Regression Predictive Models Analysis of Municipal Waste

Jiří Křupka, Miloslava Kašparová, and Ivana Knížková

***Abstract*—**This article deals with solid waste modelling for municipalities in the Pardubice region in the Czech Republic. Predictive models work with real data on municipal waste. Municipalities report the stated data into public management reporting systems. This regards „non-sorted" – and „sorted" waste used for recycling. In the first case we talk about mixed (assorted) solid waste, in the second case we talk about plastics and plastic wrappings waste, glass and glass wrapping waste, paper and card board waste. The proposed regression models provide sufficiently high coefficient of determination. Outputs obtained from this analysis may help municipal representatives to define volumes of selected types of waste for the coming period and for the determination of municipal solid waste fee per municipality inhabitant.

***Keywords*—**Waste management, municipal waste, sorted waste, non-sorted waste, predictive models, regression.

I. Introduction

WASTE management (WM) can be defined as: „an activity focused on waste generation prevention, on waste handling and on management of the place where waste is permanently deposited and on the control of the above listed activities" [26]. In the Czech Republic (CR) the WM Council [13], [14] plays a very important role in WM. This Council is as also an advisory body of the Ministry of the Environment [15].

The Region Authority (hereinafter "Authority") deals with waste issues on the regional level. The Authority each year monitors the amount of waste types reported by municipalities belonging to the region. The Authority compiles the Region WM Plan that must be in accord with the CR WM Plan [26]. The joint stock company EKO-KOM is an important part in the process of waste handling. This company represents the country wide system that provides for the utilization, recycling and sorting of packaging waste done at European level. The system is based on the cooperation of cities, municipalities and industrial enterprises. The objective of this system is to make sure that packaging waste is properly sorted, properly transported, finely sorted and then used as a secondary raw material or as source of energy [3], [17].

The CR WM Plan includes also the development of an Integrated WM Plan which could serve as a tool for the fulfilment of the CR Plan on the regional level [10]. It is possible to discuss applicability of an integrated model of municipal WM in CR [8].

In the first past of this article, for the modelling of the WM issues, we have used mixed municipal waste - hereinafter mixed waste (MW) data [1], [26], MW means waste volume without a sorted waste. Municipalities charge a fee for the MW generation, this type of waste is generated mainly by physical entities activities. Municipalities use the proceeds from this collected fee for paying for the "removal" of this type of waste from the municipality [1], [26]. This type of waste (MW), but also other types of waste (e.g. plastics, plastic wrappings, glass and glass wrapping, paper and cardboard) can be further handled (separated) or processed [26]. By such handling [26] we understand (gathering, collection, purchase, transportation, transport storage, treatment, recovery and disposal). By such processing we understand waste disposal [12], managed composting [22], incineration [5] and recycling [26]. In the second part of article "handling" municipal waste data were used. It means these following groups: plastics and plastic wrappings waste (PW), glass and glass wrapping waste (GW), paper and cardboard waste (CW).

The importance of WM mathematical modelling increases along with the increasing organization and technology complexity of this field of activity. These models serve as support tools for decision making in companies and public administration and thus they must include also the ideas and predictions of the future development and priorities. With regard to the increasing prices of energy, of raw materials and with regard to externalities generated by e.g. waste disposal on dumps there emerge trends supporting potential maximalization of waste utilization. To reach this objective economic and administration tools can be used. However, to be able to use these tools in an effective way it is essential to have an idea on how materials flow and what is the environmental impact of these flows as well as on the

Manuscript received Jun 8, 2012: Revised version received August 17, 2012. This work was supported by the project No. SGFES02 of the Ministry of Education, Youth and Sports of CR with title Research and Development Activities in the area of System Engineering and Informatics at the Faculty of Economics and Administration, University of Pardubice in 2012.

J. Křupka is with the Faculty of Economics and Administration, Institute of System Engineering and Informatics, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (e-mail: Jiri.Krupka@upce.cz).

M. Kašparová is with the Faculty of Economics and Administration, Institute of System Engineering and Informatics, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (e-mail: Miloslava.Kasparova@upce.cz).

I. Knížková is with the Faculty of Economics and Administration, Institute of System Engineering and Informatics, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (e-mail: I.Knizkova@seznam.cz).

economic impacts of the individual measures [16], [24]. A quality of environmental data and information plays important role in the process of environmental modelling [7]. In the WM area mathematical models are used for various purposes [16]:

- Creation of integrated WM systems
- Land and area planning, optimalization of locations of new facilities and forecasting their capacity
- Environmental impact and human health assessment for any specific facility and EIA alternatives selection
- Optimalization of operations on the level of one company (collection, transportation, capacity utilization, and similar).

In the area of WM modelling [20], [23] for instance the following models have been used [16]: Cost Benefit Analysis, Life Cycle Assessment, Life Cycle Costing, Decision Support Tool, logistics optimalization models and other. In the current time logistics models, that deal with the optimization of waste collection and transportation, as well as Life Cycle analysis models become more and more significant. The latter models are used for the evaluation of the impacts of the individual WM alternatives. Also models for MW integrated management are newly created [16]. An Integrated WM system includes the set of sources producing waste (in this case these are municipalities) that are interconnected by road network, the system also includes the set of facilities representing dumps, composting facilities and recycled waste treatment facilities. Optimalization can be done also for these models, however the optimalization itself depends on the facility managers and operators and on regulatory elements. The Regulator uses administrative tools (e.g. operations permissions) and economic tools (e.g. fees) in order to achieve e.g. minimum environmental impact under socially bearable costs [16].

The objective of this contribution is the proposal and the analysis of selected part of Pardubice region municipalities WM. We strive for the creation of predictive models for the "estimate" of the future volume of waste in a municipality related to the number of inhabitants. We have analyzed a couple of options for the setting of the predictive model input set of data on the basis of regression analysis. The outputs can be used by cities and municipalities governments for decision making on certain issues related to their WM.

II. Problem Formulation

In the course of the recent years integrated WM models have been elaborated. These models on the one hand minimize costs and on the other hand they minimize the impacts on the environment. However, in the CR the optimalization model has not been used for the design of integrated MW handling systems [3], [16].

There is already a number of scientific and professional works that deal with WM modelling. For instance linear regression (LR) has been used [4] and [18], mental modelling has been used [19], artificial neural networks (NNs) [21] and similar. Examples are given in Table 1. In these stated examples [4], [18], [19], [21] the methods have been marked as "possible to use", that means the methods return acceptable outputs that can be somehow interpreted and use in practical life..

