

Tomasz SZUL<sup>1\*</sup>, Jarosław KNAGA<sup>1</sup> and Krzysztof NĘCKA<sup>1</sup>

## APPLICATION OF ROUGH SET THEORY TO ESTABLISH THE AMOUNT OF WASTE IN HOUSEHOLDS IN RURAL AREAS

### ZASTOSOWANIE TEORII ZBIORÓW PRZYBLIŻONYCH DO OCENY ILOŚCI ODPADÓW Z GOSPODARSTW DOMOWYCH NA WSI

**Abstract:** The method based on rough set theory (RST) was used in the study to establish the rate of mass accumulation of waste in households in rural areas, which are characterised by different economic types, in case of which traditional statistical analyses are usually hardly reliable. The following indicators available in the General Statistical Office's statistics were used in the analysis: population density, income level, main source of income, economic type of the municipality, area of agricultural land, age of the buildings and participation of gaseous fuels in meeting heat demands. The method shown should not be considered as a competition for statistical methods, but it could complement them, especially in cases when there are few objects to analyse, the more so as it proves useful in cases where input data are general, imprecise and uncertain. As has been shown in the study, with such data and a small number of objects, the relative error of estimation was 13% on average.

**Keywords:** mass accumulation of waste rate, rough set theory, rural areas

## Introduction

Municipal waste is waste generated by households, as well as waste that does not contain hazardous waste coming from other waste producers, which are similar in nature and composition to waste generated in households. General rules of waste management system within the European Union were set out in the framework Directive no 2008/98/WE [1] on waste. The Directive obliges EU Member States to provide recovery or disposal of waste without endangering human health and harming the environment. The main objective, as far as waste management is concerned, is to prevent and minimise waste production, recover, dispose and re-use the waste as well as provide environmentally friendly storage of unused waste.

In 2014, in the whole European Union, 392 Tg ( $Tg = 10^{12} g$ ) of municipal waste was generated, which corresponds to  $772 \text{ kg} \cdot (\text{person} \cdot \text{year})^{-1}$ . The amount of accumulated waste in particular countries ranged from 52 to  $1594 \text{ kg} \cdot (\text{person} \cdot \text{year})^{-1}$  [2].

The amount of produced municipal waste in a given country depends on many factors, the most important of which are: the standard of living and population as well the scope and

<sup>1</sup> Faculty of Production and Power Engineering, University of Agriculture in Krakow, ul. Balicka 116B, 30-149 Kraków, Poland, phone +48 12 662 46 47

\* Corresponding author: T.Szul@ur.krakow.pl

intensity of goods consumption. The type of generated waste depends on the type of area in which the waste is generated, population density, buildings type, the presence of public use facilities, the presence and size of retail outlets, the type of industry or services and other [3].

The amount of produced waste in Poland changes vary rapidly. Until 2004 the volume of collected waste was decreasing, which to some extent could have been caused by mass emigration of people connected with Poland's accession to the European Union and opening of labour markets, but also with the occurrence of pathological phenomena such as: creation of illegal waste dumps, not showing waste in registers kept by entrepreneurs dealing with waste collection to avoid paying fees for using the environment as well as managing waste by households on their own - often against the requirements of environment protection. According to the General Statistical Office's data, in 2015 [4] in Poland 10.3 Tg of municipal waste was generated, which corresponds to  $283 \text{ kg} \cdot (\text{person} \cdot \text{year})^{-1}$ . Figure 1 shows the amount of municipal waste generated and collected in 2015.

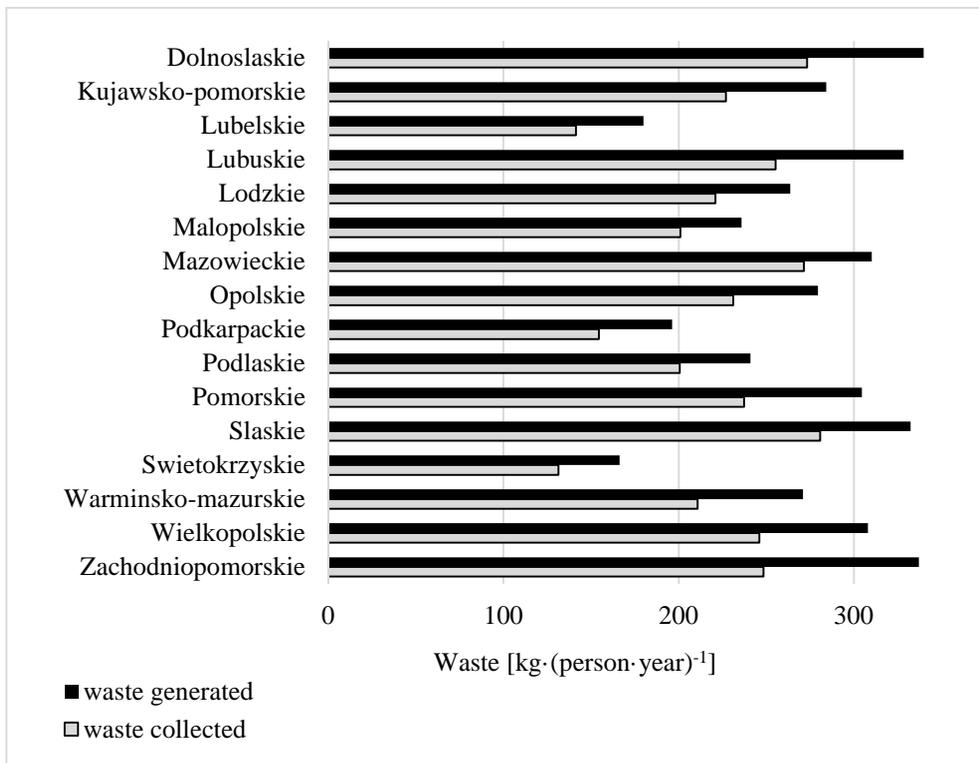


Fig. 1. Mass accumulation of municipal waste generated and collected in 2015 [5]

The greatest amounts of municipal waste in 2015 (Fig. 1) were produced in the following voivodeships: Lower Silesian, Silesian and Masovian and Pomerania (more than

330 kg·(person·year)<sup>-1</sup>). The smallest amounts of waste were produced in Swietokrzyskie and Lublin Voivodeships (160-180 kg·(person·year)<sup>-1</sup>).

