

Forecasting Municipal Solid waste Generation by Hybrid Support Vector Machine and Partial Least Square Model

Abbasi, M.*, Abduli, M.A., Omidvar, B., Baghvand, A.

Faculty of Environment, University of Tehran, Tehran, Iran

Received 12 Oct. 2011;

Revised 14 Oct. 2012;

Accepted 18 Oct. 2012

ABSTRACT: Forecasting of municipal waste generation is a critical challenge for decision making and planning, because proper planning and operation of a solid waste management system is intensively affected by municipal solid waste (MSW) streams analysis and accurate predictions of solid waste quantities generated. Due to dynamic and complexity of solid waste management system, models by artificial intelligence can be a useful solution of this problem. In this paper, a novel method of Forecasting MSW generation has been proposed. Here, support vector machine (SVM) as an intelligence tool combined with partial least square (PLS) as a feature selection tool was used to weekly prediction of MSW generated in Tehran, Iran. Weekly MSW generated in the period of 2008 to 2011 was used as input data for model learning. Moreover, Monte Carlo method was used to analyze uncertainty of the model results. Model performance evaluated and compared by statistical indices of Relative Mean Errors, Root Mean Squared Errors, Mean Absolute Relative Error and coefficient of determination. Comparison of SVM and PLS-SVM model showed PLS-SVM is superior to SVM model in predictive ability and calculation time saving. Also, results demonstrate which PLS could successfully identify the complex nonlinearity and correlations among input variables and minimize them. The uncertainty analysis also verified that the PLS-SVM model had more robustness than SVM and had a lower sensitivity to change of input variables.

Key words: Municipal Solid Waste, Support Vector Machine, Partial Least Square, Intelligent model

INTRODUCTION

Prediction of solid waste generation is the initial and major important step in planning and operation of MSW management system (Chang and Lin, 1997; H.W. Chen, 2000; Thanh and Matsui, 2011; Arshad *et al.*, 2011; Nouri *et al.*, 2011; Hyun *et al.*, 2011). Nowadays, various models have been proposed to forecast short and long term MSW generation which demonstrate difficulty of problem (Beigl and Lebersorger, 2009; Maqbool *et al.*, 2011; Chen *et al.*, 2011; Safari *et al.*, 2011). Rapid waste generation growth, lack of information, affection of variable and out of control factors on waste generation cause the forecasting to be a complex engineering problem especially in the developing countries (Abdoli *et al.*, 2012; Nada *et al.*, 2012; Rashidi *et al.*, 2012; Shafieiyoun *et al.*, 2012; Mahmoudkhani *et al.*, 2012). In conventional methods, waste generation is characterized by percapita indices with respect to the demographic and socioeconomic factors (Grossman D, 1974; Mukherjee, 1997; Niessen WR, 1972). These models can be applied to the situation in which underlying relationships have not significantly changed over time. It means they don't consider

dynamic properties in the process of MSW generation and consequently the process must be fully characterized. However, there are attempts which use the current information about input variables to forecast the future (Chang and Lin, 1997; Chang *et al.*, 1993). But many developing countries may not have sufficient budget and management task force available to maintain a long-term and large-scale sampling and analysis program. On the other hand, classic statistical models such as most commonly used multiple regression models cannot learn from new data and their precision is poor when inaccurate data are used to find information hidden in data and having a universal approximation (Zhu and ReVelle, 1993 ; Svozil, 1997; Blasco *et al.*, 1998). Based on the advantages of intelligence models, they become popular and are inherently interested in all sciences also solid waste management (Bayar *et al.*, 2009; Dong *et al.*, 2003; Karaca and Ozkaya, 2006). In these models, relation between input and output variables are first found by a learning process and then future outputs will be predicted. These data-driven models without need of complete perception of MSW generation

