FORECASTING HAZARDOUS WASTE GENERATION USING SHORT DATA SETS: CASE STUDY OF LITHUANIA

Aistė KARPUŠENKAITĖ[1], Gintaras DENAFAS[2], Tomas RUZGAS[3]

Kaunas University of Technology, Kaunas, Lithuania
E-mails: [1]medianaa@gmail.com; [2]gintaras.denafas@ktu.lt; [3]tomas.ruzgas@ktu.lt

Abstract. Due to inefficient waste sorting in primary and secondary waste generation sources Lithuania fails in trying to meet EU requirements for waste management sector regarding the amount of waste flow that reaches landfills. Especially sensitive situation is with hazardous waste, which often are disposed along with municipal solid waste and with it reaches landfills and due to the fact that mechanical and biological treatment plant are only now being established in the biggest cities of Lithuania, landfills becomes a big issue. The main purpose of this research is to find out which mathematical modelling methods could be fitted and if it is possible to forecast annual hazardous waste generation by using automotive, medical and daylight lamps waste generation statistical data. This is part of a research of medical, automotive and daylight lamps waste generation forecasting possibilities. Tests on the performance of artificial neural networks, multiple linear regression, partial least squares, support vector machines and four nonparametric regression methods were conducted on two developed data sets. The best and most promising results in both cases were demonstrated by generalized additives method ($R^2 = 0.99$) and kernel regression ($R^2 = 0.99$).

Keywords: forecasting, hazardous waste, generalized additives, kernel regression, medical waste, daylight lamps, automotive waste.

Introduction

Hazardous waste management and treatment system is considered to be flawed in Lithuania due to ineffective sorting in primal waste generation sources and inefficient waste collection and identification system. Due to these reasons major part of hazardous waste from households reach their end of life in landfills as a constituent of solid waste. Furthermore, unofficially there is an illegal practice in industry, which falsely identifies hazardous waste as other types, usually less toxic or cheaper to manage, waste. It is thought that if hazardous waste collection, management, identification and control system would work in its full efficiency, there would be 1/3 more waste than it is today. According to Denafas, there is a lack of information, not only about the nuances of improving the system but also about the characteristics of the current situation and the reason for not always achieving the desired results.

The possibility of making accurate and reasonable forecasts on various types and total hazardous waste generation is an important aspect in making waste management decisions in short and long terms. Predictions based on openly accessible socio-economic factors provides necessary data for various interested parties from public and business sectors. Such knowledge can influence the choosing

and development of effective hazardous waste utilization methods or the establishment of certain facilities.

The overview of previously conducted research on waste matter in general showed that various mathematical forecasting methods were used to predict solid waste generation. Only a few published research papers that were able to be found using a wide access to scientific journals and subscribed scientific databases uncovered successful application of mathematical prognostic methods applicable for few types of hazardous waste, but not hazardous waste in general. This paper will only overview those research papers that showed the most promising forecasting results.

A study conducted by Abdol *et al.* (2011) presented the approach to unravel the interpolating problem of various structures of artificial neural networks (ANN) for the long-term prediction of solid waste generation (SWG). Results indicate that the multilayer perception approach has more advantages in comparison with traditional methods, like MLR, in predicting the municipal SWG (Abdol *et al.* 2011).

Artificial neural networks (ANNs) and multiple linear regression (MLR) were applied to predict the total rate of medical waste generation and in different types of sharp, infectious and general waste. ANNs showed high performance measure values ($R^2 = 0,99$). Such results were attributed

to the non-linear nature of ANNs in problem solving, which provides the opportunity for relating independent variables to dependent ones non-linearly (Jahandideh *et al.* 2009).

In the Rimaitytė *et al.* (2012) study, the municipal solid waste (MSW) generation was forecasted using time series analysis. The combination of autoregressive integrated moving average (ARIMA) and seasonal exponential smoothing (SES) techniques were found to be the most accurate. This method proved to be very valuable for forecasting the weekly variation of waste generation data ($R^2 > 0,87$) (Rimaitytė *et al.* 2012).