Each municipality has to deal with WM issues. The WM issues represent a quite complicated process of matching inhabitants WM services needs with the effective and efficient use of the municipal budgets. Costs related to transportation, handling and depositing municipal waste are usually paid for jointly by the municipality and its citizens [26]. The number of inhabitants then plays a significant role for the volume of municipal waste that must be "disposed" from the municipality („disposed" is used to define the activity of removal of waste by a company that provides for waste removal) and in consequence to this such service is billed by the company to the municipality. Thus it would be beneficial to have a suitable tool for forecasting of waste volume in relation and with regard to the year on year increase of population, that would allow for effectivity in this area. This would also have a major importance for those municipalities which are subjected to the current trend of „satellite developments" when upon completion of the satellite development the number of inhabitants in the municipality may dramatically increase by these new inhabitants.

In quantitative research [12] multidimensional social and human reality is reduced to a limited number of a couple of variables and to a small number of analyzed relations between such variable. Qualitative research is a non-numerical examination and interpretation of social reality the objective of which is to uncover the meaning of the interpreted information.

Data from Pardubice region (NUTS 3) WM have been selected for the purposes of the models development. The Pardubice region is thanks to its land area of 4519 km^2 the fifth largest region of the CR. It consists of four districts – Chrudim, Pardubice, Svitavy and Ústí nad Orlicí. This region had (as of December 31, 2011) 517,164 inhabitants that live in 451 municipalities [2], [10].

Based on our discussions with a couple of representatives from Pardubice region municipalities it is possible to state that planning in the area of WM is done in the traditional way, no modelling is done and in the majority of cases no exact methods to determination of waste quantity (volume) are used. This however causes the situation that municipalities only react to an already developed situation with a time delay (that means in the following year or after they have received the bill from the disposal company) while the actual situation is already different.

The solution of this area is linked to municipal finance. Financing of WM represents major costs for MW, large volume separated waste and also the disposal of hazardous waste. There are also the costs related to the management of collecting yards and to the liquidation of illegal waste dumps. Company EKOKOM executed a study from which it issues that municipal expenditures for WM continuously increase. In year 2010 those costs increased by 1.2 % compared to year

2009. While in year 2003 the WM costs per one inhabitant were CZK 652.40 in year 2010 those costs increased to CZK 881.90. This fact could not stay unnoticed. For next year (2013) it will be already to increase the maximum fee per a citizen up to CZK 1,000 [25].

The highest part of these costs are the costs of MW transportation and disposal, they represent from 49 % to 80 % of the total costs allocated by municipalities to waste management. In year 2010, costs disbursed on separated waste represented 10.6 – 25.5 % of the overall waste related costs [25].

Table 1 Methods used in WM

Method	Description of method application
LR	Output: WM development forecast Characteristics: 8 subsystems have been researched in detail from the past development in order to describe their representative characteristics. Prognostic system has been developed that is a reflection of the actual municipal WM system used in the CR [4, pp. 28-30]
LR	Output: Household waste production rate Characteristics: Based on a questionnaire survey done in households in Port Harcourt (located in Niger delta plain) the annual household waste production rate related to the number of household members have been enumerated. Regression analysis have been used [18]
Mental models	Output: Recycling processes mental map Characteristics: Between years 2007-2008 a project was undertaken in the Michigan University which by using mental models analysis monitored the actual level of students´ and employees levels of knowledge in the area of recycling processes inside the university. Statistical analyses were used for the final model (the map). The analyses showed any potential gaps in the system. An example of the map is showed in [19]
NNs	Output: Prediction of stability of recycled mass „Reinforced Lightweight Soil" (RLS) Characteristics: NNs were used in research the objective of which was to find out what level of stability would have a mass marked RLS. This mass is the result of mixing various types of waste (e.g. : waste fishing nets and excavated land) for the purpose of their recycling. This is an ecological manner of recycling of such waste. Practical experiments have proven that in this case the NNs can be used for the prediction of stability of the created mass with regard to the ratio in which the mixed materials are represented in the created mass [21]

On the other hand municipalities receive financial means for WM mainly in the form of fees paid by inhabitants (citizens), such fees are nearly three quarters of the WM incomes. Revenues from the sale of secondary raw materials extracted from the waste are another source of revenues. On average the fees collected by municipalities were oscillating very closely under local fees upper limit. In total the total revenues in year 2010 reached the average value of CZK 620.40 per inhabitant per year. The municipality then must pay the difference between the collected fees and the actual real costs from its budget [25].

The above described situation indicates the need for effective use of the WM resources. This effectivity is then also very closely linked to the effectivity of the collection and transportation systems [25]. An example of an ineffectivity is a situation where there is a low yield from separated waste combined with high costs of MW management caused by insufficient number of containers for the separated parts of waste which can become part of the MW.

This situation, but also other difficult situations, could be prevented by using planning tools and modelling on the local level. However this is also conditioned by the will to accept „new methods of work" on the part of municipalities government.

III. Problem Solution

Based on system analysis of the data sources in the Pardubice region it was possible to get a real data set from waste collection and transportation company, from company EKO-KOM, from individual municipalities from the region, from the region administration office or from the WM information system [9].

Finally the real data matrix has been built [10], [11] based on data provided by the Environment department of the Region Authority in the Pardubice region. Then the standard data pre-processing was executed including data cleansing, selection of a suitable set, creation of data dictionary and of new derived variables. The resulting data set included 416 entries from year 2010. The data dictionary for the MW, PW, GW and CW prediction models design is showed in Table 2.

A. Prediction of MW volume

For the MW prediction the regression analysis have been used. A couple of models with various input set settings have been designed. Based on their comparison the most suitable model has been selected. Coefficient of determination R^2 has been selected as the criteria for their evaluation.

In all cases in which linear regression has been used there was first found and evaluated Pearson's correlation coefficient (CC) ρ, which is defined as the rate of mutual linear dependency of random variables x and y [6, p. 42]. After that importance test for the selection correlation coefficient has been executed.

This test tests the hypothesis H_0: $\rho = 0$. The zero value of this coefficient means "non-correlation" (linear independence) of the considered random variables. Positive sign shows direct dependency between y and x, while negative sign shows indirect (dependency)" [6, p. 42].

For the purposes of this calculation the importance coefficient value was chosen to be $\alpha = 0.05$. Variable x represented number of inhabitants and variable y_{MW} represented the volume of MW. We had assumed that these are non-dependent variables from two-dimensional space with normal distribution.