The rate of mass household waste accumulation is also dependant on the place where waste is produced. Literature sources give different values of waste accumulation coefficient, which is shown in the following Table 1.

Table 1  
Rates of mass household waste accumulation depending on the place of waste production [5, 6]

Specification	[kg·(person·year) <sup>-1</sup> ]	[kg·(person·year) <sup>-1</sup> ]
big cities (> 100 thousand inhabitants)	220-400	385.9
small and medium towns (10-50 thousand inhabitants)	180-330	346.2
rural areas	90-110	233.9

The above data prove that a city dweller produces 200 kg more waste than a rural dweller (calculated per year). Such a difference in the level of waste production results from higher consumption in the cities and higher standard of living of their inhabitants. Moreover, in the rural areas many food products come from people's own farming, whereas city inhabitants buy all such products, which entails using packaging and generating packaging waste. What is more, the majority of organic waste is further used by rural inhabitants. Participation of municipal waste generated in rural areas is between 21% (General Statistical Office 2014) and 29% (Eurostat 2014). Municipal waste in rural areas is produced mainly in households which produce 76% of mixed municipal waste (in the cities this participation is 70%). Mixed waste stream produced by an average household in Poland is approximately 160 kg·(person·year)<sup>-1</sup>. This rate is 89 kg·(person·year)<sup>-1</sup> in case of rural areas. Lower rate of waste accumulation compared to the cities is among other the consequence of the following circumstances:

- a part of municipal waste stream is recovered by people on their own (*e.g.* home composting). What is more, a part of the waste stream is managed illegally;
- not all waste management companies have the possibility of checking the weight of dustcarts entering the facilities;
- the weight of received waste is often under-reported in order to pay lower storage fees [7].

## Methods of estimating waste stream

The studies on municipal waste management touch upon the problem of the way of estimating the amount of accumulated waste. Many authors have done research on the strength and direction of influence of different economic and social factors on the mass of generated or collected waste. Statistical models that describe the relationship between socio-economic factors and waste characteristics are most often based on the information provided by municipalities or statistical offices [7-15]. Waste analyses confirm that their weight and composition change (in time and space) depending on many factors. The factors that are taken into account in statistical studies can be classified in the following way:

- socio-economic factors;
- individual consumption;
- production and trade.

The study [8] describes 45 statistical models (published before 2006), concerning the relationship between the weight of generated waste and the factors that characterise the areas where the waste was produced. Most of these studies were developed for the needs of waste management planning on the national level or for urban areas. Models based on these studies are usually multi-factor. Relevant statistical data are not available, which makes it impossible to verify the models in other areas. Analyses on the national [16, 17] or regional level show general relations between the amount of waste and population density or gross domestic product. However, these models cannot be applied for planning the management of waste on lower regional levels.

The influence of individual inhabitants' income and the national income on the amount of produced municipal waste is well documented by numerous research results. In countries with higher GDP (Gross Domestic Product) bigger amounts of waste per capita are produced.

Other factors that affect the weight of collected waste are, according to the following:

- a) the size of a given municipality (area, number of inhabitants);
- b) the level of economic development;
- c) land use structure;
- d) the system of waste collection;
- e) the level of ecological awareness of the local community.

In the research aimed at determining the direction and strength of influence of different features characterising municipalities (or regions) on the weight and composition of waste generated within them, apart from the above-mentioned factors, also the following were taken into consideration:

- a) population density - where the authors [11, 12, 18] state, that weight and quality of collected waste is strictly connected with population density;
- b) household size - the average number of people living in a house/flat [8, 11, 12];
- c) urbanisation level - the number of newly built flats, kindergartens, etc. [8, 10, 18];
- d) buildings type and heating system - in the rural areas and small towns, where the majority of buildings are detached houses (equipped with furnaces for solid fuel), the weight of waste should be lower (compared to urban areas), whereas the main component of the waste should be ash. In multi-family buildings type (in big cities), where there is no possibility of burning waste in furnaces, neither of feeding livestock, the main component of waste should be biodegradable fractions [8, 10, 11, 17];
- e) other technical and sanitary equipment of buildings [11];
- f) administrative, functional and economic type of the municipality [8];
- g) saturation of technical infrastructure facilities - density of road, sewage, water supply networks, including the number of sewage connections, etc. [8];
- h) affluence („standard of living”) and inhabitants' lifestyle - according to the affluence of inhabitants measured with GDP is one of the most important factors affecting the amount of generated municipal waste. The greater the society's affluence the bigger the volume of generated waste.
- i) municipality's income from taxes calculated per one inhabitant [19];
- j) eating habits and health indicators, such as lifespan and infant mortality, as well as the age structure of population - in [5] it was stated that the smallest amount of waste is generated in retired people's households, while the composition of waste is the most diverse in young society;
- k) tradition and people's habits [5, 12];

- l) social factors - the education level, including people with higher education, feminisation rate, birth rate, the rate of migration attractiveness, as well as the number of pharmacies or libraries per 1000 people [8, 11, 12, 18];
- m) the number of unemployed people, the level and structure of employment - in the study [11] it was stated that houses/flats of people employed in industrial sectors of economy produced more waste than those of people employed in services, which might have been the result of a higher rate of selective waste collection in the latter group of households;
- n) tourism - the number of accommodation places, hotels, guesthouses, etc. [5];
- o) climate factors (the temperature and precipitation) and the season of the year [5].

