

Selection of variables for the Croatian municipal solid waste generation model

The Mining-Geology-Petroleum Engineering Bulletin
UDC: 502
DOI: 10.17794/rgn.2017.3.6

Preliminary communication



Anamarija Grbeš¹

¹ University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, HR-10000 Zagreb, Croatia, Postdoctoral Researcher, PhD

Abstract

Municipal solid waste (MSW) is generated in households and enterprises because of everyday human consumption. The quantity (volume) of municipal solid waste depends on the number of consumers, i.e. population. Everyday consumption depends on consumers' available money for consumption – more money available to spend, more waste is generated from consumption. Consumers' ability to produce their own food, to feed domestic animals with food residue, and to destroy waste or compost it with no especial efforts or costs (in rural and suburban households), contributes to a weaker correlation of the rural population with waste generated. The urban population, characterized by a higher income and employment rate is correlated strongly with waste generated. Non-residents and visitors such as tourists also contribute to the generation of waste. All these elements of waste generation can be shown with corresponding parameters. In the testing of the waste generation hypothesis, the parameters are tested for correlations with the generated quantity of waste, promoting the parameters into potential variables for the waste model. The second step in waste modelling is to inspect how the proposed model variables correlate among them and to select the most appropriate candidates for the model. That step is performed in research described in this article. A total of 16 variables were grouped into five groups: county descriptive variables, total population variables, rural/urban population related, additional population and economy related variables. These groups are found to be correlated among each other. From each group, the appropriate representatives are proposed: length of roads, population or households, households with and without land, tourist stays at tourist accommodation, and annual income of the county. It was concluded that the latter should be modelled to represent the real income structure of the population. The sampling unit of the data for this research was the administrative unit county. It was concluded that the special administrative unit County of the City of Zagreb should, for modelling purposes, be considered as part of the County of Zagreb.

Keywords:

waste generation model, independent variable selection, correlation analysis

1. Introduction

Municipal solid waste (MSW) generation is related with the consumption of goods and their end of life. As hypothesised in first part of the research on the Croatian MSW generation mechanism in Grbeš et al. (2016), municipal solid waste is generated in households and enterprises because of everyday human consumption (packages of various consumer goods, food waste, etc.). The quantity (volume) of municipal solid waste depends on the number of consumers, i.e. population. Everyday consumption depends on consumers' available money for consumption – more money available to spend, more waste is generated from consumption. However, consumers' ability to produce their own food, to feed domestic animals with food residue, and to destroy waste or compost it with no especial efforts or costs (rural and suburban population), contributes to overall lower waste

generated per person. The urban population, higher income population and higher economic development of the country are strongly correlated with waste generated.

In that context, land-owning households depict suburban and rural areas, while the number of households without land depict an urban population. The size of agronomic land in use per county is an additional indicator for the rural population, however, the correlations result is biased: positive, negative and/or not significant correlation for the different analysis group (21 County- all counties, 20 County - counties without the City of Zagreb, Continental and Coastal counties). To represent the financial potential for consumption and economic development, the number of employees per county and their average monthly and total annual wages per county is used. Tourists also participate in everyday consumption, and Croatia is a popular tourist destination, thus nights spent at a tourist accommodation (e.g. one tourist one day) are used as a variable as well. It is also inspected whether some statistic facts, such as population density,

Corresponding author: Anamarija Grbeš
anamarija.grbes@rgn.hr

road length, the number of towns, municipalities and populated places, can be used as variables and how they correlate with waste generation.

This article aims to:

1. inspect how the variables proposed in the waste generation model hypothesised in **Grbeš et al. (2016)** correlate among each other,
2. group the variables into independent groups and
3. select from each group the most appropriate candidates for the waste generation model.

Waste generation models are used in the planning of waste management systems, specifically for strategic planning, waste collection services, infrastructure, treatment facilities, capacities, and land demand in the context of landfilling. The data from waste models have a direct influence on the collection system in terms of the personnel and truck utilisation and operational costs, as well as on the monitoring of systems, specifically for assessments of the effects of waste policy **Beigl et al. (2008)**. Categorization of waste models based on the analysis of 45 waste generation models as proposed by **Beigl et al. (2008)** and **Salhofer (2001)** based on sample unit, waste stream type, independent variable, and modelling method is shown in subsequent text. The sample units are based on administrative units ('districts'), specifically the municipality, county, city district or city. The regional scope is preferred. Modelled waste streams in analysed studies conceptually are: material streams, collection streams and fractions of households' waste. The material streams include all wastes from the final consumer (input-output analysis) with no considerations on the collection procedure. Collection streams include official waste statistics for total MSW, but also single collection stream, sums of recyclables, illegally disposed, etc. Fractions of households' waste refer to analyses of waste composition, sorting analyses. The independent variables used in models focus on product life stage: production and trade related, consumption related and disposal related variables. They identify seven methods of modelling: group comparison, correlation analysis, multiple regression analysis, single regression analysis, input-output analysis, time-series analysis and system dynamics modelling.

2. Methods

First, the regional sampling for the administrative unit 'county' is chosen. The source of data is the official statistics, mainly from the Croatian Environmental Protection Agency (**AZO, 2010-2014**) and the Croatian Bureau of Statistics (**DZS, 2010-2014**), as well as the county's administration official sites (**Ljubic, 2014**). The selected waste stream is total municipal solid waste. Then, the set of variables for pre-selection is chosen based on a draft waste generation model from the previous research (**Grbeš et al. 2016**). The subgroup analysis from same research is maintained as well. The part of the

research in this article focusses on the results of correlation analysis among all the considered variables (listed in chapter 2.1 of this article).

2.1. Correlation analysis

This chapter describes correlation analysis as found in **Bhattacharyya and Johnson (1977)**. Correlation analysis is a statistical method to analyse the strength of the relationship between two random variables, neither of which can be singled out as the cause of the other. In a random sample of n experimental subjects, observations on the variables X and Y are denoted by:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

where each pair has the same bivariate distribution and the different pairs are independent. Sample correlation coefficient, as given in **Equation 1**, equals:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right]}} \quad (1)$$

Where $(X_1, Y_1), \dots, (X_n, Y_n)$ are the n pairs of observations.

Properties of correlation coefficient r :

- r must lie between -1 and +1
- The numerical value of r measures strength of the linear relationship, and the sign of r indicates the direction of the relationship.
- r^2 is the proportion of variability in the y values that is explained by a straight line fitted by the least number of squares
- r remains unchanged if the x values are changed to $ax+b$ and the y values are changed to $cy+d$, where a and c have the same sign.

It should be stressed that the correlation coefficient r measures the strength of the linear relationship. The x and y may be strongly related but if their relationship is curvilinear, the r is zero or close to zero. In such a case, the visual inspection of scatterplot shows the curve shaped pattern. The so-called 'banana' pattern indicates that the sample correlation r is not the most suited to show these relations. If a scatter diagram breaks into two or more clusters, the data sample is not suited for correlation analysis.

When two variables correlate, it does not imply the causality or direct relationship between them. Often it implies the existence of a third ('*lurking*') variable that causes the x and y to vary in the same direction, while the x and y may in reality be unrelated or oppositely related. In such a case, when correlation is caused by a third, unknown variable, the correlation is called a '*spurious*' correlation.