Table 2 Data dictionary of the used data set

Attribute name	Attribute description
IC_obce	Municipality Identification Number
ID_stat	Municipality Identification Number according to Czech Statistical Office
SJ	Type of municipality
M_name	Official name of municipality
x	Number of inhabitants as of January 1, 2011
y_{MW}	Volume of MW (code 200301) – collected to traditional waste bins
y_{MWI}	Volume of MW recalculated per 1 inhabitant (new derived variable)
y_{PW}	Volume of PW (code 200139) – collected to traditional waste bins
y_{PWI}	Volume of PW recalculated per 1 inhabitant (new derived variable)
y_{GW}	Volume of GW (code 200102) – collected to traditional waste bins
y_{GWI}	Volume of GW recalculated per 1 inhabitant (new derived variable)
y_{CW}	Volume of CW (code 200101) – collected to traditional waste bins
y_{CWI}	Volume of CW recalculated per 1 inhabitant (new derived variable)

where data type and range is defined for the selected attribute (attribute/ data type/ data range) by the following way:
SJ / set / municipality, city and small town; x/ Range / 36-121854; y_{MW} in [t]/ Range/ 0.740000 – 16755.700000; y_{MWI} in [kg]/ Range/ 1.254237 – 1297.300302; y_{PW} in [t]/ Range/ 0.182200 – 763.980000; y_{PWI} in [kg]/ Range/ 0.970281 – 100.597015; y_{GW} in [t]/ Range/ 0.090000 – 925.016000; y_{GWI} in [kg]/ Range/ 0.584416 – 58.323529; y_{CW} in [t]/ Range/ 0.021000 – 889.529000; y_{CWI} in [kg]/ Range/ 0.112245 – 89.477612.

Results are demonstrated in Table 3. Similar values of CC were calculated for the rest of variables.

Table 3 Results for number of inhabitant vs. MW volume selection correlation coefficient

CC	P-value for importance test	Decision on hypothesis H_0
0.9828	0.0	declined

At the beginning the objectives of the model are set, that means to predict (with a sufficient level of accuracy) the volume of MW in relation to the number of inhabitants.

The independent variable x is in this case the number of inhabitants in the individual municipalities. The dependent variable y_{MW} is the volume of MW produced in year 2010 in tons for the individual municipalities. These variables are the inputs into the regression analysis that by means of the method of the smallest square (in case of the linear regression) estimates the parameters of the regression equation which we acquire at the output.

In order to be able to compare these models the indicator R^2 was calculated. In case this index was too low, the base data set was modified – for instance by eliminating "suspicious" records (that is such records on which it could be assumed that their values are outliers) or by reducing the input data sets.

The evaluation of R^2, modification of input data were executed based on expert evaluation.

From the base data set were removed step by step some entries of "suspicious" entries that acquired such high values of waste with regard to the number of inhabitants that it was assumed that with certain probability this could had been caused by some errors in the records.

Ten models were realized. In Table 4 there are demonstrated the results of only seven of them because the interpretation of the remaining three models did not reach any improvement in R^2. From the base data set in the course of the modelling 5 records were removed, these might have been outliers.

Table 4 MW LR analysis results

Model number	R^2	Description
1A	0.965	Data set - not excluded municipalities
2A	0.966	Data set without municipality Horní Třešňovec
3A	0.975	Data set without municipality Horní Třešňovec and city Seč
4A	0.976	Data set without municipality Horní Třešňovec, cities Seč and Chvaletice
5A	0.980	Data set without municipality Horní Třešňovec, cities Seč, Chvaletice and Choceň
6A	0.981	Data set without municipality Horní Třešňovec, cities Seč, Chvaletice, Choceň, Brandýs nad Orlicí
7A	0.928	Data set without statutory town Pardubice

Linear regression equation (1) corresponds to model 1A, (2) model 2A , …, (7) model 7A and defines the predicted value of MW volume $\hat{y}_{MW}$ by the following way:

$$\hat{y}_{MW} = 64.32 + 0.144x \,, \tag{1}$$
$$\hat{y}_{MW} = 64.68 + 0.144x \,, \tag{2}$$
$$\hat{y}_{MW} = 60.28 + 0.144x \,, \tag{3}$$
$$\hat{y}_{MW} = 58.99 + 0.144x \,, \tag{4}$$
$$\hat{y}_{MW} = 56.50 + 0.143x \,, \tag{5}$$
$$\hat{y}_{MW} = 55.78 + 0.143x \,, \tag{6}$$
$$\hat{y}_{MW} = 17.53 + 0.194x \,. \tag{7}$$

Outlier records have been derived from Fig. 1 where horizontal axis represents number of inhabitant and vertical axis represents Volume of MW in tons. Excluded records "points" are marked by pale points. After their exclusion any potential other remote values (see grey points) have been analyzed, in these cases however the determination index value has not improved, thus these values have been preserved in the data set.

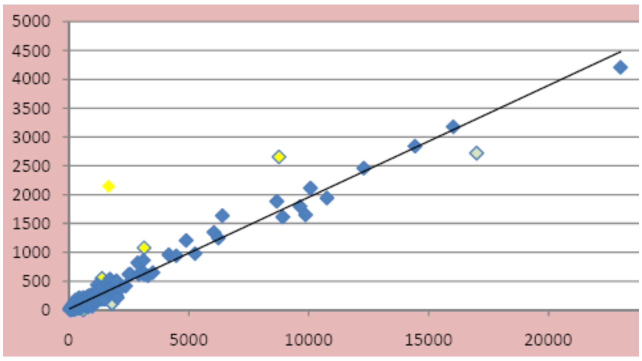


Fig. 1 MW data original structure

In Fig. 2 the MW volume data are transited by the regression line, according to (6) where horizontal axis represents Number of inhabitant x and vertical axis represents Volume of MW y_{MW} in tons. The highest value in the graph is reached by Pardubice statutory town.

While it could seem that compared to others this record cannot be compared to other records (the value of the derived attribute MW per inhabitant y_{MWI} however achieved value comparable with other cities or municipalities, and has not been considered outlier value), its exclusion from the base data set decreased R^2 (see model 7 in Table 4). Mainly due to this phenomenon this record has been retained in the data set for the remaining part of data modelling.

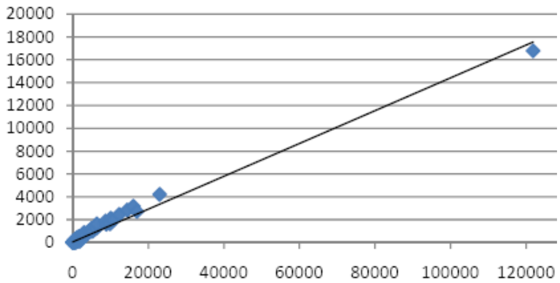


Fig. 2 Resulting LR line for $\hat{y}_{MW}$ (model 6A)

From among the above listed models, under the selected criteria, the linear regression achieved the best results on the base data set modifies by 5 listed towns (model 6A in Table 4). With further removal of values there has not been any further R^2 improvement achieved. The regression equation for the MW volume per inhabitant prediction $\hat{y}_{MW}$ has the following shape stated in (6), for R^2 is 0.981, that means that the equation explains the 98.1 % variation of the explained variable. This is sufficient level and thus this equation could be used in practical life.