Some of the mathematic models that explain the weight of generated waste list also factors directly connected with municipal waste management [8, 13, 17]:

- a) fees for waste collection and disposal calculated per one inhabitant or per one tonne and the frequency of waste collection;
- b) the number and capacity of containers calculated per one household - furnishing houses with small capacity containers motivates the inhabitants to collect waste selectively;
- c) participation of ashes and/or biodegradable waste in the mixed municipal waste stream;
- d) participation of waste from infrastructural facilities in the total weight of municipal waste;
- e) percentage of municipality's/city's inhabitants covered under waste collection system;
- f) participation of waste collected selectively;
- g) level of contamination in selectively collected waste;
- h) participation of households composting organic waste.

Each of the above-mentioned factors has a certain influence on the weight and composition of produced waste, which may be described with linear or logarithmic relationships. The authors showed in [15] the substantial influence of some factors on waste production in European countries and cities chosen for the study. The analysis of correlation between the studied factors showed that the number of indicators should be limited to GDP, one of the social factors and the rate of employment in particular sectors. Statistical models developed and described in [16] the basis of data from 55 European cities of high and low level of prosperity explain 65% of the variance.

In study [11] the authors analysed more than 30 indicators describing 509 Austrian municipalities. They applied 3 indicators for developing the formula (model explaining 74% of variability):

- a) average size of a household [people];
- b) participation of households equipped with furnace for solid fuels [%];
- c) participation in taxes comprising national budget income personal income tax [ $\text{€}(\text{person}\cdot\text{year})^{-1}$ ].

The approaches to analysing the rate of mass accumulation of waste, presented in the chosen literature, use mainly statistical methods in the form of linear regression models [5, 6], multiple regression [20] and artificial neural networks [7].

Regressive analyses are usually time-consuming and one can rarely estimate in a precise way this indicator in small areas - especially rural, due to the lack of data, inaccuracy and their uncertainty. With regard to the above, the authors of the study suggest the procedure of estimating the rate of mass accumulation of mixed municipal waste in

households with the use of rough set theory (RST), which is created for the analysis of imprecise, general and uncertain data.

## **Rough set theory**

Rough set theory created by Polish computer scientist Zdzisław Pawlak [21] is used in the study of imprecision, generality and uncertainty in the process of data analysis. The suggested method is a fine theoretical base for solving problems concerning intelligent decision making systems. It is concentrated on providing huge sets of data with necessary formulae, rules and information. Rough sets are used as a methodology in the process of discovering knowledge in databases, which is usually an iterative and interactive process (with many decisions made by a user). This theory is one of the methods of knowledge representation [22, 23] which creates a platform for effective inference.

RST is one of the fastest developing fields of artificial intelligence due to the fact that it is an important tool in data mining process. RST is a certain theory of knowledge (theory of information systems) and serves as a tool for describing uncertain, imprecise knowledge, for modelling approximation and decision making systems as well as systems of feature and classification recognition. The results of theoretical study within RST regard logics, set theory, knowledge representation, data filtering, algorithmic problems connected with information systems. Although developed a short time ago, rough set theory is used in a number of new fields of study. Nowadays it is used both in medicine, pharmacology, economics, banking, chemistry, sociology, acoustics, linguistics, general engineering as well as in diagnostics of machines, geography, land management and environmental engineering - publications of results can be found, inter alia, in [23-27].

Methodology of inference with the use of Rough Set Theory includes only the qualitative nature of objects characteristics. It is a certain hindrance and limitation with regard to the application of methods based on classical assumptions of RST to the attributes of objects which are described with regard to the occurrence of features in a quantitative form, *e.g.* population density, gross domestic product index or participation of a given feature in the whole. Changing them into qualitative form is not a big problem, but it causes a major loss of information accuracy. In this case, integrating rough sets with fuzzy sets in the form of application of valued tolerance relation - VTR, proved helpful [28]. It allowed the introduction of greater flexibility in data mining to rough set theory and the possibility to analyse observations expressed in a quantitative form. Classic assumption of RST is based on the indiscernibility relation concept as a precise equivalence relation, which is the objects will only then be indiscernible when they have similar attributes (system 0-1). Introduction of valued tolerance relation to RST allows determination of the upper and lower approximation of a set with a different level of indiscernibility relation [28]. Owing to this solution, one can compare two sets of data and achieve a result in the 0...1 range, which constitutes the level of indiscernibility relation. This range is a membership function derived from the assumptions of the fuzzy set theory. The closer the result to 1, the more similar are the objects (indiscernible) with regard to the analysed attribute, and the closer it is to 0, the more discernible they are [28].

## **Aim of the study**

The aim of the study is to introduce new method based on rough set theory (RST) to determine the rate of mass accumulation of waste from households in rural areas. This new

method of estimating the amount of waste in rural areas, based on RST, will allow the use of a quick, relatively easy procedure to determine the rate of mass accumulation in the rural municipalities and rural areas of urban and rural municipalities, where the statistical analyses are usually of little reliability. The suggested procedure uses the assumptions of rough set theory in the application of valued tolerance relation (VTR), to analyse data which is presented in a quantitative form (continuous). This method, and it has to be clearly expressed, does not constitute a competition for statistical methods, but it was created as an alternative possibility of choosing the procedure for the estimation of the amount of waste.