The probability that two variables correlate by chance is determined in a significance test of two mutually ex-

clusive hypotheses: null hypothesis $r=0$ and alternative hypothesis $r < 0$. First, the significance level is chosen, commonly $p=0.05$ (odds that the correlation is a chance occurrence is no more than 5 out of 100). Then a decision on the nature of the alternative hypothesis: one tailed ($r < 0$ or $r > 0$) or two tailed (the hypothesis does not specify direction) is made and the degrees of freedom (df) are calculated ($df=n-1$ or $df=n-2$). Then the statistical tables are used for comparison with critical values of r . If the correlation is in the “safe” interval (smaller than a negative critical value or higher than a positive critical value, the significance level is 0.05 (Trochim, 2006).

2.2. Input data for the correlation analysis

The sample unit is a county. The analysed variables are:

- County descriptive variables: number of towns, municipalities and populated areas per county, size (area) of the county, population density and length of roads in a county,
- Total population variables: total number of inhabitants, households, and annually registered population in a county
- Rural/urban related: number of households without land, and owning households and area of used agricultural land in a county
- Additional population: tourist visits and stays (number of tourist nights spent in touristic accommodations)
- Economy related: number of employees in legal entities in a county, monthly net wages per inhabitant and total annual income in a county.

The sources of data:

- Croatian Environment Agency waste registry and reports: **AZO (2009-2014)**
- Croatian Bureau of Statistics website (census data and annual reports: **DZS (2010-2014)**)
- County Road Administration (**Ljubic, 2014**).

Table 1 shows the list of administrative units – counties grouped in four analysis groups:

- I. 21 County,
- II. 20 Counties,
- III. Continental Counties,
- IV. Coastal Counties.

3. Results and discussion

This chapter gives the result of the correlation analysis. Four tables of correlation coefficients (**Tables 2-5**) are shown as follows: 21 County-, 20 Counties-, Continental Counties- and Coastal Counties analysis groups. The results are commented and discussed. In **Tables 2-5** only correlations at confidence level 95% ($p < 0.05$) are shown.

3.1. Correlations inside the group of variables

Inside the group that consists of county descriptive variables (A-F): number of towns (A), municipalities (B) and populated areas (C) per county, size or area (D) of the county, population density (E) and length of roads (F), as shown in **Tables 2-5** and **Figure 1**, the number of towns is correlated with municipalities, populated areas and size of the county, while the number of municipali-

Table 1: Analysis groups of the sample units

I.	II.	III.	IV.
21 County	20 Counties	Continental Counties	Coastal Counties
Zagreb Krapina-Zagorje Sisak-Moslavina Karlovac Varaždin Koprivnica-Križevci Bjelovar-Bilogora Primorje-Gorski kotar Lika-Senj Virovitica-Podravina Požega-Slavonia Slavonski Brod - Posavina Zadar Osijek-Baranja Šibenik-Knin Vukovar-Sirmium Split-Dalmatia Istria Dubrovnik-Neretva Međimurje City of Zagreb	Zagreb Krapina-Zagorje Sisak-Moslavina Karlovac Varaždin Koprivnica-Križevci Bjelovar-Bilogora Primorje-Gorski Lika-Senj Virovitica-Podravina Požega-Slavonia Slavonski Brod - Posavina Zadar Osijek-Baranja Šibenik-Knin Vukovar-Sirmium Split-Dalmatia Istria Dubrovnik-Neretva Međimurje	Zagreb Krapina-Zagorje Sisak-Moslavina Karlovac Varaždin Koprivnica-Križevci Bjelovar-Bilogora Virovitica-Podravina Požega-Slavonia Slavonski Brod - Posavina Osijek-Baranja Vukovar-Sirmium Međimurje	Primorje-Gorski Lika-Senj Zadar Šibenik-Knin Split-Dalmatia Istria Dubrovnik-Neretva

Table 2: Correlation coefficients for “21 County” analysis group. at confidence level 95% ($p < 0.05$). A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads, G-Population, H-Households, I-Population registered, J-Households without land, K-Land owning households, L-Used agricultural land, M-Tourist nights, N-Employees in legal entities, O-Monthly net wages, P-Annual income, Q-MSW.

I.	County descriptive						Total population			Rural/urban			Add. Pop.	Economy			MSW
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
A	1	0.59	0.58	0.48					0.23				0.67				0.31
B	0.59	1									0.51		0.39	-0.25	-0.23	-0.3	
C	0.58		1										0.39				
D	0.48			1									0.27	-0.23		-0.24	
E					1	0.71	0.84	0.86	0.82	0.88				0.95	0.63	0.96	0.75
F					0.71	1	0.85	0.86	0.85	0.84	0.52		0.23	0.83	0.72	0.82	0.86
G					0.84	0.85	1	1	1	0.98	0.58			0.95	0.7	0.93	0.94
H					0.86	0.86	1	1	0.99	0.99	0.51		0.2	0.97	0.73	0.96	0.94
I	0.23				0.82	0.85	1	0.99	1	0.97	0.57		0.2	0.94	0.71	0.93	0.94
J					0.88	0.84	0.98	0.99	0.97	1			0.22	0.98	0.77	0.97	0.94
K		0.51				0.52	0.58	0.51	0.57		1			0.37		0.33	0.43
L												1			-0.35		-0.22
M	0.67	0.39	0.39	0.27		0.23		0.2	0.2	0.22			1		0.41		0.4
N		-0.25		-0.23	0.95	0.83	0.95	0.97	0.94	0.98	0.37			1	0.73	1	0.9
O		-0.23			0.63	0.72	0.7	0.73	0.71	0.77		-0.35	0.41	0.73	1	0.74	0.77
P		-0.3		-0.24	0.96	0.82	0.93	0.96	0.93	0.97	0.33			1	0.74	1	0.88
Q	0.31				0.75	0.86	0.94	0.94	0.94	0.94	0.43	-0.22	0.4	0.9	0.77	0.88	1

Table 3: Correlation coefficients for “20 Counties” analysis group. at confidence level 95% ($p < 0.05$). A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads, G-Population, H-Households, I-Population registered, J-Households without land, K-Land owning households, L-Used agricultural land, M-Tourist nights, N-Employees in legal entities, O-Monthly net wages, P-Annual income, Q-MSW.

II,	County descriptive						Total population			Rural/urban			Add. Pop.	Economy			MSW
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
A	1	0.53	0.53			0.63	0.85	0.88	0.85	0.89	0.49		0.68	0.9	0.49	0.92	0.86
B	0.53	1			0.52	0.44	0.77	0.75	0.75	0.68	0.7		0.4	0.73		0.71	0.63
C	0.53		1			0.53			0.4				0.38	0.41	0.49	0.44	0.34
D				1	-0.55	0.76		0.45	0.39	0.48			0.27	0.35	0.56	0.39	0.52
E		0.52		-0.55	1				0.37		0.45			0.42	-0.36	0.36	
F	0.63	0.44	0.53	0.76		1	0.65	0.66	0.64	0.61	0.59		0.42	0.62	0.46	0.63	0.7
G	0.85	0.77				0.65	1	0.99	0.99	0.95	0.77		0.42	0.95	0.35	0.95	0.85
H	0.88	0.75		0.45		0.66	0.99	1	0.98	0.98	0.69		0.49	0.97	0.42	0.97	0.88
I	0.85	0.75	0.4	0.39	0.37	0.64	0.99	0.98	1	0.94	0.74		0.44	0.94	0.39	0.94	0.85
J	0.89	0.68		0.48		0.61	0.95	0.98	0.94	1	0.53		0.59	0.96	0.51	0.97	0.9
K	0.49	0.7			0.45	0.59	0.77	0.69	0.74	0.53	1	0.45		0.63		0.59	0.47
L											0.45	1	-0.43		-0.32		
M	0.68	0.4	0.38	0.27		0.42	0.42	0.49	0.44	0.59	-0.03	-0.43	1	0.57	0.62	0.62	0.68
N	0.9	0.73	0.41	0.35	0.42	0.62	0.95	0.97	0.94	0.96	0.63		0.57	1	0.38	1	0.87
O	0.49		0.49	0.56	-0.36	0.46	0.35	0.42	0.39	0.51		-0.32	0.62	0.38	1	0.46	0.55
P	0.92	0.71	0.44	0.39	0.36	0.63	0.95	0.97	0.94	0.97	0.59		0.62	1	0.46	1	0.89
Q	0.86	0.63	0.34	0.52		0.7	0.85	0.88	0.85	0.9	0.47		0.68	0.87	0.55	0.89	1