*Corresponding author E-mail:mar80162015@yahoo.com

process have high ability to model waste generation fluctuations. Noori et al. used artificial neural network (ANN) for forecasting waste generation in Mashhad, Iran (Jalili and Noori, 2008). Model results showed good coincidence between empirical data and predictions. However, ANN may not be able to precisely model non-stationary data if preprocessing of the input and/or output data is not performed. There are many data selection and preprocessing methods which minimize or convert inputs to useful information. Principal component analysis (PCA), wavelet and Gamma test were applied as data selection method in waste generation prediction by Noori et al. (Noori *et al.*, 2009). The innate disadvantage such as over-fitting training, local minimum, difficulty in determination of network architecture and poor generalizing performance remain unsolved and limit the application of the ANN approach into practice. Support vector machine, another intelligence model, developed by Vapnik can provide an effective novel approach to improve both generalization performance of neural networks and achieve global solutions simultaneously (Vapnik, 1995). Recently, the SVM ϵ -insensitive type has been extended to solve non-linear regression estimation and time series prediction (Mukherjee, 1997; Broomhead, 1998; Vapnik, 1997a, b). To provide the acceptable accuracy of prediction and speed calculations, data reduction and variable selection can be applied for preprocessing inputs of SVM. There are many data reduction technique (Zhang *et al.*, 2006; Zhang, 2007; Corcoran, 2003; Wang, 2006). The method selected for the present study is partial least square because it is an unsupervised dimension reduction technique. When the key area of application is multivariate regression, there may be considerable improvement if standardized linear combinations of predictive variables are built to capture as much information in the raw predictive variables as well as in the relation between the predictive and target variables. PLS allows us to achieve this balance and provide an alternate approach to PCA technique (Saikat and Jun, 2008).

Moreover, understanding uncertainty in a model is important to interpreting its results. This becomes especially important if the outcomes to be compared are near one another in magnitude. Literature shows that just a few methods proposed for determining the uncertainty such as bootstrap, sandwich estimator, maximum likelihood, and Bayesian inference which was proposed by Marce and et al (2004). In order to provide the uncertainty associated with the estimation of MSWG, a Monte Carlo simulation was performed herein due to its good performance (2007b; Aqil *et al.*, 2007a). Monte Carlo simulation is a flexible tool for performing uncertainty analysis of data-driven models.

The aims of the study are to develop a novel method for prediction of solid waste generation using hybrid PLS-SVM model and analysis of uncertainty in the model results.

MATERIALS & METHODS

Tehran, the capital of Iran, with population of approximately 13 millions is the most important metropolis and largest commercial and political centre of the country. Daily waste generation in Tehran amounts to over 7500 tons. The total solid waste generated in Tehran during 2004 and 2005 was 2,614,904 and 2,626,519 tons respectively and the total amount of MSW in these years was 2,561,069 and 2,570,988 tons respectively (Damghani *et al.*, 2008). This amount is 2.5-3 times more than other metropolises with the same population. Management of this large quantity of waste needs to prepare a suitable and precise model for forecasting solid waste generation. In this paper, weekly MSW generated in the period of 2008 to 2011 was used to learn the model.

To estimate the amount of generated waste in a city, seasonal patterns are more effective and applicable (Tchobanoglous, 1977). Therefore, weekly time series with 12 time lags (equal to a season) were inputted to the model. So the predicted waste amounts were based on 12 previous time series.

Partial least square was proposed by Wold originally (Wold, 1966). It can separate the information and noise of the predictors or independent variables (X). PLS works by successively extracting factors from both X and responses or dependent variables (Y) such that covariance between the extracted factors is maximized. The technique of PLS is similar to Principal component (PCA). It also produces linear combinations of the original surface parameters. However, PLS and PCA differ in the way they extract the principal directions. PCA ignores the information in Y when building the principal components and PLS produces the directions reflecting the relationship between Y and X . So, PLS results will have more practical meanings.