## Area of study

The calculations were made on the basis of Malopolska Voivodeship's statistical data [4]. In the analysed area, 1001 thousand Mg was generated, which constitutes 7.8% of the waste stream on a national scale. The indicator expressing the amount of waste produced per one inhabitant was  $236 \text{ kg}\cdot(\text{person}\cdot\text{year})^{-1}$  for Malopolska Voivodeship and it was only slightly lower than the national average, *i.e.*  $283 \text{ kg}\cdot(\text{person}\cdot\text{year})^{-1}$ . An average household in Malopolska produces  $200 \text{ kg}\cdot(\text{person}\cdot\text{year})^{-1}$ , while in the rural areas it is an average of approximately  $65 \text{ kg}\cdot(\text{person}\cdot\text{year})^{-1}$ . There are 168 rural and urban and rural municipalities in the analysed area. The characteristic feature of Malopolska Voivodeship's rural areas is a wide range of the rate of mass accumulation of mixed municipal waste produced by households. It varies between 20 and  $230 \text{ kg}\cdot(\text{person}\cdot\text{year})^{-1}$ . Also morphological composition of waste collected in the rural areas is not the same in different households which is the result of different population density, buildings type, functional type of the municipality, income level, participation of agricultural land or forests in the land management structure and the amount of ash produced in the process of burning solid fuels in households' boilers and furnaces.

Rural municipalities differ from one another as far as economic types are concerned. The literature on the subject features several classifications of rural areas, which group Malopolska's municipalities on the basis of their economic types and functional structure of cities and municipalities [29].

## Calculation results

The application of rough set theory was presented on the example of the set of 60 randomly chosen rural municipalities and rural areas of urban and rural municipalities of Malopolska Voivodeship. The number of objects within the set was chosen in a way to enable the level of confidence of 95%. On the basis of literature data, the municipalities within the rural areas of Malopolska Voivodeship were divided into three types - suburban, tourist and agricultural municipalities [29]. Then, the municipalities chosen for the analysis were divided into two subsets: the training set (information system) containing 40 objects and the test set comprising of 20 objects.

Objects within the training set were presented in the form of a decision table (Table 2) where the features characterising the municipalities (condition attributes) were marked with symbols  $c_1-c_7$ , and the rate of mass accumulation of waste in households, which is a decision attribute was marked with  $d$  symbol.

Table 2

Information system (decision table)

Municipality/ object number	Condition attributes							Decision attribute
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$d$
1	146	1.92	0.565	0.35	3	0.52	392	52.4
2	151	2.59	0.561	0.37	3	0.69	239	38.7
:	:	:	:	:	:	:	:	:
15	387	1.48	0.595	0.59	1	0.19	665	118.5
:	:	:	:	:	:	:	:	:
40	142	1.58	0.523	0.44	2	0.5	358	119.2

Source: own study on the basis of General Statistical Office's data

For the aforementioned attributes, domains were determined according to the following assumptions:

- $c_1$  - population density [people · km<sup>-2</sup>],
- $c_2$  - average area of agricultural land [ha],
- $c_3$  - building's age rate (established as a weighted arithmetic mean of the number of buildings from different periods of time *i.e.* before 1944, 1945-1970, 1971-1988, 1989-2002, 2003-2012),
- $c_4$  - participation of buildings heated with natural gas,
- $c_5$  - municipality type (1 - suburban, 2 - tourist, 3 - agricultural),
- $c_6$  - participation of households deriving income from agricultural activity,
- $c_7$  - income rate (municipalities' own income - participation in taxes comprising national budget income personal income tax [PLN·(person·year)<sup>-1</sup>],
- $d$  - the rate of mass accumulation of waste in the households [kg·(person·year)<sup>-1</sup>].

The values of particular attributes were established on the basis of statistical data included in the Regional Data Bank of General Statistical Office and General Agricultural Census available on the General Statistical Office's website [4].

The next step of the procedure of creating a decision making model according to RST was establishing decision rules. 40 decision rules can be determined from Table 2, the number equal to the analysed objects, which will be determined in the following manner:

1. if ( $c_1 = 146$ ) and ( $c_2 = 1.92$ ) and ( $c_3 = 0.565$ ) and ( $c_4 = 0.35$ ) and ( $c_5 = 3$ ) and ( $c_6 = 0.52$ ) and ( $c_7 = 392$ ) then ( $d = 52.4$ );
2. if ( $c_1 = 151$ ) and ( $c_2 = 2.59$ ) and ( $c_3 = 0.561$ ) and ( $c_4 = 0.37$ ) and ( $c_5 = 3$ ) and ( $c_6 = 0.69$ ) and ( $c_7 = 239$ ) then ( $d = 38.7$ );
- 3... 39;
40. if ( $c_1 = 142$ ) and ( $c_2 = 1.58$ ) and ( $c_3 = 0.523$ ) and ( $c_4 = 0.44$ ) and ( $c_5 = 2$ ) and ( $c_6 = 0.5$ ) and ( $c_7 = 358$ ) then ( $d = 38.7$ ).

General assumptions of RST prove correct mainly on the assumption that all attributes are qualitative data. Adopting attributes in a continuous way (necessary for determining the rate of waste accumulation), in the further analysis the use of VTR was presumed. Another stage of the procedure of creating a decision making model to determine the rate of waste accumulation is calculating the  $k$ -coefficient from the model of valued tolerance relation, which is represented by formula [28]:

$$R_j(x, y) = \frac{\max(0, \min((c_j(x), c_j(y)) + k - \max((c_j(x), c_j(y))))}{k} \quad (1)$$

where  $R_j(x,y)$  is the relation between two sets with a membership function  $[0, 1]$ ;  $c_j(x), c_j(y)$  - variable of the analysed object;  $k$  - coefficient taken as a standard deviation in the set of a given attribute of the analysed object.

Table 3 shows  $k$ -coefficient determined on the basis of standard deviation calculated for each of the attributes from Table 2.

Determined  $k$ -coefficient

Table 3

Attribute	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$d$
$k$	78.18	1.54	0.072	0.201	0.797	0.173	199.1	30.6

Having calculated the  $k$ -coefficient, one can determine matrices for each attribute from Table 2 (decision), containing the results of calculated valued tolerance relation.