Table 4: Correlation coefficients for “Continental Counties” analysis group. at confidence level 95% ($p < 0.05$). A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads, G-Population, H-Households, I-Population registered, J-Households without land, K-Land owning households, L-Used agricultural land, M-Tourist nights, N-Employees in legal entities, O-Monthly net wages, P-Annual income, Q-MSW.

III.	County descriptive						Total population				Rural/urban			Add. Pop.	Economy			MSW
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	1		0.65			0.57	0.7	0.68	0.67	0.6	0.77		0.36	0.66	0.31	0.68	0.49	
B		1					0.68	0.67	0.65	0.62	0.65		0.27	0.69		0.67	0.4	
C	0.65		1						0.39				0.5	0.29	0.63	0.36	0.32	
D				1	-0.6	0.79		0.62	0.49	0.66			0.35	0.44	0.7	0.51	0.72	
E					1									0.36	-0.5	0.28		
F	0.57		0.66	0.79		1	0.62	0.66	0.58	0.59	0.69		0.3	0.6	0.57	0.64	0.66	
G	0.7	0.68				0.62	1	0.99	0.98	0.96	0.86	0.57		0.94	0.36	0.96	0.75	
H	0.68	0.67		0.62		0.66	0.99	1	0.96	0.98	0.83	0.63	0.28	0.94	0.41	0.96	0.79	
I	0.67	0.65	0.39	0.49		0.58	0.98	0.96	1	0.93	0.83	0.51		0.88	0.4	0.91	0.72	
J	0.6	0.62		0.66		0.59	0.96	0.98	0.93	1	0.71	0.64	0.27	0.9	0.43	0.93	0.8	
K	0.77	0.65				0.69	0.86	0.83	0.83	0.71	1		0.25	0.84	0.27	0.84	0.58	
L							0.57	0.63	0.51	0.64		1		0.51		0.52	0.51	
M	0.36	0.27	0.5	0.35		0.3				0.27	0.25		1	0.35	0.34	0.37		
N	0.66	0.69	0.29	0.44	0.36	0.6	0.94	0.94	0.88	0.9	0.84	0.51	0.35	1		0.99	0.68	
O	0.31		0.63	0.7	-0.5	0.57	0.36	0.41	0.4	0.43	0.27		0.34		1	0.32	0.42	
P	0.68	0.67	0.36	0.51	0.28	0.64	0.96	0.96	0.91	0.93	0.84	0.52	0.37	0.99	0.32	1	0.71	
Q	0.49	0.4	0.32	0.72		0.66	0.75	0.79	0.72	0.8	0.58	0.51		0.68	0.42	0.71	1	

Table 5: Correlation coefficients for “Coastal Counties” analysis group. at confidence level 95% ($p < 0.05$). A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads, G-Population, H-Households, I-Population registered, J-Households without land, K-Land owning households, L-Used agricultural land, M-Tourist nights, N-Employees in legal entities, O-Monthly net wages, P-Annual income, Q-MSW.

IV.	County descriptive						Total population				Rural/urban			Add. Pop.	Economy			MSW
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	1				0.85		0.96	0.98	0.97	0.99			0.64	0.99	0.39	0.99	0.95	
B		1			0.8		0.85	0.82	0.89	0.76	0.86		0.74	0.87		0.87	0.93	
C			1						0.43				0.95	0.58	0.58	0.6	0.47	
D				1											-0.54			
E	0.85	0.8			1		0.87	0.88	0.87	0.87			0.66	0.9	0.6	0.9	0.89	
F						1			0.67		0.81	0.81	0.4	0.6		0.58	0.69	
G	0.96	0.85			0.87		1	0.99	1	0.96	0.88		0.53	0.98		0.97	0.99	
H	0.98	0.82			0.88		0.99	1	0.99	0.99	0.82		0.58	0.99		0.99	0.98	
I	0.97	0.89	0.43		0.87	0.67	1	0.99	1	0.96	0.88	0.36	0.53	0.98		0.97	0.99	
J	0.99	0.76			0.87		0.96	0.99	0.96	1			0.6	0.99	0.43	0.99	0.94	
K		0.86				0.81	0.88	0.82	0.88		1		0.38	0.78		0.75	0.89	
L						0.81			0.36			1	0.68	0.38		0.38	0.43	
M	0.64	0.74	0.95		0.66	0.4	0.53	0.58	0.53	0.6	0.38	0.68	1	0.63	0.64	0.65	0.57	
N	0.99	0.87	0.58		0.9	0.6	0.98	0.99	0.98	0.99	0.78	0.38	0.63	1	0.41	1	0.97	
O	0.39		0.58	-0.54	0.6					0.43			0.64	0.41	1	0.45		
P	0.99	0.87	0.6		0.9	0.58	0.97	0.99	0.97	0.99	0.75	0.38	0.65	1	0.45	1	0.96	
Q	0.95	0.93	0.47		0.89	0.69	0.99	0.98	0.99	0.94	0.89	0.43	0.57	0.97		0.96	1	

ties, populated areas and size of the county are not correlated between each other. Population density correlates with the number of towns, municipalities, size of the county (negative) and length of roads in a county. The

length of roads in a county correlates with every variable in this group.