Assumed X is a matrix with n rows and p columns and Y is a matrix with n rows and q columns. PLS method can work with multivariate response variables (i.e., when Y is an $n \times q$ vector with $q > 1$). However, here it supposed that Y is a single variable i.e., Y is $n \times 1$ and X is $n \times p$. To build a PLS model, it is needed to regress X onto the x-scores (T), which are used to predict the y-scores (U), which in turn are used to predict the responses Y . Thus $X = TP^T + E$ and $Y = UQ^T + F$, where $T_{n \times r}$ is X-scores, $U_{n \times r}$ is Y-scores, $P_{p \times r}$ is X-loadings, $Q_{1 \times r}$ is Y-loadings, $E_{n \times p}$ is X-residuals and $F_{n \times 1}$ is Y-residuals. Decomposition is finalized so as to

maximize covariance between T and U . The solution of this optimization problem can be found below (Lorber *et al.*, 1987; Wold *et al.*, 1984). On eigenvalue decomposition process, the first extracted T and U are of the form $T=X \cdot w$ and $U=Y \cdot c$, where w and c are the eigenvectors corresponding to the first eigenvalue of $X^T Y Y^T X$ and $Y^T X X^T Y$, respectively. It is noted $X^T Y$ denotes the covariance of X and Y . Once the first factors have been extracted the original values of X and Y are deflated as: $X_1=X - t t^T X$ and $Y_1=Y - t t^T Y$. The above process is now repeated to extract the second PLS factors. The process continues until all possible latent variables have been extracted T and U . More details of the PLS procedure can be found in Geladi and Kowalski (Geladi, 1986).

A brief description of the underlying principle of Support vector machine (SVM) is presented here and more details are described in literature (Vapnik, 1995, 1997a, 1998). In support vector machine, the input data is first mapped into high dimensional feature space by the use of kernel function and then linear regression is performed in the feature space. The non-linear feature mapping will allow the treatment of non-linear problems in a linear space. After training on set data SVM can be used to predict the objects whose values are unknown. A regression SVM model estimates the functional dependence of the dependent variable Y on a set of independent variables x . It assumes, like other regression problems, that the relationship between the independent and dependent variables is given by a deterministic function $f(x)$. Considering a set of training data $\{(x_1, y_1), \dots, (x_l, y_l)\}$, where each $x_i \subset R^n$ denotes the input space of the sample and has a corresponding target value $y_i \subset R$ for $i=1, \dots, l$ where l corresponds to the size of the training data (Vapnik, 1995; Müller *et al.*, 1997). The idea of the regression problem is to determine a function that can approximate future values accurately.

$$f(x) = (w \cdot \Phi(x)) + b \tag{1}$$

Where w , b and Φ denotes a non-linear transformation from R^n to high dimensional space. The goal is to find the value of w and b such that values of x can be determined by minimizing the regression risk:

$$R_{reg}(f) = C \sum_{i=0}^{\ell} \Gamma(f(x_i) - y_i) + \frac{1}{2} \|w\|^2 \tag{2}$$

Where $\Gamma(\cdot)$ is a cost function, C is a constant and vector w can be written in terms of data points as:

$$w = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \Phi(x_i) \tag{3}$$

By substituting equation (3) into equation (1), the generic equation can be rewritten as:

$$\begin{aligned} f(x) &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) (\Phi(x_i) \cdot \Phi(x)) + b \\ &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) k(x_i, x) + b \end{aligned} \tag{4}$$

In equation (4) the dot product can be replaced with function $k(x_i, x)$, known as the kernel function. Kernel functions enable dot product to be performed in high-dimensional feature space using low dimensional space data input without knowing the transformation Φ . All kernel functions must satisfy Mercer's condition that corresponds to the inner product of some feature space. The radial basis function (RBF) is commonly used as the kernel for regression:

$$k(x_i, x) = \exp\left\{-\gamma \|x - x_i\|^2\right\} \tag{5}$$

The ε -insensitive loss function is the most widely used cost function (Müller *et al.*, 1997). The function is in the form:

$$\Gamma(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon, & \text{for } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

By solving the quadratic optimization problem in (7), the regression risk in equation (2) and the ε -insensitive loss function (6) can be minimized:

$$\frac{1}{2} \sum_{i,j=1}^{\ell} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) - \sum_{i=1}^{\ell} \alpha_i^* (y_i - \varepsilon) - \alpha_i (y_i + \varepsilon)$$

Subject to:

$$\sum_{i=1}^{\ell} \alpha_i - \alpha_i^* = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \tag{7}$$

The Lagrange multipliers, α_i and α_i^* , represent solutions to the above quadratic problem that act as forces pushing predictions towards target value y_i . Only the non-zero values of the Lagrange multipliers in equation (7) are useful in forecasting the regression

line and are known as support vectors. For all points inside the ε -tube, the Lagrange multipliers equal to zero do not contribute to the regression function. Only if the requirement $|f(x) - y| \geq \varepsilon$ is fulfilled, Lagrange multipliers may be non-zero values and used as support vectors.