Next, on the basis of the aforementioned matrices calculated for condition attributes  $c_1$ - $c_7$  the matrix of sums for individual objects (Table 4) was created, according to the assumption:

$$R_j(x, p) = \max \left( \sum_{j=1}^n R_j(x, p) \right) \quad (2)$$

where:  $R_j$  - valued tolerance relation;  $x$  - attribute of the considered object;  $p$  - attribute belonging to the conditional part of the considered decision rule;  $n$  - number of the object attributes in the conditional part of the decision rule.

The next stage is establishing abstraction classes for a given indiscernibility relation. In the presented analysis, due to the introduction of valued tolerance relation, it was necessary to modify the general assumptions of the procedure of establishing indiscernibility relation in rough set theory. This study assumes that indiscernible objects are such, whose sum deriving from the matrices of valued tolerance relations (formula (2)) is 70% of the maximum value. This value was adopted arbitrarily, on the assumption that maximum similarity (indiscernibility) for all condition attributes is 7 (as there are 7 condition attributes) due to the uncertainty of data and because the majority of condition attributes is in a form of continuous recording (hence there are not two identical municipalities) it was assumed that the level of similarity will be sufficient.

According to the aforementioned assumptions, the following classes of abstraction (indiscernibility) were established for condition attributes, including at the beginning, for the purpose of simplification and greater clarity - 40 decision classes (the number equal to the number of objects), investigating the similarity of each object from Table 3 with every other object. Then, the reduction of indiscernibility classes of condition attributes was carried out due to the repeatability of objects in successive initial decision classes. The above-mentioned procedure allowed the reduction of the initial number of indiscernibility classes from 40 to 36. The results are given in Table 4.

The next step of the above-mentioned procedure was determining indiscernibility classes for the decision attribute. The established abstraction classes are given in Table 5, including the results from the matrices of valued tolerance relations. Due to the occurrence of maximum similarity at level 1, the sufficient level of similarity on level above 0.8 was assumed (meaning that the similarity was assumed on level above 80%) [28]. Adoption of the mentioned similarity coefficient on level of 80% can be accounted for with a very high

diversity of the rate of waste accumulation in specific municipalities and the attribute's continuity of record.

Table 4

Indiscernibility classes for condition attributes

Initial 40 indiscernibility classes				Reduced indiscernibility classes			
Decision class	The results of sums of chosen objects indiscernible due to a given object	Decision class	The results of sums of chosen objects indiscernible due to a given object	Decision class	The results of sums of chosen objects indiscernible due to a given object	Decision class	The results of sums of chosen objects indiscernible due to a given object
1	1	21	21	1	1	22	22, 24, 33
2	2, 3, 11	22	22, 24, 33	2	2, 3, 11	23	19, 23
3	2, 3, 6	23	19, 23	3	2, 3, 6	24	11, 22, 24, 25
4	4	24	11, 22, 24, 25	4	4	25	24, 25
5	3, 5, 11	25	24, 25	5	3, 5, 11	26	26
6	6	26	26	6	6	27	27, 28, 29
7	7	27	27, 28, 29	7	7	30	30
8	8, 32	28	27, 28, 29	8	8, 32	31	31
9	9	29	27, 28, 29	9	9	32	8, 11, 12, 32
10	10, 39	30	30	10	10, 39	33	11, 22, 33
11	2, 5, 11	31	31	11	2, 5, 11	34	34
12	12	32	8, 11, 12, 32	12	12	35	35
13	13, 14	33	11, 22, 33	13	13, 14	36	12, 36
14	13, 14	34	34	15	15	37	37, 38
15	15	35	35	16	16, 39	39	10, 16, 39
16	16, 39	36	12, 36	17	17	40	40
17	17	37	37, 38	18	18		
18	18	38	37, 38	19	23		
19	23	39	10, 16, 39	20	20		
20	20	40	40	21	21		

Reduction of classes for the decision attribute was not carried out, as can be seen in Table 5, due to the feature's continuity of record and the lack of repeatability of the decision attribute. Thus, each of the objects is a separate indiscernibility class, and there are 40 of them.

Having determined indiscernibility classes for condition and decision attributes it was possible to establish representative decision rules for the analysed data set. In order to distinguish representative rules,  $C$ -approximation of the sets family was calculated (Table 6) according to the following formulae:

- *Approximation accuracy:*

$$\beta_c(X) = \frac{\text{card}(\underline{O}X)}{\text{card}(\overline{O}X)} \quad (3)$$

- *Approximation quality:*

$$\gamma_c(X) = \frac{\text{card}(POS_c(U))}{\text{card}(\underline{O}X_i)} \tag{4}$$

where:  $\underline{O}X$  - the number of lower approximation objects (cardinality of the lower approximation of X set);  $\bar{O}X$  - the number of upper approximation objects (cardinality of the upper approximation of X set);  $POS_c$  - the number of objects in the indiscernibility class of a decision attribute.

Table 5

Indiscernibility classes for condition attribute

Decision class	The results of sizes of chosen objects indiscernible due to a decision attribute	Decision class	The results of sizes of chosen objects indiscernible due to a decision attribute
1	1, 5, 22, 24, 25, 30, 36	21	2, 21, 25, 38, 40
2	2, 14, 27, 34	22	6, 22, 27, 37
3	3, 8, 10, 18, 23, 33, 39	23	13, 14, 23
4	4	24	1, 5, 19, 24
5	1, 5, 22, 24, 25, 30, 36	25	16, 21, 27, 33, 37
6	6, 15, 21, 28, 40	26	20, 3
7	7, 15, 21, 40	27	3, 22, 25, 27
8	3, 8, 10, 18, 23	28	2, 9, 10, 12, 17, 20, 28, 38
9	9, 6, 7, 16, 31, 35, 36	29	11, 12, 17, 18, 29
10	10, 14, 19, 20, 30, 32, 39	30	1, 5, 10, 30, 33, 37
11	11, 16, 35, 36	31	12, 17, 31
12	3, 8, 12, 13, 29, 30, 32, 39	32	32, 4
13	6, 10, 12, 13, 18, 24	33	3, 25, 30, 33, 39
14	14, 2	34	6, 7, 10, 34
15	6, 7, 15, 17, 38, 40	35	1-8, 12, 17, 18, 35, 39
16	1, 5, 16, 26	36	1-8, 10, 15, 32, 33, 36
17	9, 3, 8, 15, 17, 29, 30, 32, 39	37	10-21, 35, 37
18	1, 5, 10, 11, 13, 18, 19, 20, 30, 39	38	12-19, 21-33, 38
19	1, 5, 9, 18, 19, 25, 38	39	21-36, 39
20	11, 20, 35, 36	40	35, 36, 40

