MENDEL
Soft Computing

HEURISTIC METHODOLOGY FOR FORECASTING OF QUANTITIES IN WASTE MANAGEMENT

Veronika Smejkalová[1], Radovan Šomplák[2,3], Vlastimír Nevrlý[3], Martin Pavlas[2,3]

[1]Institute of Mathematics
[2]Sustainable Process Integration Laboratory - SPIL, NETME Centre
[3]Institute of Process Engineering
Faculty of Mechanical Engineering, Brno University of Technology
Technická 2896/2, 616 69 Brno
Czech Republic
Vlastimir.Nevrly@vutbr.cz

Abstract: *The forecast of waste production and disposal is an important requirement for a future waste management planning. The problem is very often a short time series of the database. This paper suggests an approach to forecast the production of multiple waste types in micro-regions taking into account this challenge by combining many techniques. The heuristic methodology consisting of few steps is formulated. First, the input data are transformed and the methods from cluster analysis are repetitively applied. The second step is about a determination of quality for trend functions based on historical data. In the last step is performed the testing. The different type of representatives from cluster analysis is used to calculate indices of determination which are compared. This procedure is repeated until the criteria hit. The proposed approach reduced the computational time and managed to aggregate micro-regions with a similar trend. The forecast should have contributions in terms of building new facilities or adaptations to the existing ones, where it is necessary to estimate the production of waste for several years in advance. The article includes a case study of production forecast for several waste types in territorial units of the Czech Republic. The forecast is based on data in years 2009–2014 and following year 2015 was used to assess the quality of the final models. In the future, the database will expand and thus it will be possible to make more precise estimates and to develop statistical methods to measure this prognostic tool.*

Keywords: *forecasting, cluster analysis, waste production, short-time series, regression analysis.*

1 Introduction

In developed countries, the emphasis is placed on the use of renewable and secondary sources because of the limited number of primary sources. Within the waste management, currently, the transition from the existing waste treatment to the circular economy is preferred. The circular economy is based on reuse of waste or material recovery for the manufacture of new products. Some waste can no longer be further used (material and technical limitations: for example, a limited amount of recycling cycles, technological restrictions, economic factors, environmental and legislative restrictions) so it is intended for energy utilisation. In the worst case scenario (following waste treatment hierarchy), the waste is landfilled [1]. For the efficient reuse of waste, it is necessary to build the required processing infrastructure which is based on analysis of processing options and subsequent use of the waste in specific locations. Assessments of the current and future waste streams are indispensable in the waste management planning. Therefore, the focus is on forecasting of waste production and the creation of high-quality forecasts, which are crucial for the calculations. The study [2] deals with previously published models of municipal solid waste production. There is also mentioned a number of modelling approaches, using economic, socio-demographic or management-oriented data. The paper [3] also shows a review of models for solid waste production predictions and introduces the most common attributes affecting the production of waste. Most of the published models utilised regression analysis, where the relation between socio-economic factors and the amount of waste was investigated, see [4]. However, these factors are not always available in required detail for micro-regions. Thus, these methods are not applicable in general and so the utilisation for purposes of this paper is limited. In some cases, the only explanatory parameter is time and thus the estimation of future production is based on the analysis of time series. The study [5] compared time series models for forecasting the amount of solid waste for the upcoming years. The field of interest were autoregressive integrated moving average (ARIMA), autoregressive moving average model (ARMA) and exponential smoothing models. However, these models require a long time series that is not always available

in waste management and therefore models are unsuitable. In the paper [6] some tools for time series analysis and forecasting are proposed to evaluate municipal solid waste production. A prediction technique based on non-linear dynamics is given, comparing its performance with a seasonal ARIMA methodology, dealing with short and medium-term forecasting. A short-time series algebraic prediction technique with mixed smoothing is presented in the paper [7], several models (Moving Average - MA, Single Exponential Smoothing - SES, ARIMA, Algebraic Prediction with External Smoothing - APES, Algebraic Prediction with Internal Smoothing - APIS) were tested on data sets of different size (the shortest time series comprised 18 data points) and from different areas. The research [8] contributed to short-time series prediction based on a methodology by means of Bayesian enhanced modified approach (BEMA) utilising relative entropy for issuing a short rainfall series prediction.