The constant C introduced in equation (2) determines penalties to estimation errors. A large C assigns higher penalties to errors so that the regression is trained to minimize error with lower generalization while a small C assigns fewer penalties to errors; this allows the minimization of margin with errors, thus higher generalization ability. If C goes to infinitely large, SVR would not allow the occurrence of any error and result in a complex model, whereas when C goes to zero, the result would tolerate a large amount of errors and the model would be less complex.

Now, the value of w is solved in terms of the Lagrange multipliers. For the variable b , it can be computed by applying Karush-Kuhn-Tucker (KKT) conditions which, in this case, implies that the product of the Lagrange multipliers and constrains has to equal zero:

$$\begin{aligned} \alpha_i(\varepsilon + \zeta_i - y_i + (w, x_i) + b) &= 0 \\ \alpha_i^*(\varepsilon + \zeta_i^* + y_i - (w, x_i) - b) &= 0 \end{aligned} \quad (8)$$

and

$$\begin{aligned} (C - \alpha_i)\zeta_i &= 0 \\ (C - \alpha_i^*)\zeta_i^* &= 0 \end{aligned} \quad (9)$$

Where ζ_i and ζ_i^* are slack variables used to measure errors outside the ε -tube. Since $\alpha_i, \alpha_i^* = 0$ and $\zeta_i, \zeta_i^* = 0$ for $\alpha_i \in (0, C)$, b can be computed as follows:

$$\begin{aligned} b &= y_i - (w, x_i) - \varepsilon \quad \text{for } \alpha_i \in (0, C) \\ b &= y_i - (w, x_i) + \varepsilon \quad \text{for } \alpha_i^* \in (0, C) \end{aligned} \quad (10)$$

Putting it all together, SVM can be used without knowing the transformation.

The Monte Carlo method is just one of many methods for analyzing uncertainty propagation, where the goal is to determine how random variation, lack of knowledge, or error affects the sensitivity, performance, or reliability of the system that is being modeled. Monte Carlo method is a technique that involves repeatedly forming a random vector of parameters from prescribed probability distributions, evaluating the function, and then computing the statistics of the evaluated function.

In this research, uncertainty analysis was performed as follows:

Step 1: 1000 times randomly rearranged the database without replacement since the ratio between the training and validation sets was kept fix. Thus, 1000 weekly MSWG series were generated,

Step 2: 1000 different results for each forecasted weekly MSWG were obtained by SVM and PLS-SVM models.

Step 3: the resulting statistical performances (mean, median and variance) were collected, tabulated and their distributions were plotted. It is to be noted that only the 95% confidence intervals of estimation are reported in this study due to the fact that confidence intervals provide more information than other statistical values about the range of prediction associated with the model. The 95% confidence intervals are determined by finding the 2.5th and 97.5th percentiles of the constructed distribution.

Suppose the current time is t , $y(t+l)$ for the future time $t+l$ is predicted with the knowledge of the value $y(t-n), y(t-n+1), \dots, y(t)$ for past time $t-n, t-n+1, \dots, t$, respectively. The prediction function is expressed as:

$$y(t+l) = f(t, y(t), y(t-1), \dots, y(t-n)) \quad (11)$$

As discussed above, in this study, next week waste generation forecasted by 12 previous weekly waste generation time series. Relative Mean Errors (RME), Root Mean Squared Errors (RMSE), Mean Absolute Relative Error (MARE) and coefficient of determination (R^2) are applied as performance indices.

$$RME = \frac{1}{n} \sum_{i=1}^n |Y_i - Y_i^*| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |Y_i - Y_i^*|^2} \quad (13)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - Y_i^*}{Y_i} \right| \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y_i^*)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (15)$$

Where Y_i is the observation value, \bar{Y}_i is the average value and Y_i^* is the predicted value.