With respect to the correlation with municipal solid waste generation, the strongest correlation are with

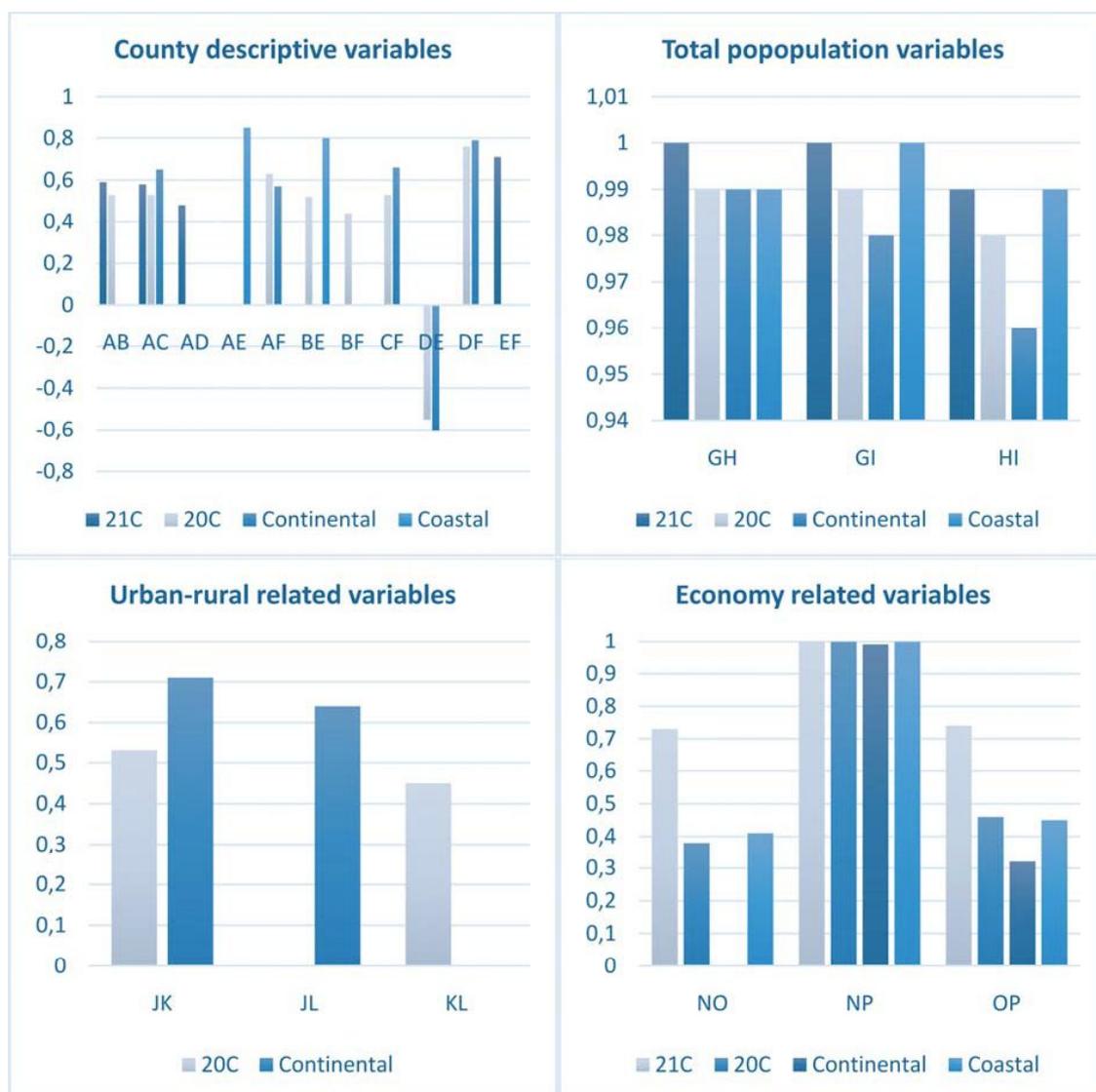


Figure 1. Graphical representation of correlations inside the group of variables: county descriptive, total population, urban-rural related and economy related variables. A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads, G-Population, H-Households, I-Population registered, J-Households without land, K-Land owning households, L-Used agricultural land, M-Tourist nights, N-Employees in legal entities, O-Monthly net wages, P-Annual income. A letter combination stands for the correlation among variables represented by each letter, for example “AB” stands for the correlation between the variables A and B (towns and municipalities), and so on.

towns, population density and length of roads at country level in analysis group I (Q with A,E,F, **Table 2**); towns, roads, and municipalities when the City of Zagreb is excluded in analysis group II (Q with A,F,B, **Table 3**); area of the county and length of roads in continental counties in analysis group III (Q with D,F, **Table 4**); and towns, municipalities, population density, and length of roads in coastal counties in analysis group IV (Q with A,B,E,F, **Table 5**).

These results imply that the length of roads which correlates well with both, variables from the country de-

scriptive group and with waste generation, could be used as the group representative in the waste generation model. In support of that consideration is also the fact that the towns and municipalities can grow for a long time before their number changes, however, their development is evidenced by road length change in current time.

In the group of total population variables (G,H,I): population, households, and population registered, the variables are correlating strongly between each other (≈ 1 , **Tables 2-4, Figure 1**), as well as with the generation of municipal solid waste in analysis groups I ($=0.94$;

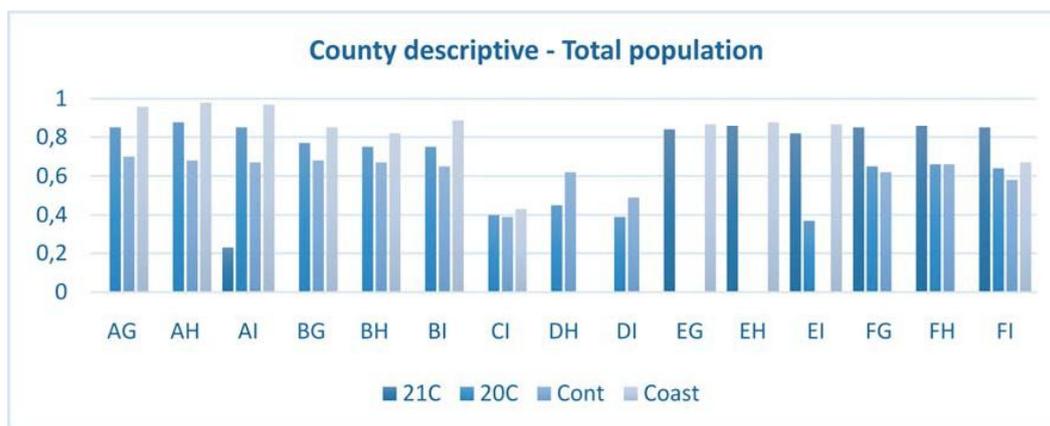


Figure 2. Graphical representation of correlation matrix for multiple variable groups: county descriptive variables (A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads) with total population variables (G-Population, H-Households, I-Population registered). A letter combination stands for the correlation among variables represented by each letter, for example "AG" stands for the correlation between the variables A and G (towns and population), and so on.

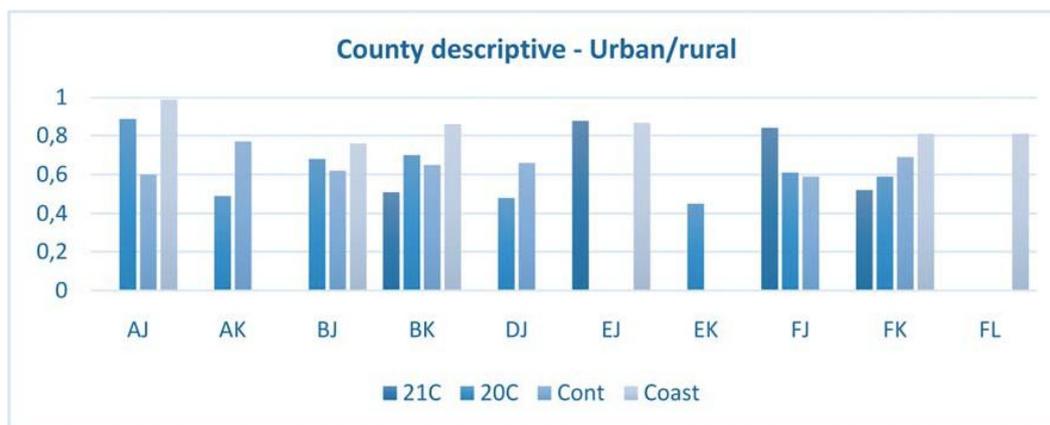


Figure 3. Graphical representation of correlation matrix for multiple variable groups: county descriptive variables (A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads) with urban/rural population related variables (J-Households without land, K-Land owning households, L-Used agricultural land). A letter combination stands for the correlation among variables represented by each letter, for example "AJ" stands for the correlation between the variables A and B (towns and households without land), and so on.