A support vector machine (SVM) as an intelligence tool, combined with partial least squares (PLS) as a feature selection tool were used to produce a weekly prediction of MSW generated in Tehran, Iran. Research showed that PLS-SVM is superior to the SVM model in predictive ability and is also calculation time saving. In addition, results demonstrate that PLS could successfully identify the complex nonlinearity and correlations among input variables and minimize these (Abbasi *et al.* 2013).

In a study conducted by Noori and others (2009), the hybrid of wavelet transform (WT) – adaptive neuro-fuzzy inference system (WT-ANFIS) and wavelet transform-artificial neural network (WT-ANN) was used to predict the weekly waste generation in Tehran. The achieved results indicate the positive effect of input variables pre-processing by WT, which led to the noticeable increase in the accuracy of two model calculations. However, the WT-ANFIS model had better results than the WT-ANN model, because of the smaller uncertainty than the WT-ANN model (Noori *et al.* 2009).

The study of Denafas (2014b) provides results from municipal waste composition research campaigns conducted during the period 2009–2011 in four cities of Eastern European countries. The quantitative estimation of seasonal variation was performed by fitting the collected data into time series forecasting models, such as non-parametric seasonal exponential smoothing, winters additive and winters multiplicative methods (Denafas *et al.* 2014b).

A different approach to predict medical waste generation was conceived by Eleyan *et al.* (2013) and Chaerul *et al.* (2008), who present a new technique using system dynamics modelling to predict generated medical solid waste. Eleyan *et al.* findings indicate that this forecasting approach may cover a variety of possible causative models and track inevitable uncertainties when traditional statistical least-squared regression methods are unable to handle such issues (Eleyan *et al.* 2013). A hospital waste management model, by Chaerul *et al.* determines the interaction among factors in the system using a software package. A simulation

was made to find out when the existing final medical waste disposal sites will reach their capacity (Chaerul *et al.* 2008).

The range of applicable mathematical modelling methods was limited by the scarce official annual data that was obtainable from the Agency of Environmental Protection, due to the institutions data collection system specifics.

The main aim of the research was to investigate the difference of traditional mathematical modelling methods performance when used to forecast medical, daylight lamps, automotive and total hazardous waste generation. Therefore, no ownership of the applied model development can be credited to the authors of this paper.

Materials and methods

Data sets

Since the hazardous waste collection and accounting system, as it is now, was only created in 2004 in Lithuania, it prevents the gaining of large sets of relevant annual data for the research. Furthermore, the current system requires that hazardous waste collection units must be submitted only on an annual data basis, therefore, it was impossible to obtain data about weekly, monthly or quarterly national hazardous waste generation.

After a short analysis of available hazardous waste annual data, two data sets were developed. First data set (with 2007 data) consists of 10 cases with 7 variables, both independent and dependent. Due to the intention to compare the results of this study with foreign countries in the future, data was normalized and adjusted to 1000 residents. Total and hazardous medical waste generation, daylight lamps waste generation, total and hazardous automotive waste generation, total waste generation – all of these variables were used as independent variables in further calculations, which were conducted by the chosen software.

The only feature that separated second (without 2007 data) data set from the first is a lesser number of cases. Due to noticed discrepancy in 2007 total automotive waste
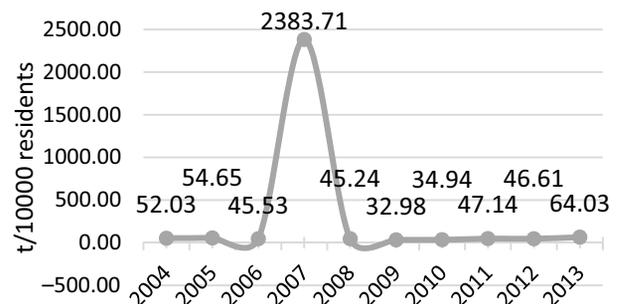

Fig. 1. Data on collected total annual hazardous waste generation 2004–2013 in Lithuania, t / 1000 residents (Agency of Environmental Protection)

358

generation, this case was removed from the data set. The reason for this discrepancy is unclear as the analysis, which was conducted in the previous research by the same authors, of independent variables showed no meaningful fluctuations or reasons for the increased generation of automotive waste.