B. Prediction of PW volume

The procedure for defining regressive equation for the amount of municipal waste plastic is analogical to the previous procedure. Even the simplified model looks the same only with the exception that in this case the dependent variable is the amount of plastics y_{PW} in tons produced in the territory of the given municipalities in year 2010.

After the elimination of zero values from the original data set 9 records were removed (the data matrix contained 407 municipalities). More than 30 models were implemented (dependent variable: amount of waste y_{PW} in tons, independent variable: number of inhabitants x). Based on the graphic illustration of data and the assessment of the variable values y_{PWI} some records were replaced (Fig. 3) that could be considered remote value. R^2 was continuously increased, however in smaller steps, up to the level of 0.98. The replacement/removal of some objects from the data matrix did not have nearly any effect on the value R^2 (this concerned mainly municipalities with very limited number of inhabitants and with a large production of plastics per 1 inhabitant). The review of observed but not excluded municipalities is mentioned in Table 5. Selected results of the modelling – values of R^2 are stated in Table 6.

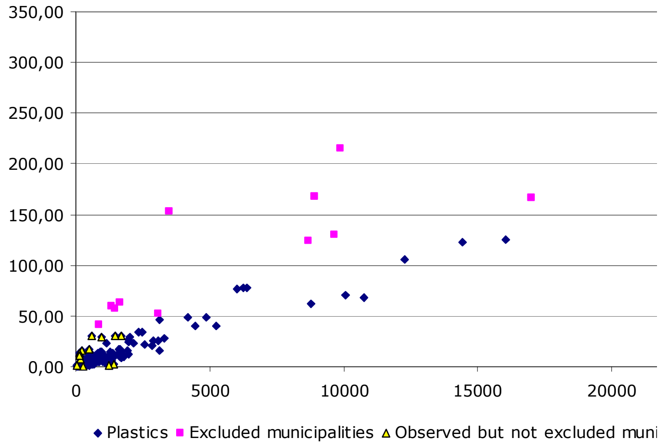


Fig. 3 The original data structure (excluded, observed but not excluded municipalities)

Note: Fig. 3 illustrates only the general/orientation location of the regression line, the value for statutory town Pardubice is missing for the reason of better visibility.

Table 5 Observed but not excluded municipalities

M_name	x	y_{PW} [t]	y_{PWI} [kg]
Ronov nad Doubravou	1707	30.21000	17.70
Býšť	1465	30.72776	20.97
Prachovice	1442	2.60000	1.80
Dlouhá Třebová	1245	1.20800	0.97
Horní Ředice	973	29.30627	30.12
Staré Ždánice	605	30.20000	49.92
Rohovládova Bělá	527	17.23500	32.70
Rozstání	257	0.25567	0.99
Rohoznice	240	15.39000	64.13
Gruna	162	6.89500	42.56
Neratov	134	13.48000	100.60
Újezd u Sezemic	128	10.77000	84.14
Vysoká	36	1.31500	36.53

Linear regression equation (8) corresponds to model 1B, (9) model 2B , …, (19) the best model 12B and defines the predicted value of PW volume $\hat{y}_{PW}$ by the following way:

Table 6 PW LR analysis results

Model number	R^2	Description
1B	0.8871	Data set - not excluded municipalities
2B	0.9028	Data set without Sezemice
3B	0.9346	Data set without Litomyšl, Sezemice, Polička
4B	0.9387	Data set without Litomyšl, Sezemice, Polička, Přelouč
5B	0.9415	Data set without Litomyšl, Sezemice, Polička, Přelouč, Seč
6B	0.9456	Data set without Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun
7B	0.9468	Data set without Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun, Živanice
8B	0.9484	Data set without Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun, Srch
9B	0.9683	Data set without Chrudim, Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun,
10B	0.9716	Data set without Chrudim, Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun, Srch
11B	0.9750	Data set without Chrudim, Litomyšl, Sezemice, Polička, Přelouč, Seč, Lanškroun, Srch, Svitavy
12B	0.9809	Data set without Chrudim, Polička, Litomyšl, Sezemice, Přelouč, Lanškroun, Seč, Srch, Živanice, Staré Hradiště, Svitavy, Třemošnice

$\hat{y}_{PW} = 4.0804 + 0.0068x,$ (8)
$\hat{y}_{PW} = 3.7922 + 0.0068x,$ (9)
$\hat{y}_{PW} = 3.3339 + 0.0067x,$ (10)
$\hat{y}_{PW} = 3.2125 + 0.0067x,$ (11)
$\hat{y}_{PW} = 3.0934 + 0.0067x,$ (12)
$\hat{y}_{PW} = 2.9759 + 0.0066x,$ (13)
$\hat{y}_{PW} = 2.8946 + 0.0066x,$ (14)
$\hat{y}_{PW} = 2.8569 + 0.0066x,$ (15)
$\hat{y}_{PW} = 2.8477 + 0.0065x,$ (16)
$\hat{y}_{PW} = 2.7276 + 0.0065x,$ (17)
$\hat{y}_{PW} = 2.6533 + 0.0064x,$ (18)
$\hat{y}_{PW} = 2.3855 + 0.0064x.$ (19)

When modelling (similar as in the previous subchapter A) the municipality Seč was again excluded. This record shows high values of the selected waste types. This can be caused by the fact that this area is a highly touristy area, in particular in summer months. Containers are used here for waste that is collected separately (this knowledge is based on a query). However, we cannot also exclude the possibility of an error in entering the data to Pardubice Regional Office database

The best modelling result (model 12B) was obtained by excluding municipalities stated in Table 7. 12 municipalities were excluded. Nevertheless it could be possible that some further changes and amendments would bring about a higher value of R^2, however this change would be in the order of thousandth.

Table 7 Excluded municipalities for model

M_name	x	y_{PW} [t]	y_{PWI} [kg]
Litomyšl	9860	214.75820	21.78
Polička	8915	168.01090	18.85
Přelouč	8665	124.47720	14.37
Sezemice	3493	153.70000	44.00
Třemošnice	3086	52.64338	17.06
Seč	1655	62.74760	37.91
Staré Hradiště	1491	56.91000	38.17
Srch	1331	59.59000	44.77
Živanice	882	41.07000	46.56
Chrudim	22999	295.11600	12.832
Svitavy	17004	166.10520	9.769
Lanškroun	9637	130.13800	13.504

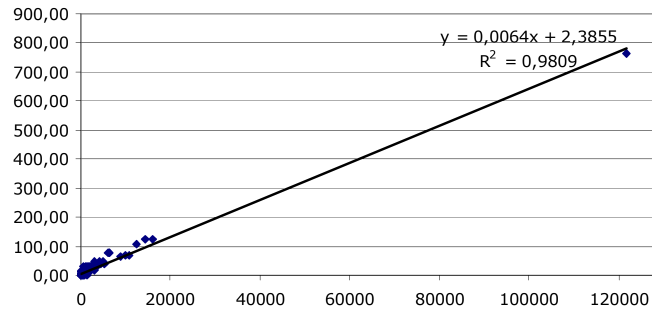


Fig. 4 Resulting LR line for $\hat{y}_{PW}$ (model 12 B)

Resulting regression equation for the estimation of the volume of variable $\hat{y}_{PW}$ in tons depending on the number of inhabitants x (Fig. 4) has the following form (19) and R^2 for this equation obtains value 0.98 which represents 98 % variability of the explained variable. As is the case with variable $\hat{y}_{MW}$ this is quite high value, thus this could be sufficient for practical utilization.