This paper presents a novel approach to the estimate of a functional relationship describing the trend of a given data set. The motivation was to develop a new method since the approaches in the previously discussed papers were insufficient: they were not suitable for short-time series and they cannot work without socio-economic or other data for regression analysis, either. Based on the nature of the data, the newly developed system will select a suitable course of the trend for various types of waste produced in various microregions. The selection uses heuristic methodology consisting of cluster analysis combined with normalisation of historic data and evaluation of conditions of the regression function quality. Resulting dendrogram illustrates the categorization of particular types of waste into clusters. The logic behind it is to find a suitable model for the specific data since each type of waste and micro-region may have different characteristics. In order to facilitate the whole process, testing is done only in time series which best characterise the course of data in each cluster (representatives). Forecasting of production of various types of waste for very short-time series is the result of this approach.

For the prediction is essential to have a quality database which contains information about the production and processing of a certain type of waste. The proposed method is tested through a case study for waste management data in the Czech Republic, where the system of reporting waste data was changed in the recent past years, which is also the problem of many other countries. In the case of the Czech Republic, a new method of the reporting has been used since 2009. This led to a reduction of data usability and so the problem arises with a short-time series. The analysis was performed with data from years 2009–2014, while 2015 was utilised to compare and select the final models. These results may be utilised in stochastic network design, e.g in [9] for Wait-and-See, in [10] for Here-and-Now see [9], or facility location problem [11].

2 Data and Procedure

Analysis of waste production is performed for several types of waste in the Czech Republic on a micro-regional level (206 microregions). The paper discusses the following groups of waste:

- Mixed municipal waste (MMW)

- Certain sorted segments of waste (paper, plastic, glass)

- Part of the municipal waste, designated as MW* in this paper (MW* = MMW + paper + plastic + glass)

- Biowaste (Bio)

- Bulky waste (BW).

The aim of this work is to identify regression functions which will well describe trends in historic data. Data regarding the above-mentioned types of waste, available for the work, come from the period of 2009—2015. With respect to this data, there is a problem with short-time series. The presented calculation system will be tested for 206 microregions of the Czech Republic, which corresponds to 1,442 time series that require a regression model. This is a rather time-consuming task, but also practical applications call for a suitable aggregation of similar waste producers. Identification of similarities between particular regions (types of waste) may also help introduce modern methods and implement novel technologies (campaigns) in waste treatment. Therefore, this paper proposes a methodology which allows to carry out an analysis only for a sub-set of data and form suitable groups of micro-regions. Fig. 1 illustrates this procedure which is based on three basic steps (S1 through S3).

1. Cluster analysis (S1): the approach is divided into several stages:

    (a) Data preparation: normalisation, difference and calculation of distance matrix for time series

    (b) Formation of clusters (noted as C#) from all input data and selection of cluster representatives (R1 type). These representatives enter S2.

(c) Formation of sub-clusters from clusters defined in phase S1b and selection of their representatives (R2 type). These representatives enter S3.

2. Selection of trends for R1 representatives (S2)

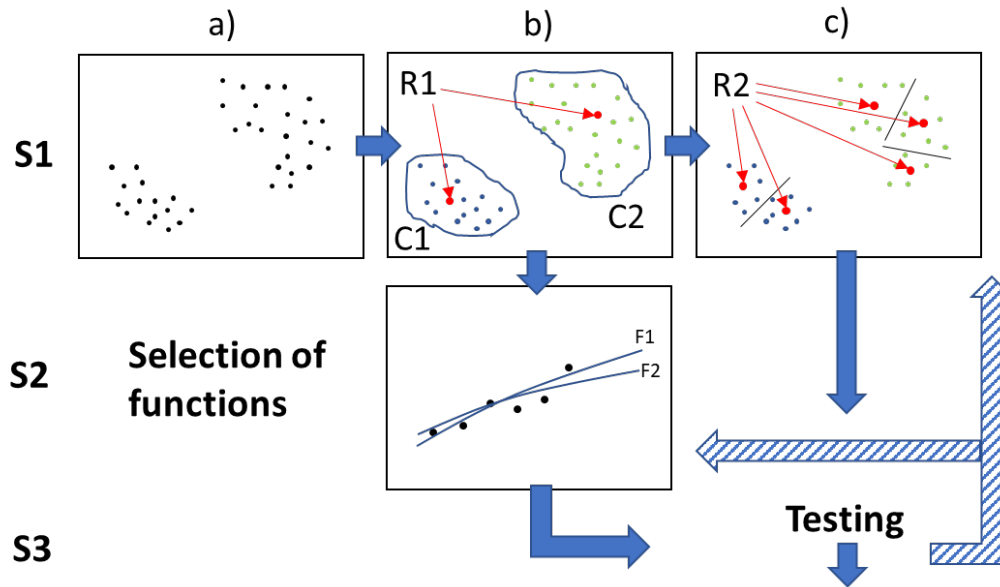3. Testing of selection from S2 on representatives R2 (S3)