Mathematical modelling methods

Artificial neural networks (ANN)

The ANN-based models are meant to interact with objects in the real world in the same way that the biological nervous system does. The calibration of ANN-based models is easier than the white-box models as fewer parameters are used in the model development process. For this reason, artificial intelligence techniques using ANN have recently become immensely popular and attractive mathematical tools for both modelling and controlling of several complex environmental processes (Yetilmezsoy *et al.* 2001).

Each neuron in the network is connected to several of its neighbours, with varying coefficients or weights representing the relative influence of the different neuron inputs to other neurons. The weighted sum of the inputs are transferred to the hidden neurons, where it is transformed using an activation function, such as a tangent sigmoid activation function. In turn, the outputs of the hidden neurons act as inputs to the output neuron where they undergo another transformation (Yetilmezsoy *et al.* 2011).

The network training is a process by which the connection weights and biases of the ANN are adapted through a continuous process of simulation by the embedded network's environment. The training function applies the inputs to the new network, calculates the outputs, compares them to the associated targets, and calculates a mean square error. If the error goal is met, or if the maximum number of epochs is reached, the training is stopped and the training function returns the new network and a training record. Otherwise, the training goes through another epoch. During the adaptation phase, the training algorithm receives part of the data (inputs and outputs) and automatically develops the ANN model. After development, the model could generate the appropriate responses for simulations with varying levels of data input. When the learning is complete, the neural network is used for prediction. The primary goal of training is to minimize an error function by searching for a set of connection strengths and biases that causes the ANN to produce outputs equal or close to the targets (Yetilmezsoy *et al.* 2011).

In order to apply the neural network method Neural Tools 6. software was used.

Multiple linear regression (MLR)

Multiple linear regression analysis is an extension of simple linear regression analysis, used to assess the association between two or more independent variables and a single continuous dependent variable. The multiple linear regression equation is as follows (Boston University School... 2013):

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p, \qquad (1)$$

where $\hat{Y}$ is the predicted or expected value of the dependent variable, X_1 through X_p are p distinct independent or predictor variables, b_0 is the value of Y when all of the independent variables (X_1 through X_p) are equal to zero, and b_1 through b_p are the estimated regression coefficients. Each regression coefficient represents the change in Y relative to a one unit change in the respective independent variable. In the multiple regression situation, b_1, for example, is the change in Y relative to a one unit change in X_1, holding all other independent variables constant (Boston University School... 2013).

Multiple linear regression analysis was conducted using SPSS software.

Partial least squares (PLS)

Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear. Note that the emphasis is on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables. For example, PLS is not usually appropriate for screening out factors that have a negligible effect on the response. However, when prediction is the goal and there is no practical need to limit the number of measured factors, PLS can be a useful tool (Tobias 1995).

PLS was developed in the 1960's by Herman Wold and in addition to spectrometric calibration has been applied to monitoring and controlling industrial processes; a large process can easily have hundreds of controllable variables and dozens of outputs (Tobias 1995).

The X- and Y-scores are chosen so that the relationship between successive pairs of scores is as strong as possible. In principle, this is like a robust form of redundancy analysis, seeking directions in the factor space that are associated with high variation in the responses but biasing them toward directions that are accurately predicted (Tobias 1995).

Minitab software was used for the appliance of PLS method.

Support vector machine (SVM)

SVM use an implicit mapping Φ of the input data into a high-dimensional feature space defined by a kernel function, i.e., a function returning the inner product Φ(x), Φ (x$_0$)i between the images of two data points x, x$_0$ in the feature space. The learning then takes place in the feature space, and the data points only appear inside dot products with other points (Karatzoglou *et al.* 2006).

One interesting property of support vector machines and other kernel-based systems is that, once a valid kernel function has been selected, one can practically work in spaces of any dimension without any significant additional computational cost, since feature mapping is never effectively performed. In fact, one does not even need to know which features are being used (Karatzoglou *et al.* 2006).

Another advantage of SVMs and kernel methods is that one can design and use a kernel for a particular problem that could be applied directly to the data without the need for a feature extraction process. This is particularly important in problems where a lot of structure of the data is lost by the feature extraction process (e.g., text processing) (Karatzoglou *et al.* 2006).

Support Vector Machine method was applied using R studio software's package called e1071.

Nonparametric regression

Nonparametric regression relaxes the usual assumption of linearity and enables a more flexible exploration of data, uncovering structure in the data that might otherwise be missed. SAS software was used to apply nonparametric regression methods to the research data sets.