C. *Prediction of GW volume*

The dependent variable y_{GW} here is the total volume of glass in tons produced in the individual municipalities of Pardubice region in year 2010, the independent value is the number of inhabitants x.

The data matrix after removal of zero values included 407 municipalities. There were realized more that 20 models (dependent variable: amount of waste y_{GW} in tons, independent variable: number of inhabitants x). Based on the graphic illustration of data (Fig. 5) and the assessment of values of the variable y_{GWI} those records that could had been consider remote values were removed. Gradually the R^2 was increased. Exclusion of some objects from the data matrix did have hardly any influence on the value of R^2 (this concerned mainly municipalities with a low number of inhabitants and at the same time a high volume of produced glass per inhabitant). The number of excluded and further observed municipalities is stated in Tables 8 and 9.

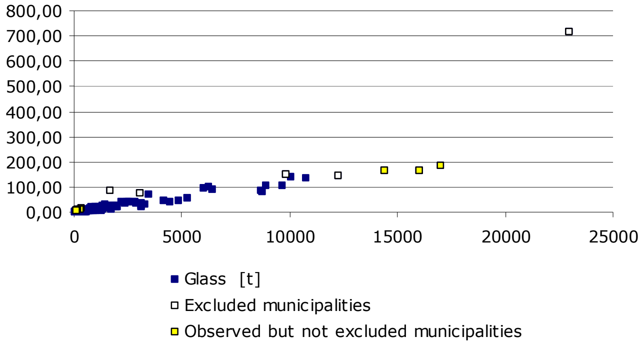


Fig. 5 Original data structure (excluded, observed but not excluded municipalities)

Note: Fig. 5 illustrates orientation location of the regression line, Statutory town Pardubice value is missing for better orientation reasons.

Table 8 Observed but not excluded municipalities

M_name	x	y_{GW} [t]	y_{GWI} [kg]
Svitavy	17004	182.39000	10.73
Ceská Trebová	16028	165.19430	10.31
Ustí nad Orlicí	14428	165.57900	11.48
Vrbatuv Kostelec	333	13.98800	42.01
Mladonovice	313	10.14900	32.42
Cenkovice	177	8.04000	45.42
Neratov	134	6.73700	50.28
Ceské Lhotice	102	5.94900	58.32

Table 9 Excluded municipalities for model

M_name	x	y_{GW} [t]	y_{GWI} [kg]
Chrudim	22999	714.73300	31.08
Vysoké Mýto	12292	143.60700	11.68
Litomyšl	9860	147.33000	14.94
Tremošnice	3086	74.33492	24.09
Sec	1655	86.26620	52.12

The value R^2 was observed under step by step elimination of selected municipalities. The most dramatic change against the model without exclusion of municipalities was recorded when municipality Chrudim was excluded (the difference is 0.1614) and after the exclusion of municipalities Chrudim and Seč (the difference against the model with excluded municipality Chrudim is 0.0043). With excluded municipalities Chrudim and Seč, Třemošnice, Vysoké Mýto a Litomyšl the highest value R^2 = 0.9741 was reached. Fig. 6 shows linear regression line for the resulting model. Selected modelling results are in Table 10.

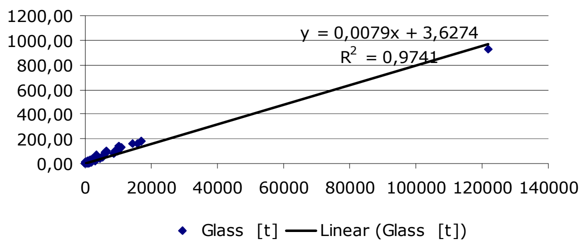


Fig. 6 Resulting LR line for $\hat{y}_{GW}$ (model 6C)

Linear regression equation (20) corresponds to model 1C, (21) model 2C , …, (25) the best model 6C and defines the predicted value of GW volume $\hat{y}_{GW}$ by the following way:

$$y\hat{y}_{GM} = 4.5012 + 0.0087x, \quad (20)$$
$$y_{GM} = 4.0925 + 0.008x, \quad (21)$$
$$y_{GM} = 3.9243 + 0.008x, \quad (22)$$
$$y_{GM} = 3.8173 + 0.008x, \quad (23)$$
$$y_{GM} = 3.7481 + 0.0079x, \quad (24)$$
$$y_{GM} = 3.6274 + 0.0079x. \quad (25)$$

Resulting regression equation for the estimation of the volume of variable $\hat{y}_{GW}$ in tons depending on the number of inhabitants x has the following form (25) and explains the 97.4 % of spread of the explained variable.

Table 10 GW LR analysis results

Model number	R^2	Description
1C	0.8016	Data set - not excluded municipalities
2C	0.963	Data set without Chrudim
3C	0.9673	Data set without Chrudim, Seč
4C	0.9692	Data set without Chrudim, Seč, Třemošnice
5C	0.9704	Data set without Chrudim, Seč, Třemošnice, Vysoké Mýto
6C	0.9741	Data set without Chrudim, Seč, Třemošnice, Vysoké Mýto, Litomyšl

D. Prediction of CW volume

In the Framework of the data set pre-processing in total 101 records with zero value were replaced. The fact that some municipalities show in this type of waste zero value is possibly caused by the fact that in some municipalities there are not paper and card board containers and there are no other collection spots where this waste could be deposited and later transported for further processing or liquidation. Despite this fact the modelling was done.

The used data matrix included in total 315 records. The dependent variable here was y_{CW} in tons for individual municipalities in year 2010, the independent variable here is x in individual municipalities of the Pardubice region. More than 40 models were executed with step by step elimination of the indicated municipalities (the exclusion was based on graphic illustration again and based on recalculated values of the variable y_{CWI}).