Q with G,H,I in **Table 2**), II (0.85-0.88, Q with G,H,I in **Table 3**) III (0.72-0.79; Q with G,H,I in **Table 4**), and IV (≈ 0.99 , Q with G,H,I in **Table 5**) which indicates that for the waste generation mechanism, from this point of view, any of the population variables could be used.

In the group of urban/rural population related variables (J,K,L): households without land, land owning households and used agricultural land; the correlations between the variables are found only in analysis group II (0.45-0.53; **Table 3**) and III (0.64-0.71; **Table 4**) as shown in **Figure 1**. Land owning households are correlated with area of agricultural land in use (KL) and land owning households (JK) in analysis group II; while in analysis group III the households without land are correlated with land owning households (JK) and used agri-

cultural land (JL). However, lack of correlations in groups I and IV, and lack of strong correlations in group II, implies that these variables might be used together in waste generation model, as independent.

When considering the correlation of urban/rural population related variables with the generation of municipal solid waste, in every analysis group, among urban/rural population related variables, the households without land are the strongest correlated with waste generation (0.94, 0.90, 0.8 and 0.94 in **Tables 2-5** respectively). Land owning households are in the groups I-III which have a much weaker correlation with waste generation than in group IV (0.43, 0.47, 0.58 and 0.89 in **Tables 2-5** respectively). Size of agricultural land in use have shown they are also correlated with waste generation, but in the

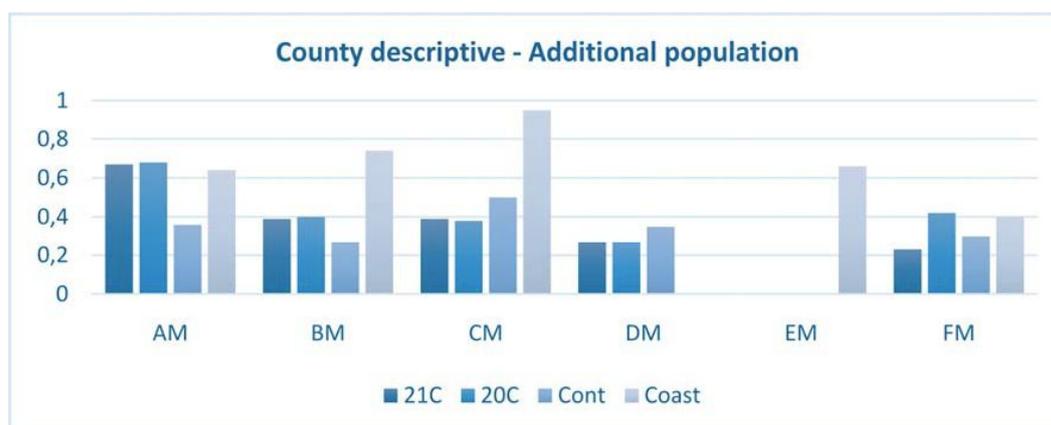


Figure 4. Graphical representation of correlation matrix for multiple variable groups: county descriptive (A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads) with additional population related variable (M-Tourist nights). A letter combination stands for the correlation among variables represented by each letter, for example “AM” stands for the correlation between the variables A and M (Towns and night spent at tourist accommodation), and so on.

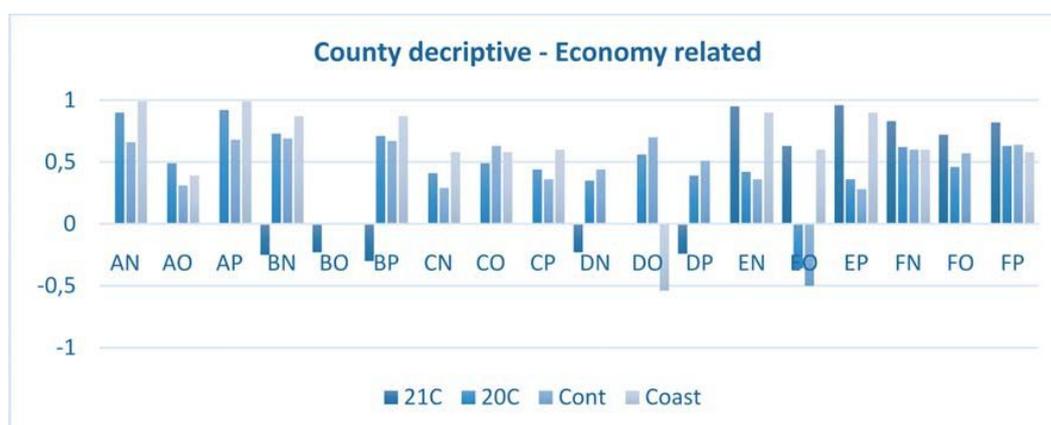


Figure 5. Graphical representation of correlation matrix for multiple variable groups: county descriptive (A-Towns, B-Municipalities, C-Populated area, D-Area, E-Population density, F-Length of roads) with economy related variables (N-Employees in legal entities, O-Monthly net wages, P-Annual income). A letter combination stands for the correlation among variables represented by each letter, for example “AN” stands for the correlation between the variables A and N (towns and employees), and so on.

biased manner (-0.22, 0.51 and 0.43 in **Tables 2, 4, and 5** respectively), that at this point of the research it appears that it is unfit for the waste generation mechanism. However, it might show itself to be useful in future research for the adjustment of the model to regional specifics. Both, the number of households without, and with land, appear relevant for the inclusion in the waste generation mechanism. That is reinforced by the fact that when added up, they give total number of households, which, as previously described, is qualified as a good variable related to the total population number.

The additional population variable – the number of touristic stays at touristic accommodation centres (short Tourist nights, M) although alone in this variable group, have shown to be an important contributor to generated

waste: where touristic activity is occurring, the variable correlates to waste generation (0.4, 0.68, 0, 0.57 in **Tables 2-5** respectively).

The economy related group of variables considered in this research includes the number of employees in legal entities in a county (N), monthly net wages per inhabitant of the county (O) and the total annual income in a county (P). These variables have shown correlations among each other as shown in **Figure 1**. For the total annual income, it is not a surprise because it is derived from the other two variables.