Generalized Additive Models (GAMs) are designed to capitalize on the strengths of ability to fit logistic and Poisson regressions without requiring the problematic steps of *a priori* estimation of response curve shape or a specific parametric response function. They employ a class of equations called "smoothers" or "scatterplot smoothers" that attempt to generalize data into smooth curves by local fitting to subsections of the data. The design and development of smoothers is a very active area of research in statistics, and a broad array of such functions has been developed (Montana State University 2005).

The idea behind GAMs is to "plot" (conceptually, not literally) the value of the dependent variable along a single independent variable, and then to calculate a smooth curve that goes through the data as well as possible, while being parsimonious. The approach generally employed with GAMs is to divide the data into some number of sections, using "knots" at the ends of the sections. Then a low order

polynomial or spline function is fit to the data in the section, with the added constraint that the second derivative of the function at the knots must be the same for both sections sharing that knot. This eliminates kinks in the curve, and ensures that it is smooth and continuous at all points (Montana State University 2005).

Local regression is a method for fitting a smooth curve between two variables, or fitting a smooth surface between an outcome and up to four predictor variables (StatsDirect Statistical Help 2016).

This is a nonparametric method because the linearity assumptions of conventional regression methods have been relaxed. Instead of estimating parameters like m and c in y = mx + c, a nonparametric regression focuses on the fitted curve. The fitted points and their standard errors represent are estimated with respect to the whole curve rather than a particular estimate. So, the overall uncertainty is measured as how well the estimated curve fits the population curve (StatsDirect Statistical Help 2016).

SAS smoothing splines procedure uses the penalized least squares method to fit the data with a flexible model in which the number of effective parameters can be as large as the number of unique design points. Hence, as the sample size increases, the model space also increases, enabling the thin-plate smoothing spline to fit more complicated situations. This allows greater flexibility in the possible form of the regression surface. In particular, this method also makes no assumptions of a parametric form for the model (SAS user guide 2016).

Kernel regression evaluates the possibilities of the relative random variable. The main task is to find nonlinear relations between random values X and Y. The main purpose of kernel regression is to calculate and use fitted weights. The calculated regression curve is finished with the development of forecasts. In the whole data sample, the kernel function is not defined. Smoothing width can be chosen, but a kernel function which would have nonparametric features cannot be chosen. This method has one problem, which occurs if X closes on the limits of the data sample, then the kernel weights become less symmetric and the accuracy of the forecasted values become irrelevant (Liutkevičiūtė 2014).

Time series

The moving-average (MA) method is not only useful in smoothing a time series to see its trend; it is the basic method used in measuring the seasonal fluctuation. In contrast to the least squares method, which expresses the trend in terms of a mathematical equation, the moving-average

method merely smooths the fluctuations in the data. This is accomplished by "moving" the arithmetic mean values through the time series (McGraw Hill Education 2016).

To apply the moving-average method to a time series, the data should follow a fairly linear trend and have a definite rhythmic pattern of fluctuations (McGraw Hill Education 2016).

The basic forecasting equation for single exponential smoothing (SES) is often given as

$$\bar{x}_{t+1} = \alpha x_t + (1-\alpha)\bar{x}_t . \quad (2)$$

It forecasts the value of x at time $t+1$ to be a weighted combination of the observed value at time t and the forecasted value at time t . Although the method is called a smoothing method, it's principally used for short-run forecasting (PennState Eberly College of Science 2016).

The value of α is called the smoothing constant. $\alpha = 0.2$ is a popular default choice of programs. This puts a weight of 0.2 on the most recent observation and a weight of $1 - 0.2 = 0.8$ on the most recent forecast. With a relatively small value of α, the smoothing will be relatively more extensive. With a relatively large value of α, the smoothing is relatively less extensive as more weight will be put on the observed value (PennState Eberly College of Science 2016).

This is simple one-step ahead forecasting method that at first glance seems not to require a model for the data (PennState Eberly College of Science 2016).