The highest R^2 increase was however observed with a model based on a file/data set where town Chrudim was not included (R^2 = 0.7751) that showed 36.94 [kg] of y_{CWI} and then the case where the statutory town Pardubice was excluded (R^2 = 0.7639) that on the other hand showed very small y_{CWI} (7.38 [kg]). With the exclusion of both municipalities we observed the decrease of the observed R^2 indicator against the previous two models (R^2 = 0.7467). After exclusion of Chrudim, Lanškroun, Litomyšl, Letohrad, Žamberk and Svitavy, that showed quite high levels of variable y_{CWI} (model 8D) we observed an increase of the given indicator ($R^2 = 0.9034$) with regression equation (33) for the estimation of the volume of paper and card board $\hat{y}_{CM}$. When only Statutory town Pardubice was excluded, because of its different characteristics , and other municipalities were kept in the given data set (model 2D), the regression equation (27) was obtained for $\hat{y}_{CM}$.

Fig. 7 shows the basic structure of the data including excluded and observed but not excluded municipalities. The overview of excluded municipalities is in Table 11.

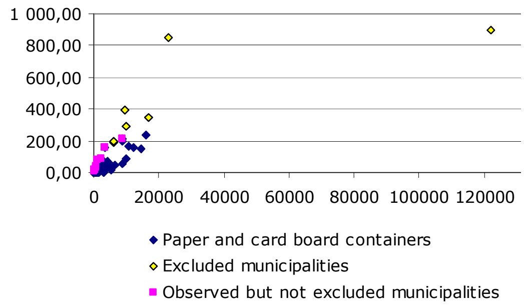


Fig. 7 Original data structure (excluded, observed but not excluded municipalities)

Table 11 Excluded municipalities for model

M_name	x	y_{CW} [t]	y_{CWI} [kg]
Letohrad	6228	189.81500	30.48
Žamberk	6043	198.56100	32.86
Pardubice	121854	899.52900	7.38
Chrudim	22999	849.51800	36.94
Svitavy	17004	348.13670	20.47
Litomyšl	9860	293.18730	29.74
Lanškroun	9637	395.77400	41.07

Table 12 shows the overview of the selected modelling results. Fig. 8 shows the graphic illustration of the model (8D).

Table 12 CW LR analysis results

Model number	R^2	Description
1D	0.6551	Data set - not excluded municipalities
2D	0.7639	Data set without Pardubice
3D	0.7751	Data set without Chrudim
4D	0.8244	Data set without Chrudim, Lanškroun
5D	0.8344	Data set without Chrudim, Lanškroun, Letohrad
6D	0.8603	Data set without Chrudim, Lanškroun, Letohrad, Litomyšl
7D	0.8746	Data set without Chrudim, Lanškroun, Letohrad, Litomyšl, Žamberk
8D	0.9034	Data set without Chrudim, Lanškroun, Letohrad, Litomyšl, Žamberk , Svitavy

Linear regression equation (26) corresponds to model 1D, (27) model 2D , …, (33) the best model 8D and defines the predicted value of CW volume $\hat{y}_{CW}$ by the following way:

$$\hat{y}_{CM} = 6.1848 + 0.0092x\,, \quad (26)$$
$$\hat{y}_{CM} = 8.8475 + 0.0225x\,, \quad (27)$$
$$\hat{y}_{CM} = 5.4381 + 0.0084x\,, \quad (28)$$
$$\hat{y}_{CM} = 4.6807 + 0.0082x\,, \quad (29)$$
$$\hat{y}_{CM} = 4.3092 + 0.0082x\,, \quad (30)$$
$$\hat{y}_{CM} = 3.7967 + 0.0080x\,, \quad (31)$$
$$\hat{y}_{CM} = 3.3855 + 0.0080x\,, \quad (32)$$
$$\hat{y}_{CM} = 2.9947 + 0.0078x\,. \quad (33)$$

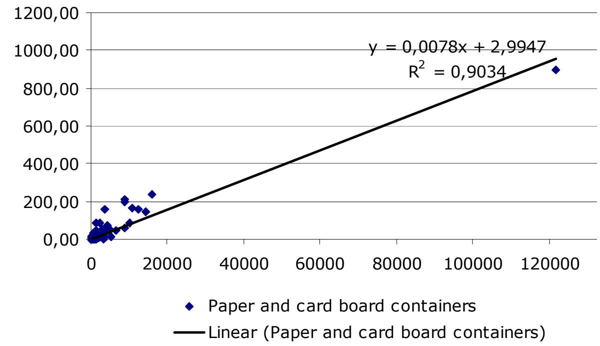


Fig. 8 Resulting LR line for $\hat{y}_{CW}$ (model 8D)

Based on the graphical illustration of the data there was also applied a polynomial trend (this trend curve is used when data fluctuates), that with regard to the fluctuation of data showed to be appropriate. On the base data file a higher value of the observed indicator ($R^2 = 0.8378$) was achieved. Upon exclusion of the statutory town Pardubice there was observed decline to ($R^2 = 0.8267$), when town Chrudim was excluded there was an increase ($R^2 = 0.8829$). Upon exclusion of both municipalities at the same time there was observed decline ($R^2 = 0.7471$) with regard to the basic data set. By exclusion of the above stated municipalities (see Table 11) a model was created (model 8D1) with the highest value of the indicator ($R^2 = 0.941$), regression equation (41) and a corresponding curve on Fig. 9. Table 13 shows selected results of modelling with the utilization of the polynomial trend.

Regression equation (34) corresponds to model 1D1, (35) model 2D1 , …, (41) the best model 8D1 and defines the predicted value of CW volume $\hat{y}_{CW}$ by the following way:

$$\hat{y}_{CM} = -\,9.6832 + 0.0241x - 1\text{E-}07x^2, \quad (34)$$
$$\hat{y}_{CM} = 1.045 + 0.0073x + 1\text{E-}06x^2, \quad (35)$$
$$\hat{y}_{CM} = -\,5.3745 + 0.0189x - 9\text{E-}08x^2, \quad (36)$$
$$\hat{y}_{CM} = -\,5.335 + 0.0189x - 9\text{E-}08x^2, \quad (37)$$
$$\hat{y}_{CM} = -\,4.5939 + 0.0175x - 8\text{E-}08x^2, \quad (38)$$
$$\hat{y}_{CM} = -\,4.5109 + 0.0171x - 8\text{E-}08x^2, \quad (39)$$
$$\hat{y}_{CM} = -\,3.9858 + 0.0161x - 7\text{E-}08x^2, \quad (40)$$
$$\hat{y}_{CM} = -\,2.7693 + 0.0141x - 6\text{E-}08x^2. \quad (41)$$

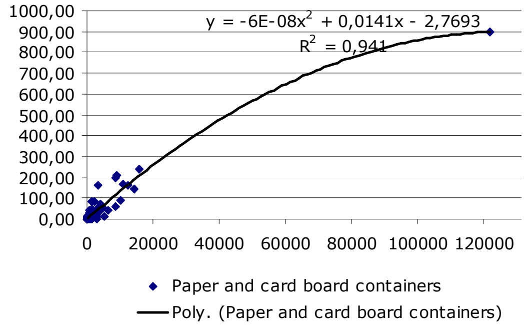


Fig. 9 Resulting regression curve polynomial trend for $\hat{y}_{CW}$ (model 8D1)