Figure 1: Illustration of the procedure

Selection of regression functions is done using R1 representatives and software TableCurve 2D which employes Grammatical Evolutiuon algorithm, see [12]. Subsequent testing (S3) is performed on sub-cluster representatives R2 using already selected functions for R1. The logic behind this is to find a suitable regression function for each cluster and thus also for all time series. Based on the chosen criterion, the procedure is repeated, see Algorithm 1 which shows pseudocode.

3 Cluster Analysis (S1)

The reason for integration of S1 and also the reason for this whole section is the necessity to reduce the number of scenarios for where is aim to find the trend course. The cluster analysis was used; time series from all micro-regions and all types of waste were clustered on a principle of similar course in time. Each series of historic data was normalised to interval $\langle 0; 1 \rangle$ by Chebyshev distance in the first step. The aim of this step was to transform the time series so that the size of waste production didn't matter. Normalisation was based on the relation:

$$\tilde{x}_{i,t} = \frac{|x_{i,t}|}{\max_t |x_{i,t}|} \qquad \forall i \in I, \forall t \in \{1, \ldots, T\}. \tag{1}$$

Now, time series with a characteristic production trend will be aggregated within the cluster analysis. This rule is now applied to 6 points of historic data (2009–2014) which creates 5 differences for each time series according to equation (2).

$$\tilde{\Delta}_{i,t} = \tilde{x}_{i,t+1} - \tilde{x}_{i,t} \qquad \forall i \in I, \forall t \in \{1, \ldots, T-1\}. \tag{2}$$

These differences are used to calculate a distance matrix of time series (distance between two time series i and j). Considering future applications with more data available, it was opted for d_1 Manhattan distance. A commonly used Euclidean distance behaves differently in high dimensionality, and cannot describe the differences in data there, see [13], where d_1 distance is better to be applied.

Cluster analysis offers a series of methods which at the core may be described as hierarchical and non-hierarchical. Suitability of the methods was verified in experiments; considering the silhouette values (see below), Ward's method is deemed as a satisfactory hierarchical method, and k-means and PAM (Partitioning around medoids) as adequate non-hierarchical methods [14]. The Ward's method is implemented in S1b. In