Holt (1957) extended simple exponential smoothing to allow forecasting of data with a trend. This method is called Holt's and it involves a forecast equation and two smoothing equations (one for the level and one for the trend) (Hyndman, Athanasopoulus 2014):

Forecast equation
$$\bar{y}_{t+h|t} = l_t + hb_t . \quad (3)$$

Level equation
$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}). \quad (4)$$

Trend equation
$$b_t = \beta * (l_t - l_{t-1}) + (1 - \beta^*)b_{t-1,} \quad (5)$$

where l_t denotes an estimate of the level of the series at time t, b_t denotes an estimate of the trend (slope) of the series at time t, α is the smoothing parameter for the level, $0 \le \alpha \le 1$ and β^* is the smoothing parameter for the trend, $0 \le \beta^* \le 1$ (Hyndman, Athanasopoulus 2014).

StatTools software was used to calculate all three time series mathematical modelling methods.

Performance evaluation

In order to evaluate the performance of the applied model, four statistical indices were used: the root mean square error (RMSE) (see Eq. (6)), coefficient of determination (R^2) (see Eq. (7)), the mean absolute error (MAE) (see Eq. (8)) and the mean absolute percentage error (MAPE) (see Eq. (9)). They are calculated using the outputs given by used models or calculated along with other calculations while running models.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_0 - Y_p\right)^2} ; \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_0 - Y_p\right)^2}{\sum_{i=1}^{n}\left(Y_0 - Y_{ave}\right)^2} ; \quad (7)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_0 - Y_p\right| ; \quad (8)$$

$$MAPE = 100\% \times \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_0 - Y_p}{Y_0}\right| , \quad (9)$$

where Y_0 is the observed value of medical waste generation, Y_{ave} is the average value of observed medical waste generation and Y_p is the predicted value of medical waste generation. The R^2 represents the proportion of the overall variance explained by the model. Additionally, the MAE represents the most absolute and relative meaningful measure of the model's error, and the RMSE is a measure for the remaining measurement variance not explained by the model. MAPE measures prediction accuracy, usually expressed as a percentage, of a forecasting method in statistics.

Results and discussions

Table 1. Applied models performance results by using first (with 2007 data) data set of total hazardous waste generation

| Methods		RMSE	R^2	MAE
ANN	PN/GRN auto-testing	–	–3.881	–
	MLF auto-testing	–	–1280509.49	–
MLR	Enter method	13126.0462	0.94659	127.374
PLS	(after 4 selections)	13135.912	0.946	127.382
SVM	Classification	–	–29614.9	–
	Nonlinear ϱ regression	–	–19134.41	–
Nonparametric regression	Genera-lized additive	132.396	0.999	12.694
	Local regression	–	–	–
	Smoothing splines	245602.189	0.0005	420.356
	Kernel regression	1.068	1	0.938
Time series	MA	521458.4	1	676.011
	SES	302095.539	0.999	267.337
	Holt's	301212.455	0.999	269.872
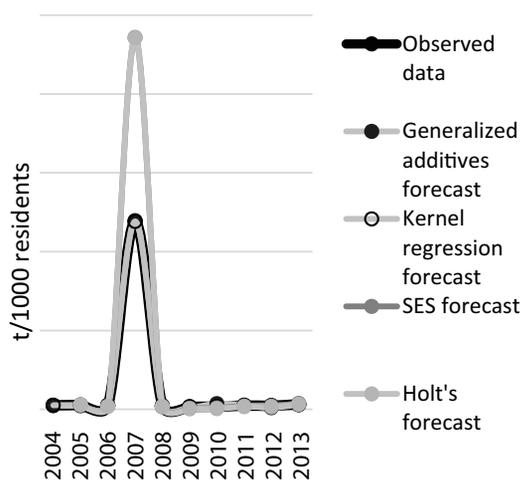
Fig. 2. Observed and most accurately predicted using different mathematical modelling methods on total hazardous waste generation data (first data set with 2007 data)