Table 13 CW polynomial regression analysis results

Model number	R^2	Description
1D1	0.8378	Data set - not excluded municipalities
2D1	0.8267	Data set without Pardubice
3D1	0.8829	Data set without Chrudim
4D1	0.7471	Data set without Pardubice, Chrudim
5D1	0.9118	Data set without Chrudim, Lanškroun
6D1	0.9161	Data set without Chrudim, Lanškroun, Letohrad
7D1	0.9272	Data set without Chrudim, Lanškroun, Letohrad, Litomyšl
8D1	0.941	Data set without Chrudim, Lanškroun, Letohrad, Litomyšl, Žamberk ,Svitavy

From the above stated it is clear that the polynomial trend represents a better approximation of the given data that reached quite visibly better R^2 values. Upon the exclusion of the statutory town Pardubice record and potentially the exclusion of some other municipalities the data, after graphic illustration, more pointed to a linear trend.

IV. Conclusion

The possibility to forecast MW volume is important especially for negotiation of contract conditions with waste collection and transportation companies. The estimated MW value jointly with other types of waste may serve to define the level of fees to be paid by municipality inhabitants as well as for reservation of sufficient municipal finance for this area from the municipal budget. And last but not least it allows to plan rent (or purchase) of new (larger or smaller) waste bins/containers and in case of separated waste also their number.

The achieved R^2 values than speak about the explanation of 98 % (98 %, 97 % and 90 %) variability of real data set speak about the usability or about the suitability of regression models for MW – the model 6A (PW – the model 12B, GW – the model 6C and CW – the model 8D) prediction on the regional level. It is possible to discuss application the polynomial trend represents

There is a decision making issue arising: to respect reality when modelling, that is to work with the status (real situation) in other municipalities and only exclude statutory town Pardubice, that dramatically differs in the number of inhabitants (there is about 5 times more inhabitants than in Chrudim) and where we can assume that the low value of the given variable is given by the fact that a large number of people does not separate paper and card board, and thus to select as the winning model the model (2D) with value R2 = 0.7639 which represents only 76.39 % of the explained variable, or to keep this model and to exclude other municipalities with higher number of inhabitants where the amount of paper and card board per 1 inhabitant is very high and to use model (8D) with observed indicator R^2 better by 0.1701, that is R^2 = 0.9034 that explains 90.34 % spread of the explained variable in case of the linear regression equation, or potentially to use model (8D1) with polynomial trend that explains up to 94.1 % spread of the volume of paper and card board in the observed municipalities.

In case of the variable Glass, also the polynomial trend was used based on detailed graphical illustration where very good results were obtained. With the resulting model the R^2 values for the polynomial and the linear trend reach high values (the difference is only 0.0182). To predict the volume of glass both models could be possibly used. However due to complexity of the calculation and also due to large rounding in the case of the polynomial trend the approximation under linear regression is more suitable despite the fact that before „cleansing“ it returns lower R^2 values.

In the future it is possible to deal with the issue of prediction by means of NNs, or potentially deal with municipality classification by the volume of MW, PW, GW and CW on the basis of cluster analysis.

References

[1] N. Adamičková, M. Königová, “Tisícovka i víc za svoz odpadu? Poslanci dali obcím volnou ruku” [The thousand and more for waste disposal? Members of the Parliament of the Czech Republic gave municipalities a free hand], *Právo: Nezávislé noviny*. Praha: Borgis, a.s., ISSN 1211-2119. (2012, April 04). 2012. (in Czech) [Online]. Available: http://www.novinky.cz/domaci/258592-tisicovka-i-vic-za-svoz-odpadu-poslanci-dali-obcim-volnou-ruku.html

[2] *Aktuality*. ČSÚ. Český statistický úřad [Czech Statistical Office], (2011, Oct. 15). 2011. (in Czech) [Online]. Available: http://www.czso.cz/xe/redakce.nsf/i/home

[3] R. Baťa, P. Kadlecová, “Modelování ekonomicko-environmentálních dopadů energetického využití papíru” [Modeling of economic and environmental impact of energy generation using of waste paper], *E+M Ekonomie a Management / E&M Economics and Management*, 2011, No. 2, Vol. XIV, pp. 90-98. ISSN 1212-3609. (in Czech)

[4] B. Černík, L. Benešová, M. Doležalová, “Prognóza vývoje odpadového hospodářství ČR” [Waste management forecasting in the Czech republic]. *Odpadové fórum: Waste forum*. 2011, No. 5, pp. 461-471. ISSN 1212-779. (2012, April 01). (in Czech) [Online]. Available: http://www.wasteforum.cz/cisla/WF_5_2010.pdf

[5] *Česká televize* [online]. c1996-2010 [cit. 2010-01-25]. “Nedej se” [Do not be] (2010, Jan. 25). (in Czech) [Online]. Available: http://www.ceskatelevize.cz/program/10267429305-11.02.2010-09:25-2-nedejse.html?online=1

[6] P. Hebák, *Regrese* [Regression]. Vyd. 1. Praha: Vysoká škola ekonomická v Praze, 1998, 138 pp. ISBN 80-707-9909-9. (in Czech)

[7] M. Hejča, J. Hřebíček, "Primary Environmental Data Quality Model: Proposal of a Prototype of Model Concept", *iEMSs 2008: International Congress on Environmental Modelling and Software Integrating Sciences and Information Technology for Environmental Assessment and Decision Making 4th Biennial Meeting of iEMSs*, M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.) International Environmental Modelling and Software Society (iEMSs). 2008. [Online]. Available: http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings

[8] J. Hřebíček, J. Soukopová, "Integrated Model of Municipal Waste Management of the Czech Republic", *International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting*, Ottawa, Canada David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.), International Environmental Modelling and Software Society (iEMSs) 2010. [Online]. Available: http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

[9] Informační systém odpadového hospodářství (ISOH) [Information system of waste management], *CENIA*. (2012, Febr. 26). 2006. (in Czech) [Online]. Available: http://www.cenia.cz/web/www/webpub2nsf/$pid/CENMSFURMQON

[10] I. Knížková, *Modelování vybrané části životního prostředí. Modely odpadového hospodářství obcí Pardubického kraje* [The modelling of selected parts of the environment. Municipalities waste management models in the Pardubice region]. Pardubice, Fakulta ekonomicko-správní, Univerzita Pardubice, 2012, 116 pp. (in Czech)

[11] J. Křupka, I. Knížková, M. Kašparová, "Data Analysis of Municipal Solid Waste Models", *Recent Sesearches in Environmental and Geological Sciences*, *Proc. of the 7th WSEAS International Conference on Energy, & Environment (EE '12)*, Kos Island, Greece, July 14-17, 2012, WSEAS Press, N. Altawell, K. Volkov, C. Matos, P.F. de Arroyabe (Eds.), 2012, pp. 217-222.