[15], the use of the Manhattan distance for Ward's method is discussed. Final dendrogram with highlighted clusters is in Fig. 2. There are no universal rules available for identifying a suitable number of clusters. In general, the best cut in a dendrogram is the value on the vertical axis with the significantly higher distance of elements in the cluster as opposed to a cut with a higher number of clusters. A C-index criterion was opted to choose the best cut point, see [16]. In the following dendrogram, there is a cut for 10 clusters: C1 through C10 (see Fig. 2). Categorization of particular types of waste into clusters is in Table 1. The value shows a number
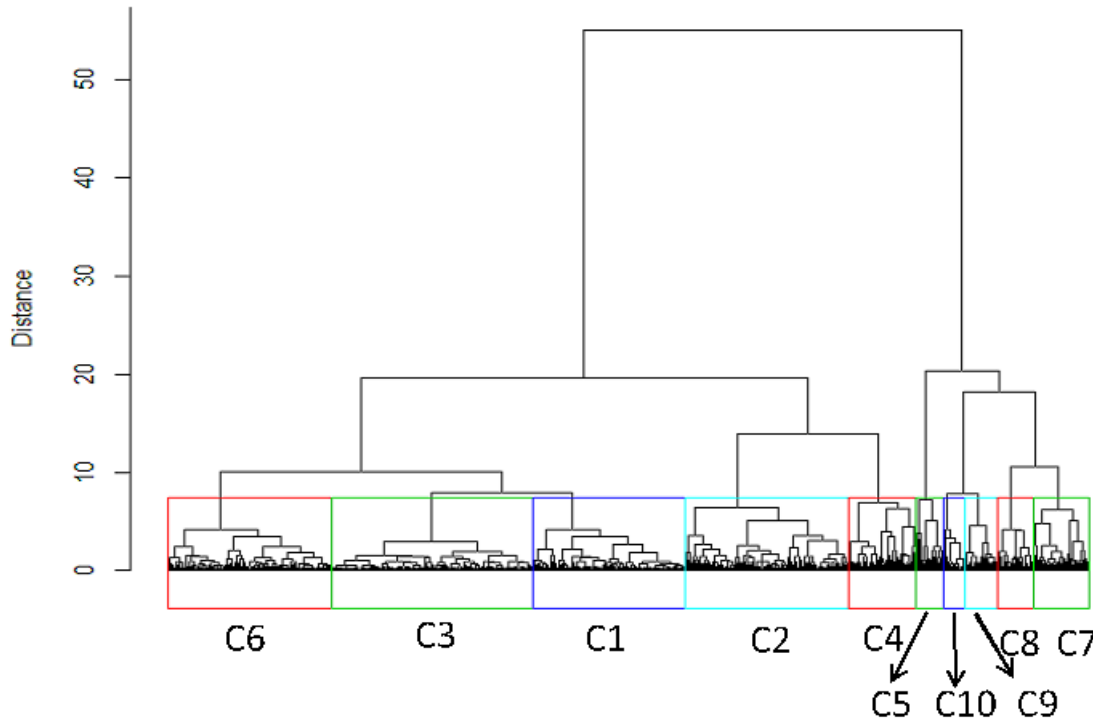

Figure 2: Illustration of the procedure

of percentage points of all time series for a given type of waste which has been aggregated under a particular cluster. Results prove that various types of waste have a similar trend (a character of historic data) in various microregions and therefore emerge repeatedly in certain clusters. It may be observed that MMW has in most cases a progress similar to cluster C1 or C3. A similar character may be observed between paper and plastic. An in-depth analysis of clusters (analysis results are not illustrated in Table 1) proves that production of paper and plastic in 31.6% of microregions is included in the identical cluster. In addition, results for biowaste are quite interesting as the biowaste is quite different from others and forms its own clusters C8–C10 (see Table 1).

Table 1: Grouping of particular waste types into clusters

	MMW	Paper	Plastic	Glass	MW*	Bio	BW
C1	41.75%	10.68%	4.37%	16.02%	25.73%	0.49%	16.50%
C2	7.28%	29.13%	36.41%	30.10%	8.74%	1.46%	11.65%
C3	45.63%	5.34%	6.80%	27.67%	60.68%	0.97%	5.83%
C4	3.88%	6.80%	1.94%	6.31%	2.91%	3.40%	25.73%
C5	0.97%	1.46%	0%	0%	0.97%	5.34%	12.62%
C6	0.49%	37.86%	41.75%	18.45%	0.97%	11.17%	13.59%
C7	0%	7.77%	7.77%	1.46%	0%	14.08%	11.17%
C8	0%	0.49%	0.49%	0%	0%	24.27%	1.46%
C9	0%	0.49%	0.49%	0%	0%	23.30%	1.46%
C10	0%	0%	0%	0%	0%	15.53%	0%

After that, representatives are selected so that the amount of time series, which are to be tested for suitability of the model, may be reduced. The representative should cover the variability of the whole cluster. This situation is illustrated in Fig. 1b) with the highlighted element that has been selected as a representative of type R1

of the given cluster. The representative of each cluster is the element with the smallest distance from other elements.

Regression models (see section 4) were identified for the cluster representatives R1. But the model should suit all time series which fall under this cluster. In addition, the proposed procedure for testing (see section 5) requires formation of sub-clusters and identification of their representatives (R2), see Fig. 1c). This means that each basic cluster (in different colours in the figure) was further partitioned, which formed more representatives for the final evaluation.