RESULTS & DISCUSSION

A kernel function has to be selected from the qualified functions. Radial basis function was applied due to its benefits over other kernel functions (Han, 2004; Dibike *et al.*, 2001). Additionally, many works in modeling and forecasting have demonstrated the successful application of the radial basis function in support vector regression (Liong and Sivapragasam, 2002; Choy and Chan, 2003; Yu, 2004). The SVM parameters (C capacity, ϵ and γ kernel-specific parameter) are interdependent, and their (near) optimal values are often obtained by a trial and error method. Optimization of these parameters has been done by a systematic grid search of the parameters using leave-one-out cross-validation on the training set. In this grid search, first, a broad range of parameters settings are investigated with large steps. Here, optimized

values of C and ϵ for a specified γ were obtained and then γ was changed. Second, after identifying a promising region, this region is searched in more detail. The test set is used as an independent set to calculate the final prediction error. Furthermore, the test error is not used to select the optimal model but its size is compared to test set errors with other settings to identify possible overtraining.

RME, RMSE, MARE and R^2 were used to find optimum. The optimal parameters (C, ϵ, γ) = (150, 0.001, 0.6) were obtained at RME=1467, RMSE=2070, MARE=0.027 and $R^2=0.761$. Fig. 1 shows this optimal value.

75 percentages of total inputs were used for train and rest of them was used for test. Observations mapped via predictions in Fig. 2. Fig. 3 illustrate coefficient of determination in training and test stage.