The analysis of mathematical modelling methods performance results (Table 1) on first data set (with 2007 data) reveals significant fluctuations between MLR, PLS models forecast results and differences comparing to observed data (MLR, PLS, MA, smoothing splines methods). The assumption was made, that this "chaos" was caused by an observation made in 2007 which stood out in whole data set because it was 3722% higher than the next observed case (64 t/1000 residents) (Fig. 2). After the repeated analysis of Lithuania's Agency of Environmental protection data reports it was clear that in 2007 the unusual amount of total automotive and other types of waste (not used as independent variables in this research) were collected and reported. In the theory of mathematics cases like this are called exceptions and are usually excluded from further research in the very first, data set development, stage as it will cause distortions in future results. In the frames of this research, it was chosen to test the capabilities of mathematical methods to work with exceptions in data sets. Used models performance results and forecasted values showed, that kernel regression was the most accurate (RMSE = 1.068, R^2 = 1, MAE = 0.938, MAPE = 1.84%) (Fig. 3). Next best results were demonstrated by generalized additives. The performance of kernel regression and generalized additives methods must be tested outside of currently used data sample, as the accuracy of any regression forecasting ability outside of used sample is unknown until tested (Fig. 3).

The development of second data set and mathematical modelling methods performance on it gave more positive results (Table 2). Although 6 out 13 used models performance results were negative or insufficient, generalized

additives (RMSE = 0.018, R^2 = 999, MAE = 0.152, MAPE = 0.34%), kernel regression (RMSE = 0.242, R^2 = 0.993, MAE = 0.492, MAPE = 0.95%) demonstrated exceptional effectiveness. Next best results were demonstrated by MLR and PLS models (Fig. 4). The performance of kernel regression, generalized additives, PLS and MLR methods must be tested outside of currently used data sample for previously mentioned reasons.
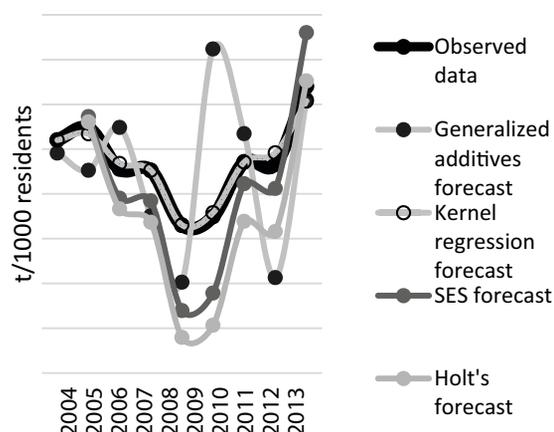

Fig. 3. Observed and most accurately predicted using different mathematical modelling methods on total hazardous waste generation data (first data set without 2007 data and forecast in the figure)

Table 2. Applied models performance results by using second data set (without 2007 data) of total hazardous waste generation

| Methods		RMSE	R^2	MAE
ANN	PN/GRN auto-testing	–	–4.057	–
	MLF auto-testing	–	–22.453	–
MLR	Enter method	1.118	0.969	1.288
PLS	(after 4 selections)	1.111	0.969	1.284
SVM	Classifica-tion	–	–0.803	–
	Nonlinear ϱ regression	–	–0.803	–
Nonpara-metric regression	Genera-lized additive	**0.018**	**0.999**	**0.152**
	Local regression	–	–	–
	Smoothing splines	28.446	0.211	5.950
	Kernel regression	0.242	0.993	0.492
Time series	MA	65.277	0.999	9.445
	SES	43.483	0.999	6.967
	Holt's	42.468	0.999	7.052
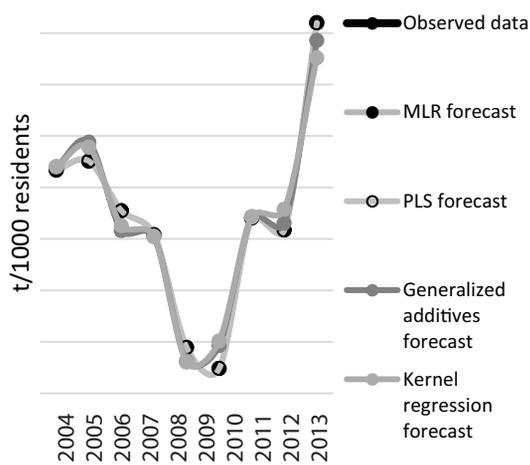
Fig. 4. Observed and most accurately predicted using different mathematical modelling methods on hazardous waste generation data (second data set without 2007 data)

Table 3. Mean absolute percentage errors (MAPE) of models tested with first and second data sets

MAPE, %	MLR	PLS	Generalized additives	Smoothing splines	Kernel regression	Time series MA	Time series SES	Time series Holt's
With 2007	261.29	261.27	28.42	477.99	1.84	935.56	31.03	39.24
Without 2007	2.91	2.90	0.34	13.54	0.95	21.13	15.08	15.51

MAPE represents the mean absolute percentage error of independent variable, in this case hazardous waste generation, value forecasted with different models on different data sets (Table 3).