A number of sub-clusters in Fig. 1c) and consequently also a number of tested representatives was identified on the basis of cluster compactness. In other words, fewer representatives R2 were selected so that a higher-quality cluster may be formed. Less compact cluster requires more representatives R2. Silhouettes were used as the evaluation method for the cluster compactness. Values of silhouettes are calculated as follows:

$$s_{k,j} = \frac{b_{k,j} \cdot a_{k,j}}{\max(a_{k,j}, b_{k,j})} \qquad \forall j \in C_k, \forall k, \tag{3}$$

where j is the element (time series) of the cluster k. Parameter $a_{k,j}$ is an average distance of j element to all other elements in the cluster k and parameter $b_{k,j}$ is the lowest average dissimilarity of j to any other cluster k, of which j is not a member. Silhouette values $s_{k,j}$ range between the interval $\langle -1; 1 \rangle$. Higher $s_{k,j}$ values are typical for elements which belong to a cluster with more elements and shorter distances. Therefore, a compact and well-partitioned cluster has a high $s_{k,j}$ value. A negative value indicates that the element may not be properly grouped in the cluster [17]. The following visualisation (Fig. 3) shows that cluster C3 and C10 have well-grouped data and fewer representatives R2 will be subject to testing, as opposed to cluster C4, for example.
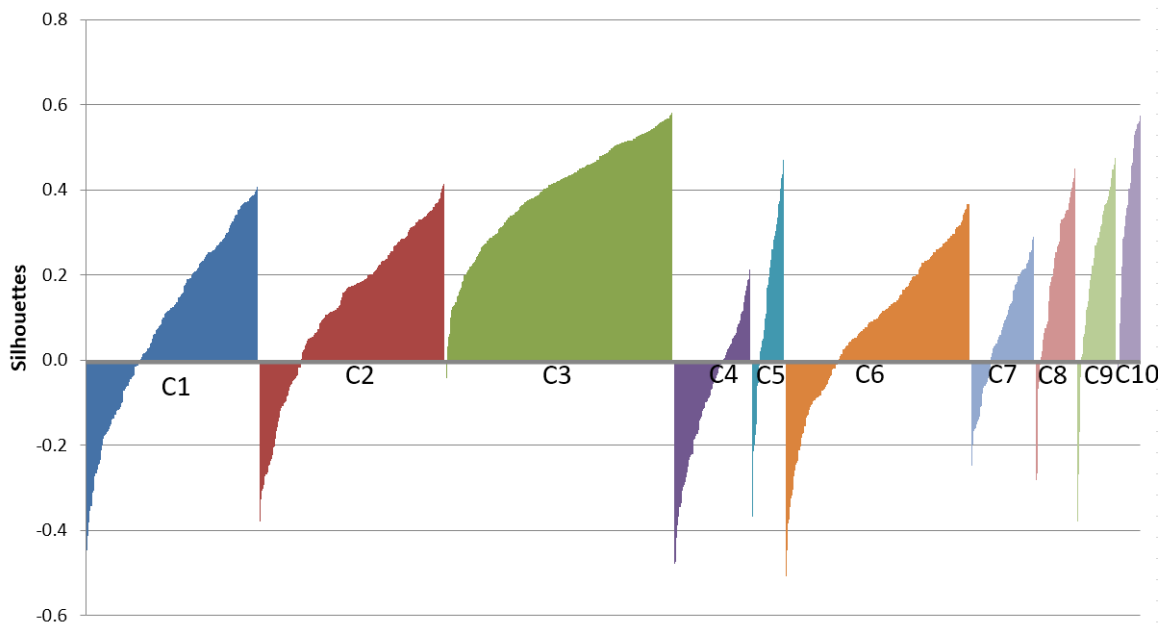

Figure 3: Silhouettes of clustering

For clusters designated C1–C10, number of representatives R2 ranges from 3 to 5 based on the average silhouette value of all cluster elements (selection of the range was done with respect to the data characteristics, see Table 2). Based on the selected criterion, the numbers of representatives of type R2 correspond to individual clusters, see Table 3. It was decided to apply the k-means method to identify sub-clusters and their representatives as it is more suitable for this type of a cluster analysis (number of clusters is given). In the S3, these representatives (of type R2) serve for testing evaluation.