The strength of correlation of economy related variables with the generation of municipal solid waste (N,O,P with Q) for the number of employees per county is rather strong (0.9, 0.87, 0.68, 0.97, groups I-IV, **Ta-**

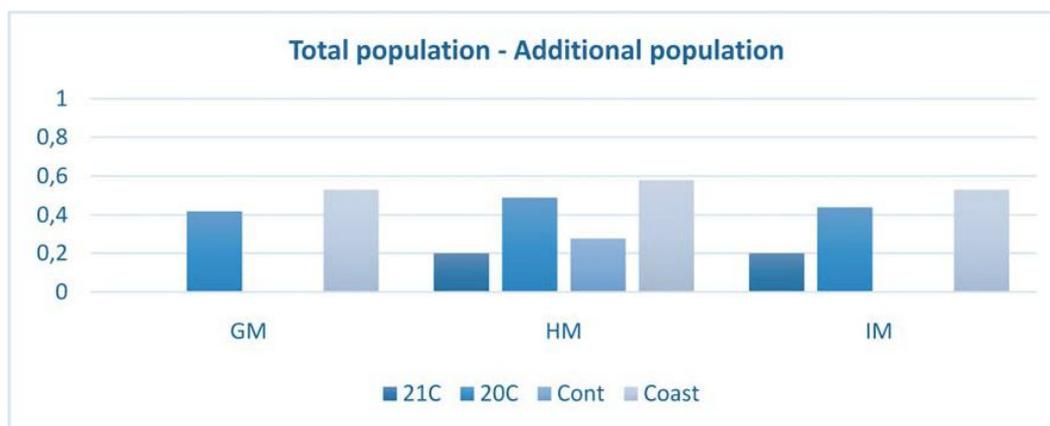


Figure 6. Graphical representation of correlation matrix for multiple variable groups: total population (G-Population, H-Households, I-Population registered) with additional population variable (M-Tourist nights). A letter combination stands for the correlation among variables represented by each letter, for example "GM" stands for the correlation between the variables G and M (population and tourist nights), and so on.

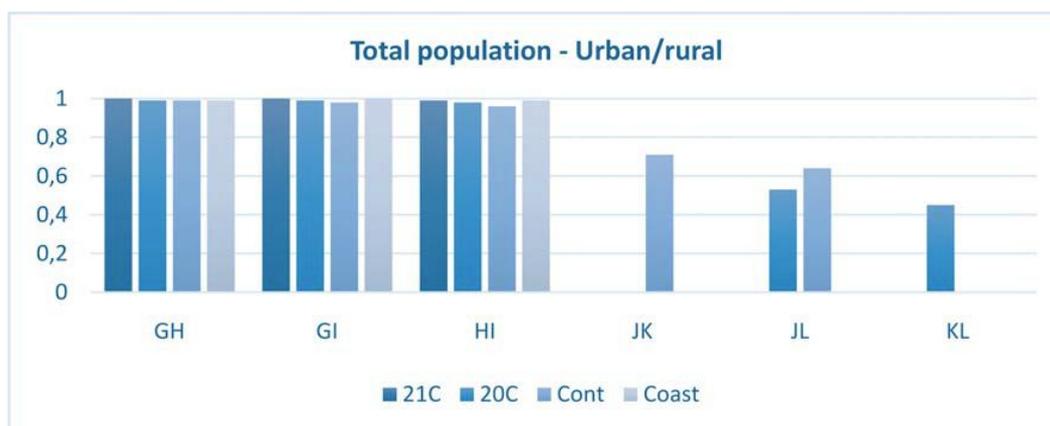


Figure 7. Graphical representation of correlation matrix for multiple variable groups: total population (G-Population, H-Households, I-Population registered) with urban/rural population variables (J-Households without land, K-Land owning households, L-Used agricultural land). A letter combination stands for the correlation among variables represented by each letter, for example "GH" stands for the correlation between the variables G and H (population and households without land), and so on.

bles 2-5 respectively) as well as for total annual income of the county (0.88, 0.89, 0.71, 0.96, groups I-IV, **Tables 2-5** respectively). The average monthly net wages per inhabitant of the county is showing a decrease in strength of correlations (0.77, 0.55, 0.42, -, groups I-IV, **Tables 2-5** respectively) with a decrease in average monthly wages or lack of correlations when the wages are not the only (or primary) source of income in a household as discussed in **Grbeš et al. (2016)** where tourism was recognised as an additional source of income.

These results are implying that at the current state in these matters, the total annual income alone, or number of employees and monthly wages together could be used in waste generation modelling, providing that tourism is also included. However, with the intention to build a

model that includes possible changes in structure income, sources of income other than employment in legal entities and related wages should be considered. The model should include total monthly or annual income in a county, comprised of all relevant sources of income.

3.2. Correlations among the groups of variables

The county descriptive group of variables (A-F) is correlated with the total population variables (G,H,I), as shown in **Figure 2**. It is not a surprise because towns, municipalities and populated areas, as well as population density and roads, are indicators of people's existence and activity. The size of the county is apparently independent of the population variables, however, there

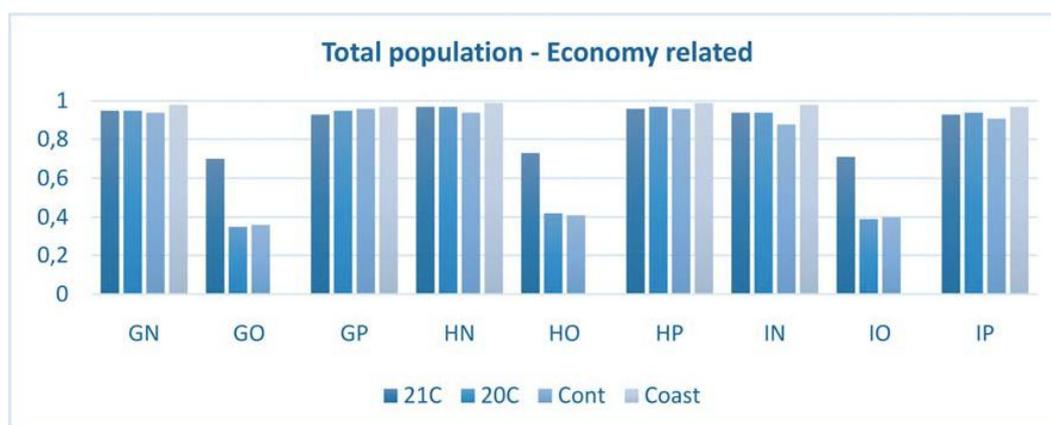


Figure 8. Graphical representation of correlation matrix for multiple variable groups: total population (G-Population, H-Households, I-Population registered) with economy related variables (N-Employees in legal entities, O-Monthly net wages, P-Annual income). A letter combination stands for the correlation among variables represented by each letter, for example “GN” stands for the correlation between the variables G and N (population and employees), and so on.



Figure 9. Graphical representation of correlation matrix for multiple variable groups: urban/rural population related (J-Households without land, K-Land owning households, L-Used agricultural land) with additional population related variable (M-Tourist nights). A letter combination stands for the correlation among variables represented by each letter, for example “JM” stands for the correlation between the variables J and M (households without land and tourist nights), and so on.

are weak correlations of population variables with the size of the county (DH, DI).

County descriptive variables are correlated with urban/rural population variables, specifically the towns, municipalities, area and population density with the households with and without land (A, B, D, E and F with J and K) as shown in **Figure 3**. Variable C (populated area) is lacking the correlations, while variable L (agr. land in use) only correlates with roads (F) in coastal counties.

County descriptive variables towns, municipalities and populated area, are correlated with the additional population variable tourist nights (A, B, C with M) in all four analysis groups. In coastal counties, all county descriptive variables except the size of the county with touristic stays are expressed (see **Figure 4**). The touristic

stays, as well as the roads, are inevitably a part of the existence of the towns, municipalities and populated places, but these correlations in terms of waste modeling can be considered to be a spurious rather than a causal relationship. The invisible variable here would be the reason for touristic visits (sea, history, marketing, policy, etc.)

