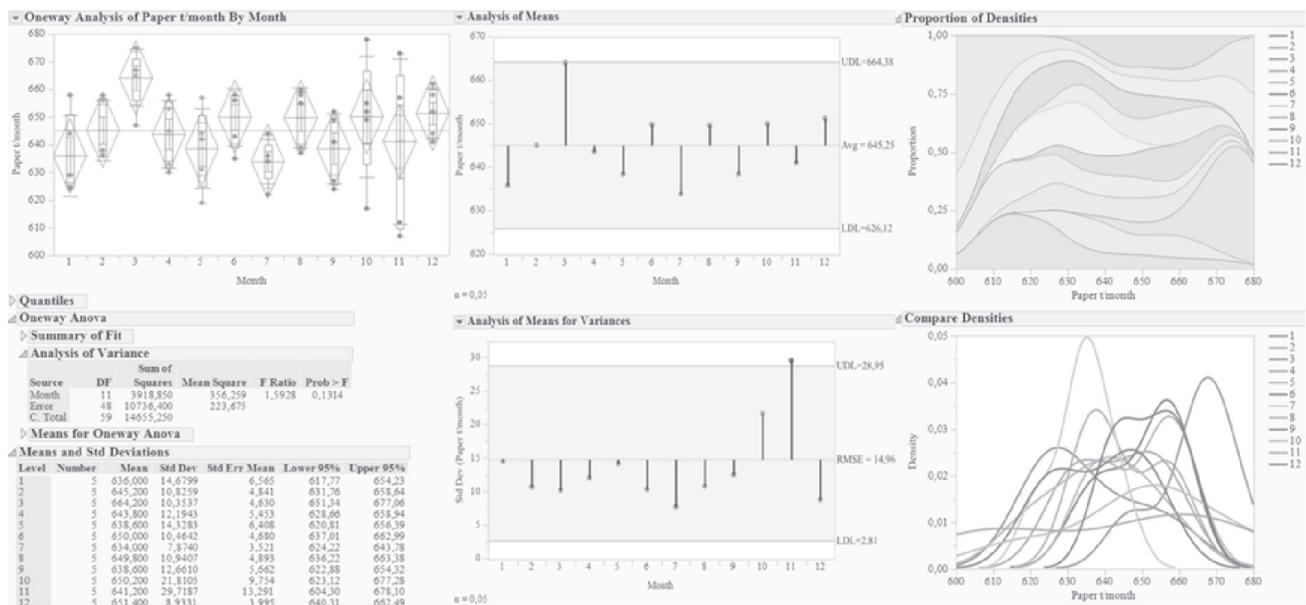


Fig. 5. Analysis of paper variability.