Table 2: Number of sub-clusters in relation to the clusters' silhouette values

Number of Representatives	Average of $s_{k,j}$
5	< 0.1
4	$\langle 0.1, 0.2 \rangle$
3	> 0.2

Table 3: Number of sub-cluster for clusters C1–C10

Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Average of s_{kj}	0.10	0.13	0.40	-0.09	0.13	0.08	0.07	0.21	0.23	0.37
Number of Representatives R2	5	4	3	5	4	5	5	3	3	3

4 Functions Selection (S2)

One representative (of type R1) was selected for each of the 10 clusters on the basis of the smallest distance from all other cluster elements, see Fig. 1b. A trend (regression) function was identified for those 10 time series (representatives R1) based on data 2009–2014 using TableCurve 2D software [18]. The software has a database of more than 3,600 regression functions and suggests suitable models for the input data. However, only one time series may be subject to analysis in one calculation cycle. This is one of the reasons why the data had to be clustered. There are 8 categories of regression function databases in the TableCurve 2D: Linear, Simple, Robust Straight Line, Poly/Ratl, WaveForm, Peak, Transition, Kinetic. Some of the functions may be grouped into more categories (for example, Linear and Simple). With respect to the model characteristics, only functions from the Linear, Simple, Peak, and Transition equations categories were used, as the requirement was a monotonous trend for a long forecasting period. For testing was chosen a reasonable time horizon (ca. 10 years from the last available data) for waste forecasts. Models whose extreme lies beyond the mentioned time interval were accepted.

12 different models were assigned to each cluster based on the data of representative R1 (3 best models for each functions category). Various functions with the highest determination coefficient were results of the selection process. In the most cases, functions with a high index of determination were identified (over 0.95), with the exception of cluster C9, for example, which changed the trend over the time (see below). Selected regression functions from S2 are further tested in S3 using representatives of type R2. This either proves the suitability of the selected regression functions or unsuitability thereof (the procedure must be then repeated with representatives R2, see Algorithm 1). The S2 includes 10 out of 1,442 elements, and S3 30 out of 1,442. This significantly increases time-efficiency of the process.

5 Testing (S3) and Results

As mentioned above, 12 models were selected for each representative R1 in S2 (3 models from each of the considered function categories). These models are in S3 evaluated using the criterion, that is the ratio of indices of determination for representatives R1 and R2. The ratio was set to value 0.85, which permitted 15% deterioration. Benefits of this approach, based on the cluster analysis, come from a properly selected regression function for most of the elements in each of the clusters. This is also the reasoning behind the evaluation of the model's selection. Testing was performed on time series of representatives of particular sub-clusters R2 (see Fig. 1b)). Input data for testing included only 2009—2014. Evaluation of the proper cluster formation and suitable function identification is done using a developed criterion, see Algorithm 1. If the criterion is not met, the calculation goes back to S1c) with representatives R2 and the process starts again. Values from 2015 served for final assessment (upon completion of Algorithm 1) of models for all waste types, which resulted in Table 4.

Major errors in the final assessment of representatives occurred in cluster C9; this means 3 testing elements with errors of -15.2%, -36.5% and -60.8% in comparison with 2015. This is data of biowaste which has witnessed a staggering increase since 2014. In 2014, an amended act providing for waste ordered municipalities to establish an obligatory location for biowaste collection. Trend analysis is obviously unable to reflect legislation interventions: all of the ca. 3,600 regression functions failed to well describe trends in cluster C9. A more detailed study of the trend is recommendable. Other tested clusters showed an error of up to 10% (ca. 60% of results had an error of less than 5%); this error had a symmetrical distribution around 0. Therefore, the error might be caused by random factors which force the data to oscillate around this trend.

In general, a universal relation cannot be identified for a description of the waste production trend. The extrapolation should be always performed in all nodes (microregions) for all types of waste considered. At that point, however, it is necessary to forecast functions shape for identification of the regression coefficients; this is of course time-consuming. Time requirements of the analysis may be greatly reduced, as described in this paper. Two options may cover this situation:

- Option A: one universal regression function is identified; analysis in all nodes is performed for this function.

- Option B: TableCurve 2D or similar software is used for all microregions and all data, the best relation is always identified, but it is very time-consuming.

The approach presented in this paper is a compromise between the two options; it tries to save time by identifying the one relation, for a cluster involving more microregions and types of waste.