In the course of calculations, certain modification of the general assumptions of decision rules on the basis of  $c$ -accurate sets was made, according to [24]. Due to the continuity of attributes and application of VTR, such decision rules were adopted, whose approximation quality and accuracy was higher than 0. Thus it was assumed that decision rules are correct if they have at least one element in the lower approximation of a set. The number of lower approximation objects ( $\underline{O}$ ) was determined *i.e.* the set of objects, for whom it was possible to assign, on the basis of condition attributes, an accurate value, as well as the number of upper approximation objects ( $\bar{O}$ ), thus also those objects whose condition attributes did not always lead to the same decision. The results of the analyses are given in Table 6, rejected rules are written in bold type. For the establishment of the rate of mass accumulation of waste of the objects set (municipalities) 35 decision making procedures were used, with the following numbers: 1, 2, 4-10, 12-22, 24-26, 28-31, 33-40.

Applying valued tolerance relation (VTR), the author checked to which of the chosen decision rules the analysed municipality has the highest level of membership (Table 7).

Table 6

C-approximation of the classification of sets of the  $d$  decision attribute family

Decision attribute number	Number of objects in the indiscernibility class of a decision attribute	Number of lower approximation objects	Number of upper approximation objects	C-accuracy of approximation	C-quality of approximation
1	7	2	11	0.18	0.29
2	4	1	9	0.11	0.25
<b>3</b>	<b>7</b>	<b>0</b>	<b>14</b>	<b>0.00</b>	<b>0.00</b>
4	1	1	1	1.00	1.00
5	7	2	9	0.22	0.29
6	5	4	7	0.57	0.80
7	4	4	4	1.00	1.00
8	5	1	10	0.10	0.20
9	7	5	9	0.56	0.71
10	7	1	13	0.08	0.14
<b>11</b>	<b>4</b>	<b>0</b>	<b>11</b>	<b>0.00</b>	<b>0.00</b>
12	8	1	17	0.06	0.13
13	6	2	13	0.15	0.33
14	2	1	3	0.33	0.50
15	6	5	7	0.71	0.83
16	4	2	7	0.29	0.50
17	9	3	17	0.18	0.33
18	10	4	17	0.24	0.40
19	7	3	10	0.30	0.43
20	4	2	10	0.20	0.50
21	5	2	9	0.22	0.40
22	4	1	9	0.11	0.25
<b>23</b>	<b>3</b>	<b>0</b>	<b>4</b>	<b>0.00</b>	<b>0.00</b>
24	4	1	9	0.11	0.25
25	5	1	11	0.09	0.20
26	2	2	2	1.00	1.00
<b>27</b>	<b>4</b>	<b>0</b>	<b>10</b>	<b>0.00</b>	<b>0.00</b>
28	8	3	16	0.19	0.38
29	5	2	13	0.15	0.40
30	6	2	9	0.22	0.33
31	3	2	5	0.40	0.67
<b>32</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>0.00</b>	<b>0.00</b>
33	5	1	12	0.08	0.20
34	4	2	5	0.40	0.50
35	13	7	18	0.39	0.54
36	13	5	17	0.29	0.38
37	14	5	23	0.22	0.36
38	21	7	26	0.27	0.33
39	17	6	23	0.26	0.35
40	3	2	4	0.50	0.67

Having distinguished representative decision rules the author determined the rate of mass accumulation of waste. For this purpose the municipalities from the test set (Table 8)

were used. The estimation procedure was presented on an exemplary municipality, marked with number 11 in the test set.

Table 7

## Choosing the appropriate decision rule

Decision rule number	Level of membership in condition attribute							Sum
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	
1	0.00	0.91	0.22	0.14	0.00	0.00	0.33	1.60
...	...	...	...	...	...	...	...	....
<b>15</b>	<b>0.00</b>	<b>0.62</b>	<b>0.64</b>	<b>0.67</b>	<b>1.00</b>	<b>1.00</b>	<b>0.30</b>	<b>4.23</b>
...	...	...	...	...	...	...	...	....
40	0.00	0.69	0.00	0.59	0.00	0.00	0.16	1.43

Table 7 shows that the rule which is the most similar to the attributes in the estimated municipality is rule no 15. Thus, the value of the rate of mass waste accumulation should equal  $118.5 \text{ kg} \cdot (\text{person} \cdot \text{year})^{-1}$  (object no 15 Table 2). According to the General Statistical Office data the actual value of this rate in 2012 was  $111.9 \text{ kg} \cdot (\text{person} \cdot \text{year})^{-1}$ , the discrepancy between the actual and estimated value being 6.6 kg. Approximation error of the estimation is 5.6%.