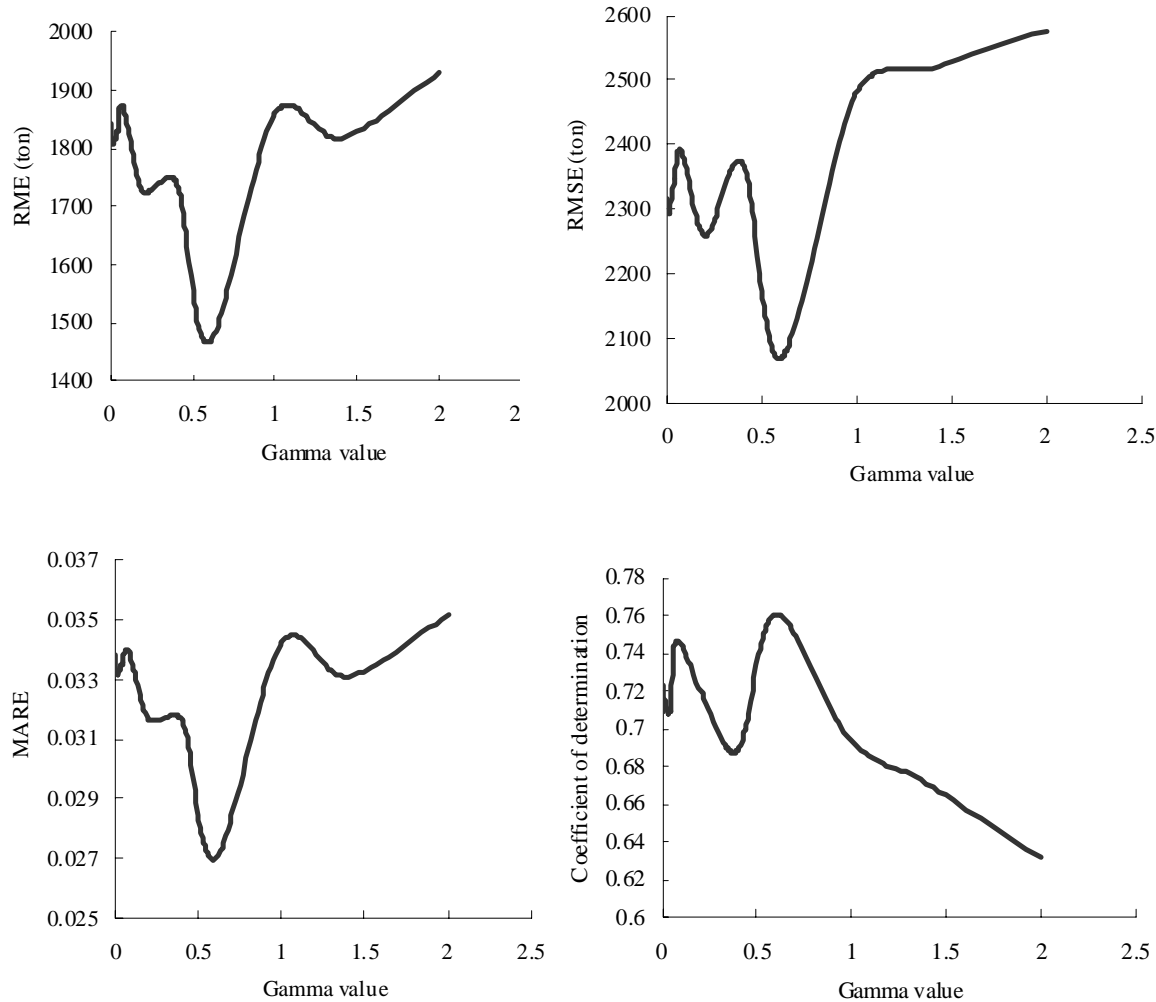


Fig. 1. Statistics indexes via γ values to find SVM optimums a) RME, b) RMSE, c) MARE and d) R^2