Algorithm 1

1: **S1a)**
2: **Calculate** distance matrix from normalised differences
3: **S1b)**
4: **Do** hierarchical cluster analysis using Ward's method
5: Choose number of clusters based on criterion $\Rightarrow p$ representatives of type $R1$
6: **S1c)**
7: **for all** clusters **do**
8:     **Calculate** Silhouettes
9:     Choose number of subclusters k based on Table 1
10:     **Do** cluster analysis using k-means method $\Rightarrow k$ representatives of type $R2$
11: **end for**
12: **S2**
13: **for all** representatives of type $R1$ **do**
14:     **Do** function selection using TableCurve 2D
15: **end for**
16: **S3**
17: **for all** representatives p of type $R1$ **do**
18:     **for all** respective representatives k of type $R2$ **do**
19:         **Calculate** coefficients of determination I for $R1_p$ and $R2_k$
20:         **if** $\frac{I_{R2_k}}{I_{R1_p}} \leq 0.85$ **and** more then one time series in cluster **then**
21:             **GoTo S1c)** with cluster of $R2_k$
22:         **end if**
23:     **end for**
24: **end for**

The proposed algorithm resulted in Table 4, which summarises the most frequent functions suitable for analysed waste types. These functions were selected on the basis of the quality of the estimations in 2015. The table also includes average of indices of determination for all time series of respective waste type. The paper and plastic are described by the same functions and even the glass has a similar character. The numbers of functions are ordered according to the percentage of represented time series. It is worth noting Bio, where model $y = 0$ (number 10) described micro-regions that have not yet reported any waste.

Table 4: Characteristic functions for analysed waste types

Waste type	MMW	paper	plastic	glass	MW*	Bio	BW
Function #	11, 3, 4, 8	2, 9, 1, 5, 6	2, 9, 1, 5, 6	2, 9, 7	3, 8, 4, 7	10, 1, 7	5
Average of R^2	0.750	0.798	0.842	0.867	0.677	0.806	0.763

$$1 \quad y = a + \frac{b}{\left(1+e^{-\frac{x-d\ln(2^{\frac{1}{e}}-1)-c}{d}}\right)^e}$$

$$2 \quad y = a + be^{\left(-e^{-\left(\frac{x-c}{d}\right)}-\frac{x-c}{d}-1\right)}$$

$$3 \quad y = a + \frac{b}{\left(\frac{1+(x-c)^2}{d^2e}\right)^{\frac{e}{2}+\frac{1}{2}}}$$

$$4 \quad y = a + \frac{b}{1+e^{-\frac{(x-c)}{d}}}$$

$$5 \quad y = a + b\frac{atan(\frac{x-c}{d})+\frac{\pi}{2}}{\pi}$$

$$6 \quad y = a + bx^{1.5} + cx^{2.5} + d\frac{\ln x}{x}$$

$$7 \quad y^{-1} = a + be^{-x}$$

$$8 \quad \ln y = a + be^{-x}$$

$$9 \quad y = a + bx + \frac{c}{x} + \frac{d}{x^2}$$

$$10 \quad y = 0$$

$$11 \quad y^{-1} = a + \frac{b}{x^2}$$

6 Conclusion

In this paper, the procedure for forecasting a large number of very short-time series is presented. In order to reduce the time necessary for the calculations, the approach based on a cluster analysis was applied. The aim of the presented calculation system was to forecast a trend in historic data. This has been achieved by choosing data representatives using a cluster analysis. The trend was identified using regression functions in TableCurve 2D software (ca. 3,600 regression functions have been tested) and twelve of these functions were selected. Regression functions have been selected for the clusters on the basis of their representatives, and have been later tested only for a partial amount of the analysed time series (ca. 2.7%). The process of cluster analysis,

functions selection and testing was repeated until the criterion hit. The procedure of heuristic methodology was described in a case study dealing with the production of certain types of municipal waste in the Czech Republic. Altogether, 7 types of waste in 206 microregions were forecasted, that is 1,442 time series. The model was then compared to 2015 data which has not been incorporated in the analysis (trend was identified only for 2009–2014 data) and the best functions were selected. Results of this approach are perfectly suitable for further reconciliation, which was proposed in [19]. The developed forecasting system may be applied to data in various other disciplines. It is necessary to perform a complete testing before each application due to a specific nature of the data.

References

[1] Act no. 185/2001 Coll., on Waste and the Amendment of Some Other Act. In: Collection of Laws of the Czech Republic, part two, § 9a.