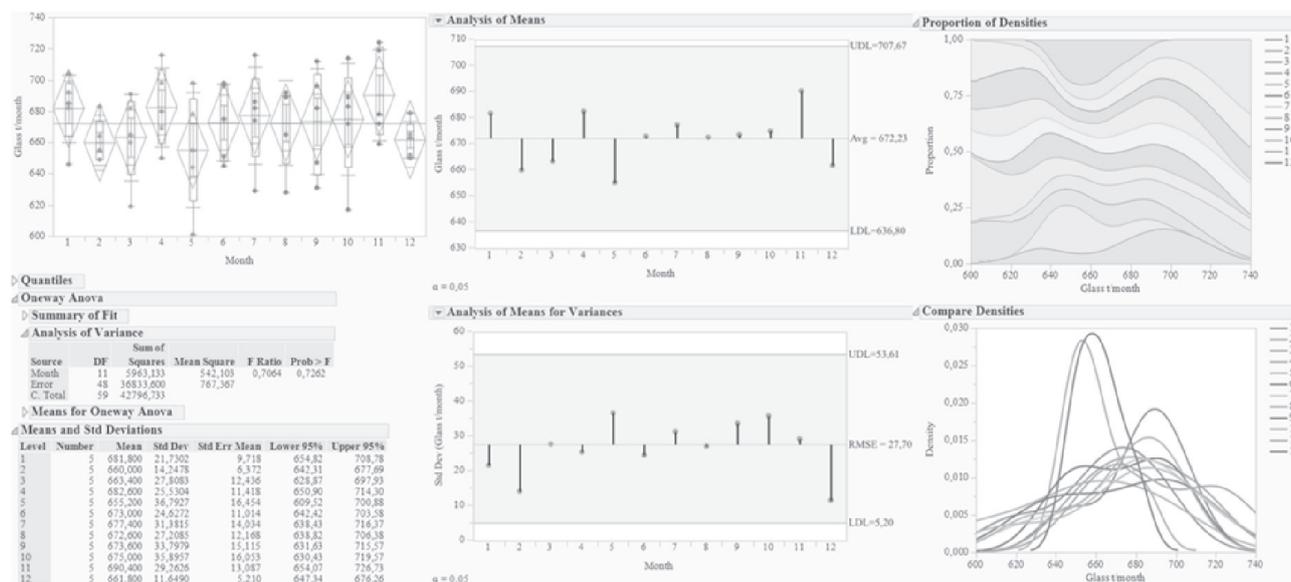


Fig. 6. Analysis of glass variability.

Conclusions

Data that represents 15 types of waste during 5 years was input for analysis. It defined types by means of descriptive statistics: the average, the standard deviation, the minimum, the maximum and the sum of waste that forms the most important part of the total waste production. Thanks to that fact, plastic granules with an average of 755.05 t/month were identified as the most important, followed by glass with an average of 672.233 t/month and paper with an average of 645.25 t/month. The persistence of waste generation was examined through the variation coefficient in order to reduce the risk of supply of these secondary raw materials in the downstream supply chain. From this point of view, already selected waste components can be considered relatively stable with a variation coefficient in the range 2.4-4.1%) to the least stable types, like electronic dust, with a coefficient of variation of almost 23%.

Linearity of the relationship among different types of waste is determined by using multidimensional pair correlation analysis. Five statistically significant negative and four positive dependencies were identified from the 15 types of waste. When repeating the analyses by data filtering according to the seasons (spring / summer / autumn / winter), there was a significantly higher dependence that differs in each season. That is why the analysed ANOVA and ANOM variability of production of individual components of waste by months was provided, but the results were not statistically significant in most cases. It would be appropriate to further explore this area by other data sorting, e.g., ¼ year, ½ year, etc. Finally, we found common features of each type of waste by means of cluster analysis, and clustering represented by one type of waste on the basis of similar behaviour, which resulted in the reduction of

six waste components: digestate, glass, non-recyclable waste, rubber granules, plastic granules and nonferrous sand.

Acknowledgements

The submitted paper is part of the project “Research and development of new smart solutions based on principles of Industry 4.0, logistics, 3D modelling and simulation for production streamline in the mining and building industry” VEGA 1/0317/19, funded by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences, and KEGA 002TUKE-4/2017 Innovative Didactic Methods of Education Process at University and their Importance in Increasing Education Mastership of Teachers and Development of Students Competences.

Conflict of Interest

The authors declare no conflict of interest.

References

1. STRAKA M., KHOURI S., ROSOVA A., CAGANOVA D., CULKOVA K. Utilization of computer simulation for waste separation design as a logistics system. *International Journal of Simulation Modelling*, **17** (4), 583, **2018**. doi:10.2507/IJSIMM17(4)444
2. STRAKA M., KHOURI S., PAŠKA M., BUŠA M., PUŠKAŠ D. Environmental assessment of waste total recycling based on principles of logistics and computer simulation design. *Polish Journal of Environmental Studies*, **28** (3), 1, **2019**. doi:10.15244/pjoes/89540