County descriptive variables (A-F), have shown to be correlated with economy related variables such as employees in legal entities, average monthly wages and total annual income (O, P, Q) in all four analysis groups in a rather interesting manner. Towns are correlated with all three economic variables in all analysis groups, except where the City of Zagreb is excluded (Group I – 21 Counties, or 21C in the graph in **Figure 5**). In analysis

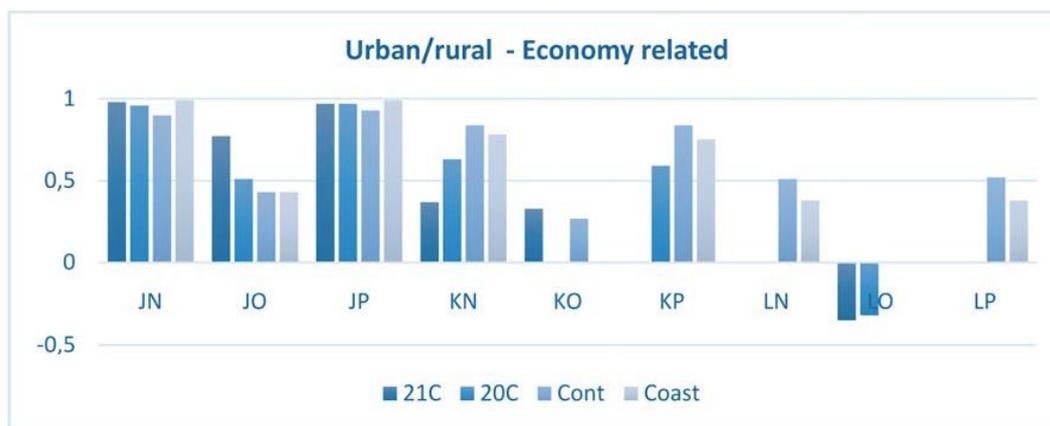


Figure 10. Graphical representation of correlation matrix for multiple variable groups: urban/rural population related (J-Households without land, K-Land owning households, L-Used agricultural land) with economy related variables (N-Employees in legal entities, O-Monthly net wages, P-Annual income). A letter combination stands for the correlation among variables represented by each letter, for example "JN" stands for the correlation between the variables J and N (households without land and employees), and so on.



Figure 11. Graphical representation of correlation matrix for multiple variable groups: additional population related (M-Tourist nights) with economy related variables (N-Employees in legal entities, O-Monthly net wages, P-Annual income). A letter combination stands for the correlation between the variables represented by each letter, for example "MN" stands for the correlation between the variables M and N (tourist nights and employees), and so on.

group I – all counties, the municipalities are negatively correlated with economy related variables suggesting that counties with a higher number of municipalities are correlated with lower employment, wages and income. The data for the City of Zagreb, due to its above average employment and wages, and the fact that it is an administrative unit without municipalities, is causing the change in the inclination of the correlation at a country level. However, when it's excluded, the correlations become positive. A similar situation arises for the size of the county (DN and DP) where, at the country level, the data for the City of Zagreb, due to its special status, causes the correlation graph to change its inclination.

In the coastal counties group, the size of the county becomes negatively correlated with average wages,

showing us that in the coastal part of the Country, the larger counties coincide with lower wages in legal entities. Then a similar situation arises, in analysis groups 20 counties and continental counties where the higher population density apparently coincides with lower wages.

The problem of correlation changing the sign (inclination) in BN, BO, BP and DN is implying that the City of Zagreb should be observed, at an analysis level, as part of the County of Zagreb. The negative DO for coastal counties is probably related to the geomorphology of the coastal area (a large area that is inhabitable). The negative EO for 20 counties and continental counties is probably capturing a different aspect of the reality of the underdevelopment of continental counties.

Total population variables (G,H,I) are somewhat correlated with the additional population variable (tourist stays) as shown in **Figure 6**, but keeping in mind similar considerations as for the Country descriptive-Additional population relations, with perspective to waste modelling the variables can be seen as coincidentally related.

Total population variables have shown to be strongly correlated with households without land (see **Figure 7**) reflecting the fact that most of the population in Croatian counties is urban population. In continental counties, the total number of households is correlated with land owning households (JK) and with land owning (JL and KL). These correlations do not appear to be causal, except for KL -land owning and the size of the land used, which suggests that either K (land owning households) or L (agr. land in use) should be used in waste management modelling.

Total population variables have shown to be strongly correlated with economy variables (see **Figure 8**). The correlation of population with the number of employees or annual income is a matter of mathematic operations – if the employment rate does not differ to high between the counties, the counties with a higher population number will inevitably have more employees. However, the wages do correlate at a lower strength with population, and in coastal counties, they lack the significant correlation of employment with population variables.

Although correlated, from the perspective of waste generation modelling, population variables and economy related variables, could be seen as independent. Similar considerations apply for the relationship of urban/rural population related variables with the additional population variable, urban/rural – economy, additional population – economy, which although they correlate (as shown in **Figures 9, 10 and 11** respectively), they could be considered independent with respect to waste management modelling.

4. Conclusions

- From the group that consists of county descriptive variables (A-F) such as number of towns (A), municipalities (B) and populated areas (C) per county, size or area (D) of the county, population density (E) and length of roads in a county(F); length of roads in a county correlates well with each of these variables and with waste generation, it could be used as the group representative in the waste generation model.
- In the group of total population variables (G,H,I): population, households, and population registered, the variables are correlating strongly between each other as well with the generation of municipal solid waste in analysis groups which indicates that for waste generation mechanism, any of the population variables could be used.

- In the group of urban/rural population related variables (J,K,L): households without land, land owning households and used agricultural land, land owning households and households without land can be used together instead of total households.
- The area of agricultural land in use may be used instead of the land-owning households or for adjustment of the model to regional specifics.
- Additional population variable – the number of tourist stays at touristic accommodation centres (short Tourist nights, M) although alone in this variable group, it is an important contributor to generated waste: where touristic activity is occurring, the variable correlates to waste generation, and hence it should be used in a waste generation model.
- Economy related group of variables considered in this research includes the number of employees in legal entities in a county (N), monthly net wages per inhabitant of the county (O) and total annual income in a county (P). This research is implying that at the current state in these matters, the total annual income alone, or the number of employees and monthly wages together could be used in waste generation modelling, providing that tourism is also separately included.
- With considerations of future changes in the structure of income, sources of income other than employment in legal entities and related wages should be considered. The model should include total monthly or annual income in a county, comprised of all the relevant sources of income.
- With respect to the potential dependency of the considered groups of variables in the context of municipal solid waste generation modelling, the correlated groups will be considered independent.
- In further MSW generation modelling, it should be revised whether the inclusion of data from the “special” administrative unit the “City of Zagreb” into the administrative unit “County of Zagreb”, to form a sample unit similar to the others, would result in less disturbances in correlations.