Conclusions

1. The research described in this paper is highly experimental and its aim was to test how different ANN, PLS, MLR, SVM, nonparametric regression, time series methods will respond to data sets developed from previously conducted similar research where same models were used to forecasting medical, daylight lamps and automotive waste generation. The only difference in them is data sets and relations which bound independent and dependent variables with each other in those data sets. Data sets for this research were developed using previous research dependable variables as independent and the dependable variable was the total annual hazardous waste generation.

2. Results gained in this study can be divided to those showing real and theoretical capabilities of certain mathematical modelling methods. Specifications of artificial neural networks and support vector machine allow demonstrating true model potential due to the fact that developed data sets are divided into a train and test data samples. This could not be done with other nine methods because data sample is already small enough for them to perform well. ANN and SVM models performance results were negative and it raised more serious doubts about the relevance of those models which showed near to perfect forecast accuracy.

3. Therefore, another stage is needed to test MLR, PLS, nonparametric regression and time series methods capability to make accurate forecasts outside of currently used data sets using 2014 and subsequent year data which is in the progress of being formed by Lithuania's Agency of Environmental or does not yet exist.

References

Abbasi, M.; Abdul, M. A.; Omidvar, B.; Baghvand, A. 2013. Forecasting municipal solid waste generation by hybrid support vector machine and partial least square model, *International Journal of Environmental Engineering* 7(1): 27–38.

Abdol, M. A.; Nezhad, M. F.; Sede, R. S.; Behboudian, S. 2011. *Longterm forecasting of solid waste generation by the artificial neural networks*. Wiley Online Library.

Agency of Environmental Protection [online]. 2015 [cited 29 June 2015]. Available from Internet: http://gamta.lt/cms/index?lang=en

Boston University School of Public Health. 2016. *Multiple linear regression analysis* [online]. [cited 23 March 2016]. Available from Internet: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Multivariable/BS704_Multivariable7.html

Chaerul, M.; Tanaka, M.; Shekdar, A. V. 2008. A System Dynamics approach for hospital waste management, *Waste Management* 28: 442–449.
http://dx.doi.org/10.1016/j.wasman.2007.01.007

Denafas, G.; Kazragytė, I.; Martišiūtė, M.; Gustaitienė, R. 2014a. Medical waste generation forecast feasibilies: case study for Lithuania, in *4th International Conference "Industrial and Hazardous Waste Management"*, 2–5 September 2014, Crete, Greece.

Denafas, G.; Ruzgas, T.; Martuzevičius, D.; Shmarin, S.; Hoffmann, M., *et al.* 2014b. Seasonal variation of municipal solid waste generation and composition in four East European cities, *Resources, Conservation and Recycling* 89: 22–33.
http://dx.doi.org/10.1016/j.resconrec.2014.06.001

Eleyan, D.; Al-Khatib, I. A.; Garfield, J. 2013. System dynamics model for hospital waste characterization and generation in developing countries, *Waste Management & Research* 31(10): 986–995. http://dx.doi.org/10.1177/0734242X13490981

Hyndman, R. J.; Athanasopoulos, G. 2014. *Forecasting: principles and practice.* Otexts.

Jahandideh, S.; Jahandideh, S.; Asadabadi, E. B.; Askarian, M.; Movahedi, M. M.; Hosseini, S.; Jahandided, M. 2009. The use of artificial neutral networks and multiple linear re-

gression to predict rate of medical waste generation, *Waste Management* 29: 2874–2879. http://dx.doi.org/10.1016/j.wasman.2009.06.027

Karatzoglou, A.; Meyer, D.; Hornik, K. 2006. Support vector machines in R, *Journal of Statistical Software* 15(9): 1–28. http://dx.doi.org/10.18637/jss.v015.i09

Liutkevičiūtė, V. 2014. *Neparametrinių regresinių metodų lyginamasis tyrimas:* magistro baigiamasis darbas. Kaunas.