[2] Beigl, P., Lebersorger, S., Salhofer, S.: Modelling municipal solid waste generation: A review. Waste Management **28**(1), 200–214 (2008). DOI 10.1016/j.wasman.2006.12.011

[3] Kolekar, K. A., Hazra, T., Chakrabarty, S. N.: A Review on Prediction of Municipal Solid Waste Generation Models. Procedia Environmental Sciences **35**, 238–244 (2016). DOI 10.1016/j.proenv.2016.07.087

[4] Ojeda-Benítez, S., Vega, C. A., Marquez-Montenegro, M. Y.: Household solid waste characterization by family socioeconomic profile as unit of analysis. Resources, Conservation and Recycling **52**(7), 992–999 (2008). DOI 10.1016/j.resconrec.2008.03.004

[5] Mwenda, A., Kuznetsov, D., Mirau, S.: Time Series Forecasting of Solid Waste Generation in Arusha City - Tanzania. Mathematical Theory and Modeling. **4**(8), 29–39 (2014).

[6] Navarro-Esbri, J., Diamadopoulos, E., Ginestar, D.: Time series analysis and forecasting techniques for municipal solid waste management. Resources, Conservation and Recycling. **35**(3), 201–214 (2002). DOI 10.1016/S0921-3449(02)00002-2

[7] Palivonaite, R., Lukoseviciute, K., Ragulskis, M.: Short-term time series algebraic forecasting with mixed smoothing. Neurocomputing. **171**(1), 854–865 (2002). DOI 10.1016/j.neucom.2015.07.018

[8] Rivero, C. R., Pucheta, J. A., Baumgartner, J. S., Laboret, S. O., Sauchelli, V. H., Patino H. D.: Short-series Prediction with BEMA Approach: application to short rainfall series IEEE Latin America Transactions. **18**(8), 3892–3899 (2016). DOI 10.1109/TLA.2016.7786377

[9] Hrabec, D., Popela, P., Roupec, J., Jindra, P., Haugen, K., Novotny, J., Olstad, A.: Hybrid algorithm for wait-and-see network design problem. In: Proceedings of 20th International Conference on Soft Computing − MENDEL 2014, pp. 97–104 (2014)

[10] Roupec, J., Popela, P., Hrabec, D., Novotný, J., Olstad, A., Haugen, K.K.: Hybrid algorithm for network design problem with uncertain demands. In: Proceedings of the World Congress on Engineering and Computer Science, WCECS, Lecture Notes in Engineering and Computer Science, **1**, pp. 554-559 (2013)

[11] Hrabec, D., Viktorín, A, Šomplák, R., Pluháček, M., Popela, P. A heuristic approach to the facility location problem for waste management: A case study. In: Proceedings of 22th International Conference on Soft Computing − MENDEL 2016, pp. 61–66 (2016)

[12] Miškařík, K., Matoušek, R., Dobrovský, L., Hanna, G.: About controller design by means of grammatical evolution and bees algorithm. In: Proceedings of 21st International Conference on Soft Computing − MENDEL 2015, pp. 65–70 (2015).

[13] Aggarwal, C. C., Hinneburg, A., Keim, D. A.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. In: International Conference on Database Theory, pp. 420–434 (2001)

[14] Everitt, B.: Cluster analysis, fifth edn. Chichester, West Sussex, U.K (2011)

[15] Strauss, T., von Maltitz, M. J.: Generalising Ward's Method for Use with Manhattan Distances. PLoS ONE. **12**(1), (2017). DOI 10.1371/journal.pone.0168288

[16] Hubert, L., Schultz, J.: Quadratic assignment as a general data-analysis strategy. British Journal of Mathematical and Statistical Psychology. **29**(2), 190–241 (1987). DOI 10.1111/j.2044-8317.1976.tb00714.x

[17] Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis Journal of Computational & Applied Mathematics. **20**, 53–65 (1987). DOI 10.1016/0377-0427(87)90125-7

[18] TableCurve 2D. systatsoftware.com/products/tablecurve-2d (2017). [Online; accessed 15-May-2017]

[19] Nevrlý, V., Šomplák, R., Popela, P., Pavlas, M., Osička, O., Kůdela, J.: Heuristic challenges for spatially distributed waste production identification problems. In: Proceedings of 22nd International Conference on Soft Computing − MENDEL 2016, pp. 109–116 (2016)