5. References

- Beigl, P., Lebersorger, S., Salhofer, S. (2008): Modelling municipal solid waste generation: A review. *Waste Management*, 28, 200-214.
- Bhattacharyya, G.K., Johnson, R.A. (1977): *Statistical Concepts and Methods. Correlation: A Measure of Linear Relationship*. University of Wisconsin. John Wiley and Sons Inc., p. 400-419
- Grbeš, A., Ljubić, I., Veinović, Ž. (2016): Study of correlations of statistical parameters with collected municipal solid waste in Croatia in period 2009–2013. *Conference Proceedings of SEM2016 / Katančić, Zvonimir; Koprivanec, Natalija; Lončarić Božić, Ana; Kušić, Hrvoje; Hrnjak-Murgić, Zlata (ur.)*. University of Zagreb, Faculty

- of Chemical Engineering and Technology, Zagreb. 2016, 51-60
- Salhofer, S. (2001): Kommunale Entsorgungslogistik: Planung, Gestaltung und Bewertung entsorgungslogistischer Systeme für kommunale Abfälle (*Municipal Waste Logistic: Planning, Design and Evaluation of Waste Logistic Systems for Municipal Wastes.*) Erich Schmidt, Berlin. (in German)
- Sircar, R., Ewert, F., Bohn, U. (2003): Ganzheitliche Prognose von Siedlungsabfällen. (*Holistic Prognosis of Municipal Wastes*). Müll und Abfall. 1, 7-11. (in German)
- Trochim, W.M. (2006): The Research Methods Knowledge Base. Correlation. 2nd Edition. URL: <http://www.social-researchmethods.net/kb/statcorr.php> (20/01/2017).
- Governmental documents:**
- AZO, Agencija za zaštitu okoliša (2010): Izvješće o komunalnom otpadu za 2009. godinu. [Croatian Environment Agency (2010): Report on Municipal Waste for 2009] URL: <http://www.azo.hr/Izvjesca25> (4/2015)
- AZO, Agencija za zaštitu okoliša (2011): Izvješće o komunalnom otpadu za 2010. godinu. [Croatian Environment Agency (2010): Report on Municipal Waste report for 2010] URL: <http://www.azo.hr/Izvjesca25> (4/2015)
- AZO, Agencija za zaštitu okoliša (2012): Izvješće o komunalnom otpadu za 2011. godinu. [Croatian Environment Agency (2010): Report on Municipal Waste report for 2011] URL: <http://www.azo.hr/Izvjesca25> (4/2015)
- AZO, Agencija za zaštitu okoliša (2013): Izvješće o komunalnom otpadu za 2012. godinu. [Croatian Environment Agency (2010): Report on Municipal Waste for 2012] URL: <http://www.azo.hr/Izvjesca25> (4/2015)
- AZO, Agencija za zaštitu okoliša (2014): Izvješće o komunalnom otpadu za 2013. godinu. [Croatian Environment Agency (2014): Report on Municipal Waste for 2013] URL: <http://www.azo.hr/Izvjesca25> (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2010a): Priopćenje. Zaposleni po područjima djelatnosti I županijama stanje 31. Ožujka 2009. Zagreb XLVI 9.2.4. ISSN 1334-0565 [Croatian Bureau of Statistics (2010): First release. Persons in Paid Employment, by Activities and Counties, Situation as on 31 March 2009] URL: http://www.dzs.hr/default_e.htm
- DZS, Državni zavod za statistiku Republike Hrvatske (2010b): Turizam u 2009. ISSN 1331-341X [Croatian Bureau of Statistics (2010): Tourism, 2009] URL: http://www.dzs.hr/default_e.htm
- DZS, Državni zavod za statistiku Republike Hrvatske (2010c): Zaposlenost i plaće u 2009. ISSN 1331-5048 [Croatian Bureau of Statistics (2010): Employment and Wages, 2009] URL: http://www.dzs.hr/default_e.htm
- DZS, Državni zavod za statistiku Republike Hrvatske (2011a): Popis stanovništva, kućanstava i stanova 2011.godine. Kućanstava i obitelji. 1583. ISSN 1333-1876 [Census of Population, Households and Dwellings 2011, Households and Families 1583. ISSN 1333-1876] URL: http://www.dzs.hr/Hrv/censuses/census2011/results/htm/H01_01_03/H01_01_03.html (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2011b): Površina korištenoga poljoprivrednog zemljišate brojeva i peradi privatnih kućanstava, popis (Area of Used Agricultural Land, Livestock and Poultry Items of Private Households, 2011 Census). URL: http://www.dzs.hr/Hrv/censuses/census2011/results/htm/h02_01_21/H02_01_21.html (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2011c): Priopćenje. Zaposleni po područjima djelatnosti I županijama stanje 31. Ožujka 2010. Zagreb XLVII 9.2.4. ISSN 1330-0350 [Croatian Bureau of Statistics (2011): First release. Persons in Paid Employment, by Activities and Counties, Situation as on 31 March 2010] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2011d): Turizam u 2010. ISSN 1331-341X [Croatian Bureau of Statistics (2011): Tourism, 2010] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2011e): Zaposlenost i plaće u 2010. ISSN 1331-5048 [Croatian Bureau of Statistics (2011): Employment and Wages, 2010] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2011f): Zbirni pregledi: Županije, površina, stanovništvo, gradovi, općine i naselja (Collective Overview, Counties, Population, Towns, Municipalities and Populated Places). URL: http://www.dzs.hr/Hrv/censuses/census2011/results/htm/H02_02/H02_02.html (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2012a): Priopćenje. Zaposleni po područjima djelatnosti i županijama stanje 31. Ožujka 2011. Zagreb XLVIII 9.2.4. ISSN 1330-0350 [Croatian Bureau of Statistics (2012): First release. Persons in Paid Employment, by Activities and Counties, Situation as on 31 March 2010] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2012b): Turizam u 2011. ISSN 1331-341X [Croatian Bureau of Statistics (2012): Tourism, 2011] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2012c): Zaposlenost i plaće u 2011. ISSN 1331-5048 [Croatian Bureau of Statistics (2012): Employment and Wages, 2011] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2013a): Priopćenje. Zaposleni po područjima djelatnosti i županijama stanje 31. Ožujka 2013. Zagreb XLIX 9.2.4. ISSN 1330-0350 [Croatian Bureau of Statistics (2013): First release. Persons in Paid Employment, by Activities and Counties, Situation as on 31 March 2012] URL: http://www.dzs.hr/default_e.htm (4/2015)
- [DZS, Državni zavod za statistiku Republike Hrvatske (2013b): Turizam u 2012. ISSN 1331-341X [Croatian Bu-

- reau of Statistics (2013): Tourism, 2012] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2013c): Zaposlenost i plaće u 2012. ISSN 1331-5048 [Croatian Bureau of Statistics (2013): Employment and Wages, 2012] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2014a): Priopćenje. Zaposleni po područjima djelatnosti i županijama stanje 31. Ožujka 2013. Zagreb XLI 9.2.4. ISSN 1330-0350 [Croatian Bureau of Statistics (2014): First release. Persons in Paid Employment, by Activities and Counties, Situation as on 31 March 2013] URL: http://www.dzs.hr/default_e.htm (4/2015)
- [DZS, Državni zavod za statistiku Republike Hrvatske (2014b): Turizam u 2013. ISSN 1331-341X [Croatian Bureau of Statistics (2014): Tourism, 2013] URL: http://www.dzs.hr/default_e.htm (4/2015)
- DZS, Državni zavod za statistiku Republike Hrvatske (2014c): Zaposlenost i plaće u 2013.. ISSN 1331-5048 [Croatian Bureau of Statistics (2014): Employment and Wages, 2013] URL: http://www.dzs.hr/default_e.htm (4/2015)