McGraw Hill Education. 2016. *Time series and forecasting* 16 chapter. [online], [cited 23 March 2016]. Available from Internet: http://www.mheducation.ca/

Montana State University. 2016. *Modeling species/environment relations with generalized additive models* [online], [cited 23 March 2016]. Available from Internet: http://ecology.msu.montana.edu/labdsv/R/labs/lab5/lab5.html#intro

Noori, R.; Abdoli, M. A.; Farokhnia, A.; Abbasi, M. 2009. Results uncertainty of solid waste generation forecasting by hybrid of wavelet transform-ANFIS and wavelet transform-neutral network, *Expert Systems with Application* 36: 9991–9999. http://dx.doi.org/10.1016/j.eswa.2008.12.035

PennState Eberly College of Science. 2016. *Smoothing time series* [online], [cited 23 March 2016]. Available from Internet: https://onlinecourses.science.psu.edu/stat510/node/70

Rimaitytė, I.; Ruzgas, T.; Denafas, G.; Račys, V.; Martuzevičius, D. 2012. Aplication and evaluation of forecasting methods for municipal solid waste generation in an eastern-European city, *Waste Management & Research* 30(1): 89–98. http://dx.doi.org/10.1177/0734242X10396754

SAS user guide 2015. [online], [cited 29 June 2015]. Available from Internet: http://support.sas.com/documentation/

StatsDirect Statistical Help. 2016. *LOESS curve fitting* [online], [cited 23 March 2016]. Available from Internet: http://www.statsdirect.com/help/Default.htm#nonparametric_methods/loess.htm

Tobias, R. D. 1995. An introduction to partial least squares, in *Proceedings of SAS User Group International Conference*, 2–5 April 1995, Orlando, Florida.

Yetilmezsoy, K.; Ozkaya, B.; Cakmakci, M. 2011. *Artificial intelligence-based prediction models for environmental engineering*. Neural Network World.

PAVOJINGŲJŲ ATLIEKŲ SUSIDARYMO PROGNOZAVIMAS NAUDOJANT TRUMPAS DUOMENŲ IMTIS: LIETUVOS ATVEJIS

A. Karpušenkaitė, G. Denafas, T. Ruzgas

Santrauka

Dėl nepakankamai efektyvaus pirminio ir antrinio atliekų rūšiavimo jų susidarymo šaltiniuose, Lietuva neatitinka ES atliekų tvarkymui keliamų reikalavimų, kurie apibrėžia į sąvartynus patenkančių atliekų srauto procentinę dalį. Pavojingosios atliekos yra ypač daug dėmesio reikalaujantis atliekų tvarkymo sektoriaus aspektas, nes didelė dalis pavojingųjų atliekų kartu su komunalinių atliekų srautu patenka į sąvartynus. Mechaninio ir biologinio apdorojimo įrenginiai, kurie padėtų spręsti šią ir sąvartynų pavojingumo aplinkai problemą, daugelyje didžiųjų šalies miestų tik dabar baigiami įrengti ar pradėti statyti. Pagrindinis šio tyrimo tikslas yra išsiaiškinti, kurie matematinio modeliavimo metodai galėtų būti pritaikyti prognozuojant metinį susidarančių pavojingųjų atliekų kiekį remiantis medicininių, automobilinių ir dienos šviesos lempų atliekų susidarymo duomenimis. Tai yra medicininių, automobilinių ir dienos šviesos lempų atliekų susidarymo prognozavimo galimybių tyrimo dalis. Atliekant tyrimą su dviem duomenų imtimis buvo išbandyti dirbtinių neuronų tinklų, daugialypės tiesinės regresijos, dalinių mažiausių kvadratų, atraminių vektorių, neparametrinės regresijos ir laiko eilučių metodai. Abiejų duomenų imčių atvejais geriausi rezultatai buvo pasiekti taikant bendrųjų adityvų ($R^2 = 0{,}99$) ir branduolinės regresijos ($R^2 = 0{,}99$) metodus.

Reikšminiai žodžiai: prognozavimas, pavojingosios atliekos, bendrieji adityvai, branduolinė regresija, medicininės atliekos, dienos šviesos lempos, automobilinės atliekos.