[12] M. Kuraš, Optimalizace následné péče o skládky odpadů [Optimization of after-care of landfills], *Odpadové fórum: Waste forum*. 2011, No. 2, pp. 22-23. ISSN 1212-779. (2012, Aug. 03). (in Czech) [Online]. Available: http://www.odpadoveforum.cz/2011/22011.pdf

[13] *Ministerstvo životního prostředí* [Ministry of Environment CR]. "Rozšířené teze rozvoje odpadového hospodářství v ČR" [Advanced thesis of development of waste management in the Czech republic]. 52 pp. (2012, April 04). 2010.(in Czech) [Online]. Available: http://www.mzp.cz/cz/rozvoj_odpadoveho_hospodarstvi

[14] Ministerstvo životního prostředí. [Ministry of Environment CR]. STATUT rady pro odpadové hospodářství. MŽP, 2010, 2010 (2011 Oct. 10). (in Czech) [Online]. Available: http://www.mzp.cz/cz/rada_odpadoveho_hospodarstvi

[15] Ministerstvo životního prostředí: [Ministry of Environment CR]. Historie a poslání MŽP. Ministerstvo životního prostředí [online]. © 2008-2011 (2011 Dec. 28). (in Czech) [Online]. Available: http://www.mzp.cz/cz/ministerstvo

[16] L. Nondek, "Matematické modely nejsou na hraní" [Mathematical models are not playing]. *Odpadové forum: odborný měsičník o odpadech a druhotných surovinách* [Waste Management Forum: Specialised monthly journal o n wastes and secondary materials]:. České ekologické manažerské centrum, 2011, No. 6, pp. 14-16. ISSN 1212-7779. (in Czech)

[17] O společnosti systému EKO-KOM. EKO-KOM, © 2011 (2012, Jan. 28). (in Czech) [Online]. Available: http://www.ekokom.cz/cz/ostatni/o-spolecnosti/systemekokom/o-systemu

[18] D.N. Ogbonna, G.T. Amangabara, T.O, Ekere, "Urban solid waste generation in Port Harcourt metropolis and its implications for waste management", *Management of Environmental Quality: An International Journal*. 2007, Vol. 18, Iss. 1, pp. 71-88.

[19] L. Olson, J. Arvai, L. Thorp, "Mental models research to inform community outreach for a campus recycling program", *International Journal of Sustainability in Higher Education*, 2011, Vol. 12, Iss. 4, pp. 322-337.

[20] I. Para, D. Stanciu, "Timisoara's Waste Sorting Station", *Recent Sesearches in Environmental and Geological Sciences*, *Proc. of the 7th WSEAS International Conference on Energy, & Environment (EE '12)*, Kos Island, Greece, July 14-17, 2012, WSEAS Press, N. Altawell, K. Volkov, C. Matos, P.F. de Arroyabe (Eds.), 2012, pp. 52-56.

[21] H.I. Park, Y.T. Kim, "Prediction of strength of reinforced lightweight soil using an artificial neural network", *Engineering Computations*, 2011, Vol. 28, Iss. 5, pp. 600-615.

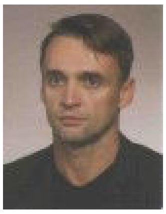
[22] P. Plíva, "Všechno kolem kompostovacích ploch" [Everything around composting sites]. *Odpadové fórum: Waste forum*. 2011, No. 3. pp. 17-19. ISSN 1212-779. (in Czech)

[23] F. Rhoma, Z. Zhang, Y. Luo, B.Noche, "Environmental & Economical Optimization for Municipal Solid Waste Collection Problems, a Modeling and Algorithmic Approach Case Study", *Proc. of the 12th WSEAS International Conference on Mathematical Methods, Computational Techniques And Intelligent Systems (MAMECTIS '10)*, Kantaoui, Sousse, Tunisia, 2010 May 3-6, 2010, pp. 205-211.

[24] B. Sekerka, R. Bata, "Linear Models in Regional and Interregional Modeling", *WSEAS Transactions on Mathematics*, 2011, Vol. 10, Iss. 1, pp. 32-46.

[25] M. Vrbová, Odpady v obcích: Rozdíly mezi náklady a příjmy jsou obrovské. Regionální odpadové centrum Pardubického kraje [online]. ©2006-2010 [cit. 2012-02-26]. (in Czech) [Online]. Available: http://www.odpadypk.cz/index.php/aktuality/clanky/188-odpady-v-obcich-rozdily-mezi-naklady-a-pijmy-jsou-obrovske.html

[26] *Zákon č.185/2001 Sb., o odpadech a o změně některých dalších zákonů* [Law No. 185/2001 about waste]. [online]. (in Czech) [Online]. Available: http://www.mzp.cz/www/platnalegislativa.nsf/d79c09c54250df0dc1256e8900296e32/8fc3e5c15334ab9dc125727b00339581?OpenDocument



Jiří Křupka was born in Prostějov (CR) in 1962. He graduated from the Military Technical University in Liptovský Mikuláš (Slovakia) in 1985. From 1985 till 1990 he worked in the Department of Technical Support System's and Automation in the Air Defense. From 1990 till 2004 he worked as a lecturer, a senior lecturer, and vice-dean for education at the Faculty of Air Defense at the Military Academy in Liptovský Mikuláš. There he finished his doctoral thesis in 1995 and habilitated in 1997. Since 2004 he is working as associated professor and head of Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, CR.

Assoc. Prof. Křupka has published parts of book and a number of papers concerning with fuzzy decision, fuzzy control, case based reasoning, and rough set theory. Nowadays he is focusing on modelling of environmental and social systems.

Miloslava Kašparová was born in Klatovy (CR) in 1976. She is a senior lecturer at Institute of System Engineering and Informatics at Faculty of Economics and Administration at the University of Pardubice. There she received the Master's degree in economy in 2000, and in 2005 she finished the doctoral thesis in the field of informatics in public administration. She has dealt with the modelling of processes in the public administration and nowadays she focuses on an application of selected data mining methods in environmental and social systems.

Dr. Kašparová has published papers concerning an application of data mining methods in various areas.

Ivana Knížková was born in Pardubice (CR) in 1987. She received the Master's degree in System engineering and informatics at the Faculty of Economics and Administration, University of Pardubice in 2012.

Her principal research interests are analysis and modelling of waste management systems in CR, and the area of sustainable development and related fields.