Table 8

## Set of test objects (test table)

No of test object	Condition attributes							Estimated value	Actual value	Approximation error of the estimation
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$			
1	223	1.37	0.673	0.59	1	0.35	838	104.4	112.5	7.8
2	164	2.06	0.406	0.48	1	0.71	333	77.2	61.6	20.2
3	117	1.68	0.507	0	2	0.47	272	97.4	84.0	13.8
4	178	3.74	0.541	0.23	2	0.37	529	90.7	87.7	3.3
5	296	1.74	0.637	0.69	1	0.13	875	113.4	99.8	12.0
6	139	1.48	0.621	0.4	1	0.73	328	77.2	95.6	23.8
7	81	2.31	0.557	0	3	0.47	225	72.8	59.3	18.5
8	163	16.9	0.461	0.58	3	0.48	269	54.3	55.3	1.8
9	153	1.38	0.538	0.156	3	0.6	317	52.4	40.4	22.9
10	63	2.36	0.73	0	2	0.22	285	112.0	135.7	21.2
11	300	2.06	0.621	0.523	1	0.19	526	118.5	111.9	5.6
12	450	1.99	0.496	0.57	2	0.18	485	97.4	88.3	9.3
13	116	1.73	0.521	0	2	0.34	402	97.4	107.1	10.0
14	97	2.73	0.532	0.27	3	0.5	173	50.5	50.0	1.0
15	73	3.5	0.412	0.33	3	0.52	169	25.6	33.0	28.9
16	141	2.78	0.539	0.31	3	0.41	307	90.7	72.3	20.3
17	101	2.94	0.493	0.29	2	0.75	296	112.7	106.4	5.6
18	161	1.86	0.6	0.3	3	0.37	428	52.0	40.7	21.7
19	91	2.11	0.591	0.289	3	0.7	412	70.3	76.0	8.1
20	211	1.38	0.508	0.33	1	0.5	505	77.2	73.2	5.2
<b>Average:</b>									<b>13</b>	

Source: own study on the basis of General Statistical Office's data

For the set of twenty municipalities constituting the test objects, condition attributes  $c_1$ - $c_7$  were assumed, then with the application of VTR levels of membership in decision rules were determined (as in Table 7) and the appropriate value of  $x$  indicator was chosen.

Table 8 shows also the actual values of  $x_0$  indicator, as well as the approximation error of estimation  $\delta$ .

Analysing the results contained in the test table one can state that the weight of waste in rural areas expressed with the rate of mass accumulation of waste can be estimated quite accurately. Approximation error of the estimation in individual municipalities ranges between 1 and 29%, with the average value of 13%. That, in comparison to the methods spoken of in chapter 2 - of estimating waste stream (26-48%) can be considered a small value and comparable with the values presented in [11] with a much smaller objects set.

## Conclusions

Presented in this paper algorithm of the construction of model for estimating the size of waste stream in rural areas can be applied in municipalities of different economic types. According to the authors, condition attributes that were subject to the analysis characterise rural areas in a proper way as they contain the information about population density, income level, main income source, economic type, area of agricultural land and also indirectly about the scope of thermal energy use (building's age rate) and solid fuels participation in meeting heat demands (the lower the participation of gas for heating, the higher the participation of solid fuels and hence more ash is produced). Unfortunately, General Statistical Office's data on the level of municipalities lack information about tourist facilities (rural tourism) and the number of accommodation places, which makes it impossible to fully characterise the municipalities - especially those of a tourist profile. The authors assume that the method shown, based on RST and valued tolerance relation, should not be considered as a competition for statistical methods, but it could complement them, especially when there are few objects to analyse, the more so as it proves useful in cases where input data are general, imprecise and uncertain. As has been shown in the study, with such data and a small set of objects, the relative error of estimation was 13% on average. Nevertheless, for further research it is advised to test the method on a bigger set of data (municipalities) gathered from several voivodeships, so as to verify the achieved results.

## References

- [1] Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives. <http://eur-lex.europa.eu/eli/dir/2008/98/oj>.
- [2] European Union Statistics. Eurostat. Waste statistics. 2016. [http://ec.europa.eu/eurostat/statistics-explained/index.php/waste\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/waste_statistics).
- [3] Ministerstwo Ochrony Środowiska. Sprawozdanie z realizacji krajowego systemu gospodarki odpadami - plan 2010. (The Ministry of Environmental Protection. Report on the implementation of the national waste management plan 2010). [http://www.mos.gov.pl/g2/big/2011\\_11/935528872b22be4e27c786bc41a3c5fc.pdf](http://www.mos.gov.pl/g2/big/2011_11/935528872b22be4e27c786bc41a3c5fc.pdf).
- [4] Główny Urząd Statystyczny. Bank Danych Lokalnych. (Central Statistical Office. Local Data Bank). [http://stat.gov.pl/bdl/app/strona.html?p\\_name=indeks](http://stat.gov.pl/bdl/app/strona.html?p_name=indeks).
- [5] KPGO Krajowy plan gospodarki odpadami 2022. (National plan of the waste disposal 2022). 2016. [https://bip.mos.gov.pl/fileadmin/user\\_upload/bip/strategie\\_plany\\_programy/DGO/Krajowy\\_plan\\_gospodarki\\_odpadami\\_2022\\_\\_\\_\\_M.P.\\_poz.\\_784\\_.pdf](https://bip.mos.gov.pl/fileadmin/user_upload/bip/strategie_plany_programy/DGO/Krajowy_plan_gospodarki_odpadami_2022____M.P._poz._784_.pdf).
- [6] Boer E, Jędrzak A, Kowalski Z, Kulczycka J, Szpadt R. A review of municipal solid waste composition and quantities in Poland. *Waste Manage.* 2010;30:369-377. DOI: 10.1016/j.wasman.2009.09.018.
- [7] Bach H, Mild A, Natter M, Weber A. Combining socio-demographic and logistic factors to explain the generation and collection of waste paper. *Resources Conserv Recycl.* 2004;41(1):65-73. DOI: 10.1016/j.resconrec.2003.08.004.
- [8] Beigl P, Lebersorger S, Salhofer S. Modelling municipal solid waste generation: a review. *Waste Manage.* 2008; 28(1):200-214. DOI: 10.1016/j.wasman.2006.12.011.