3. TANG Q.Y., ZHANG CH.X. Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. *Insect Science*, **20** (2), 254, **2012**. doi:10.1111/j.1744-7917.2012.01519.x
4. JONES B., SALL J. JMP statistical discovery software. *Wiley Interdisciplinary Reviews: Computational Statistics*, **3** (3), 188, **2011**. doi:10.1002/wics.162
5. SMITH R.I., DICK J. MCP., SCOTT E.M. The role of statistics in the analysis of ecosystem services. *Environmetrics*, **22** (Special Issue), 608-617, **2011**. doi:10.1002/env.1107
6. WANG Y., HULSTIJN J., TAN Y.H. Data quality assurance in international supply chains: an application of the value cycle approach to customs reporting. *International Journal of Advanced Logistics*, **5** (2), 76, **2016**. doi:10.1080/2287108X.2016.1178501
7. BOHÁCS G., GÁSPÁR D., KÁNYA D. Conception an intelligent node architecture for intralogistics. *Acta logistica*, **5** (2), 31, **2018**. doi:10.22306/al.v5i2.82
8. WANG G., GUNASEKARAN A., NGAI E.W.T., PAPADOPOULOS T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, **176** (6), 98, **2016**. doi:10.1016/j.ijpe.2016.03.014
9. BARNAGHI P., SHETH A., HENSON C. From data to actionable knowledge: big data challenges in the web of things. *IEEE Intelligent Systems*, 67, **2013**.
10. GENPACT (NYSE: G) stands for "generating business impact. GENERATING SUPPLY CHAIN IMPACT: Driving supply chain excellence through Data-to-Action Analytics. *Genpact*, 1, **2015**.
11. LITTLE R.J.A., RUBIN D.B. *Statistical Analysis with Missing Data*. John Wiley & Sons, Wiley Series in Probability and Statistics, **2014**.
12. STRAKA M., LENORT R., KHOURI S., FELIKS J. Design of large-scale logistics systems using computer simulation hierarchic structure. *International Journal of Simulation Modelling*, **17** (1), 105, **2018**. doi:10.2507/IJSIMM17(1)422
13. BELLER M., BHOLANATH R., MCINTOSH S., ZAIDMAN A. Analyzing the State of Static Analysis: A Large-Scale Evaluation in Open Source Software. 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER), Suita, Osaka, Japan, 470, **2016**. doi:10.1109/SANER.2016.105
14. CHEBA K., KIBA-JANIAK M., SANIUK S., WITKOWSKI K. Modeling transportation preferences of urban residents: The case of Poland. 1st International Summit on Internet of Things, IoT360 2014, Rome, Italy, 27-28 October 2014, 78, **2014**.
15. WIELAND R., VOSS M., HOLTSMANN X., MIRSCHEL W., AJIBEFUN I. Spatial Analysis and Modeling Tool (SAMT): 1. Structure and possibilities. *Ecological Informatics*, **1** (1), 67, **2006**. doi:10.1016/j.ecoinf.2005.10.005
16. MALINDZAK D. Models and simulation in logistics. *Acta Montanistica Slovaca*, **15** (SI.1), 1, **2010**.
17. SOBRINO D.R.D., KOŠTÁL P., CAGÁŇOVÁ D., ČAMBÁL M. On the possibilities of intelligence implementation in manufacturing the role of simulation. *Applied Mechanics and Materials*, **309**, 96, **2013**. doi:10.4028/www.scientific.net/AMM.309.96
18. TREBUNA P., HALCINOVA J. Experimental modelling of the cluster analysis processes. *Modelling of mechanical and mechatronics systems, Procedia Engineering*, **48**, 673, **2012**.
19. BAILEY T.C., GATRELL A.C. *Interactive spatial data analysis*. Longman Scientific & Technical, **1995**.
20. KNAPCÍKOVÁ L., DUPLÁKOVÁ D., RADCHENKO S., HATALA M. Rheological Behavior Modelling of Composite Materials used in Engineering Industry. *TEM Journal*, **6** (2), 242, **2017**. doi:10.18421/TEM62-07
21. BAZELEY P., JACKSON K. *Qualitative data analysis with NVivo*. 2nd edition, SAGE Publications Ltd., London, **2013**.
22. STRAKA M., MALINDZAKOVA M., ROŠOVA A., TREBUNA P. The simulation model of the material flow of municipal waste recovery. *Przemysl Chemiczny*, **95** (4), 773, **2016**. doi:10.15199/62.2016.4.12
23. SANIUK A., JASIULEWICZ-KACZMAREK M.,SAMOLEJOVA A., SANIUK S., LENORT R. Environmental favourable foundries through maintenance activities. *Metalurgija*, **54** (4), 725, **2015**.
24. TREBUNA F., SIMČAK F., BOČKO J., TREBUNA P., SARGA P. Experimental and numerical analysis of failure in press supporting structure. 48th International Scientific Conference on Experimentalni Analýza Napeti Experimental Stress Analysis, Velke Losiny, Czech Republic, 31 May-03 June 2010, 467, **2010**.
25. BUŠA M., KAZIMÍROVÁ I., PAŠKA M., PUŠKÁŠ E., FARKAS C. Planning and economic perspective of material flow. *Acta logistica*, **5** (1), 25, **2018**. doi:10.22306/al.v5i1.88
26. SANIUK S., SANIUK A. Challenges of industry 4.0 for production enterprises functioning within cyber industry networks. *Management systems in production engineering*, **26** (4), 212, **2018**.
27. STRAKA M., TREBUNA P., ROŠOVA A., MALINDZAKOVA M., MAKYSOVA H. Simulation of the process for production of plastics films as a way to increase the competitiveness of the company. *Przemysl Chemiczny*, **95** (1), 37, **2016**. doi:10.15199/62.2016.1.3