- [9] Hage O, Soderholm P. An econometric analysis of regional differences in household waste collection: the case of plastic packaging waste in Sweden. *Waste Manage.* 2008;28(10):1720-1731. DOI: 10.1016/j.wasman.2007.08.022.
- [10] Hockett D, Lober DJ, Pilgrim K. Determinants of per capita municipal solid waste generation in the Southeastern United States. *J Environ Manage.* 1995;45(3):205-217. DOI: 10.1006/jema.1995.0069.
- [11] Lebersorger S, Beigl P. Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel. *Waste Manage.* 2011;31:1907-1915. DOI: 10.1016/j.wasman.2011.05.016.
- [12] Miller I, Lauzon A, Wattle B, Ritter M, Hood J. Determinants of municipal solid waste generation and recycling in western New York communities. *J Solid Waste Technol Manage.* 2009;35(4):209-236. DOI: 10.5276/JSWTM.2009.209.
- [13] Passarini F, Vassura I, Monti F, Morselli L, Villani B. Indicators of waste management efficiency related to different territorial conditions. *Waste Manage.* 2011;31(4):785-792. DOI: 10.1016/j.wasman.2010.11.021.
- [14] Poll A.J. Variations in the composition of household collected waste. AEA/ENV/R/1839, AEA Technology, Harwell, UK. 2004.
- [15] Purcell M, Magette WL. Prediction of household and commercial BMW generation according to socio-economic and other factors for the Dublin region. *Waste Manage.* 2009;29(4):1237-1250. DOI: 10.1016/j.wasman.2008.10.011.
- [16] Daskalopoulos E, Badr O, Probert SD. Municipal solid waste: a prediction methodology for the generation rate and composition in the European Union countries and the USA. *Resources Conserv Recycl.* 1998;24(1):155-166. DOI: 10.1016/S0921-3449(98)00032-9.
- [17] Mazzanti M, Zoboli R. Waste generation, waste disposal and policy effectiveness: evidence on decoupling from the European Union. *Resources, Conserv Recycl.* 2008;52(10):1221-1234. DOI: 10.1016/j.resconrec.2008.07.003.
- [18] Emery AD, Griffiths AJ, Williams KP. An in depth study of the effects of socio-economic conditions on household waste recycling practices. *Waste Manage Res.* 2003;21(3):180-190. DOI: 10.1177/0734242X0302100302.
- [19] Keser S, Duzgun S, Aksoy A. Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. *Waste Manage.* 2012;32:359-371. DOI: 10.1016/j.wasman.2011.10.017.
- [20] Armstrong JS. *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston: Kluwer Academic Publishers; 2001. <https://marketing.wharton.upenn.edu/files/?whdmsaction=public:main.file&fileID=794>.
- [21] Pawlak Z. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Dordrecht: Kluwer Academic Press; 1991.
- [22] Nutech Solution - Science for Business. 2005. <http://www.nutechsolutions.com.pl/>.
- [23] Nguyen HS. Tolerance Rough Set Model and Its Applications in Web Intelligence. *IEEE/WIC/ACM Int Conf Web Intelligence and Intelligent Agent Technol. Workshop Proceedings.* IEEE Computer Society. 2013:237-244. DOI: 10.1109/WI-IAT.2013.189.
- [24] Pawlak Z. Rough set approach to knowledge-based decision support. *Europ J Operat Res.* 1997;99(1):48-57. DOI: 10.1016/S0377-2217(96)00382-7.
- [25] Polkowski L, Skowron A. Rough mereological calculi of granules: A rough set approach to computation. *Computat Intelligence.* 2001;17(3):472-492. DOI: 10.1111/0824-7935.00159.
- [26] Renigier-Biłozor M, Wiśniewski R. The effectiveness of real estate market versus efficiency of its participants. *Europ Spatial Res Policy.* 2012;19(1):95-110. DOI: 10.2478/v10105-012-0008-5.
- [27] d'Amato M. Rough Set Theory as Automated Valuation Methodology: The Whole Story. *Int Seminar Advances in Mass Appraisal in Deft.* 2006. [https://scholar.google.pl/scholar?q=d%EF%80%99Amato+M.+Rough+Set+Theory+as+Automated+Valuation+Methodology:+The+Whole+Story.&hl=pl&as\\_sdt=0&as\\_vis=1&oi=scholar&sa=X&ved=0ahUKEwjByNqFiYnNAhVI2ywKHQ4LDqIQgQMIGjAA](https://scholar.google.pl/scholar?q=d%EF%80%99Amato+M.+Rough+Set+Theory+as+Automated+Valuation+Methodology:+The+Whole+Story.&hl=pl&as_sdt=0&as_vis=1&oi=scholar&sa=X&ved=0ahUKEwjByNqFiYnNAhVI2ywKHQ4LDqIQgQMIGjAA).
- [28] Renigier-Biłozor M. Zastosowanie teorii zbiorów przybliżonych do masowej wyceny nieruchomości na małych rynkach (Application of rough set theory for mass valuation of real estate in small markets). *Acta Scientiarum Polonorum Administratio Locorum.* 2008;7(3):35-51. <http://wydawnictwo.uwm.edu.pl/artukul/14/czytelnia.html>.
- [29] Bański J. Współczesne typologie obszarów wiejskich w Polsce - przegląd podejść metodologicznych (Contemporary typologies of rural areas in Poland - an overview of methodological approaches). *Przegląd Geograficzny.* 2014;84(4):441-470. [http://www.rcin.org.pl/Content/51257/WA51\\_70537\\_r2014-t86-z4\\_Przeg-Geogr-Banski.pdf](http://www.rcin.org.pl/Content/51257/WA51_70537_r2014-t86-z4_Przeg-Geogr-Banski.pdf).