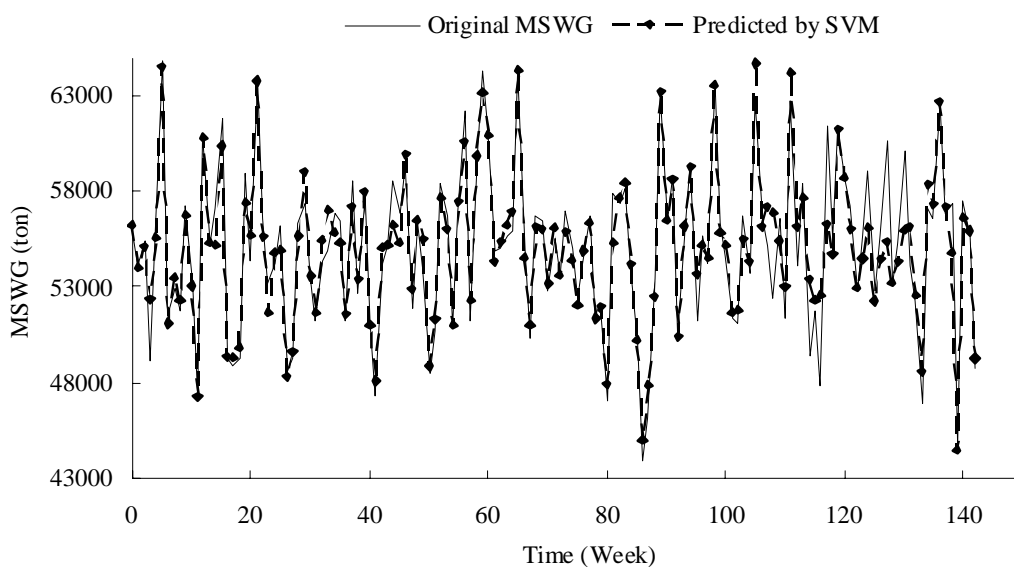


Fig. 2. Forecasting results of MSWGs by SVM model

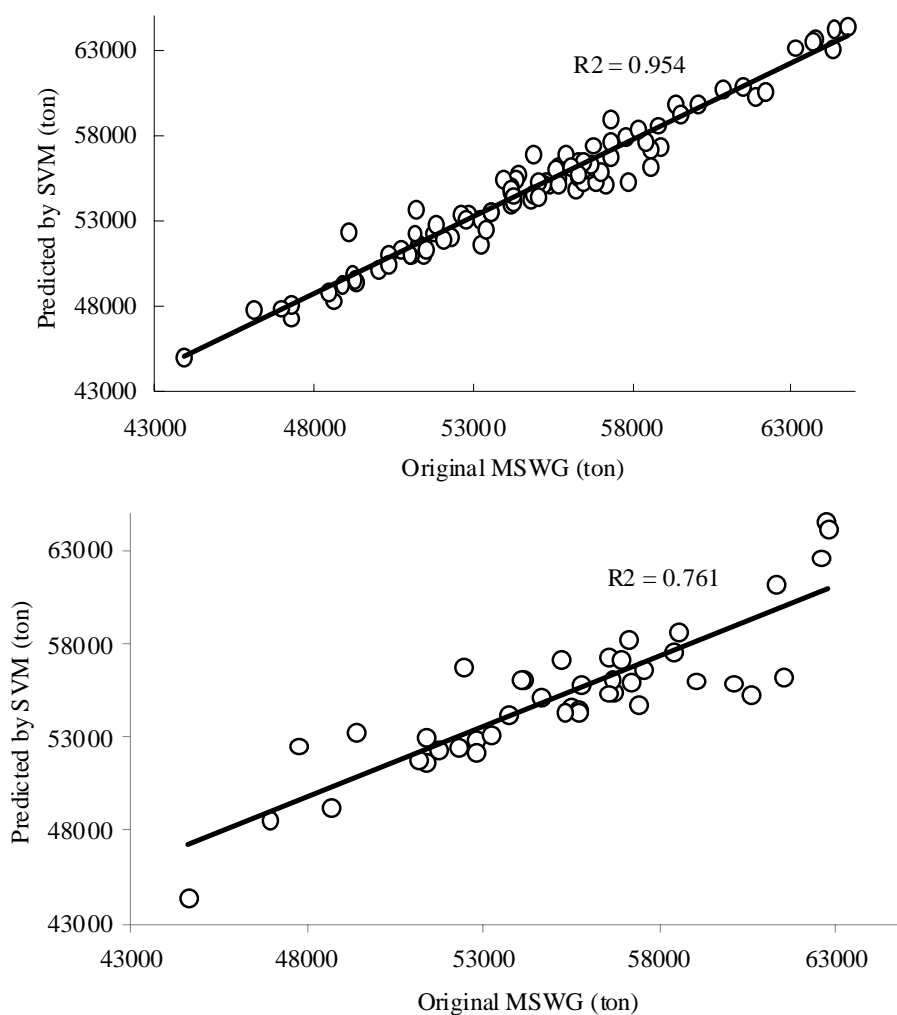


Fig. 3. Observations via predictions by the SVM model during a) train and b) test stages

As seen in Fig. 3, SVM could forecast MSW generation with coefficient of determination 0.761 and coincidence between observations and predictions are acceptable. It means SVM have a good ability for MSW generation prediction.

In PLS method, number of components (*NCs*) should be determined properly. Any method used for determining *NCs* should take into account not only the goodness of fit but also the complexity taken to achieve that fit. In other words, when building model components, a balance between *NCs* and the ability to accurately predict data should be considered. A model with an insufficient *NC* cannot predict the data accurately enough. A model with too many components has more components that it needs in predicting the data. *V*-fold cross validation method can actually find the optimal *NC* which completely described by Stone and Brooks (Stone and Brooks, 1990).

Here, *X* scores produced by the *plsregress* function in the Statistics Toolbox of MATLAB (R2009) were used to predict MSWGs.

To find optimum number of components, *V*-fold cross validation was applied. Here *V* set to 10. Then, PLS has built four components which eigenvalues of the components are illustrated in Table 1.

X-scores of these six components were replaced to original data. It is noted, the components evolve 98% variance.

The PLS-SVM optimal parameters were found by similar procedure to SVM pointed at $(C, \epsilon, \gamma) = (215, 0.125, 0.077)$ as seen in Fig. 4. In addition, the model results and coefficient of determination for training and test stages of the model is shown in Fig. 5 and 6 respectively.

Table 1. Eigenvalues of PLS components

Component No.	Eigenvalue	% Total variance
1	4.759	4.759
2	2.246	2.246
3	0.793	0.793
4	0.508	0.508

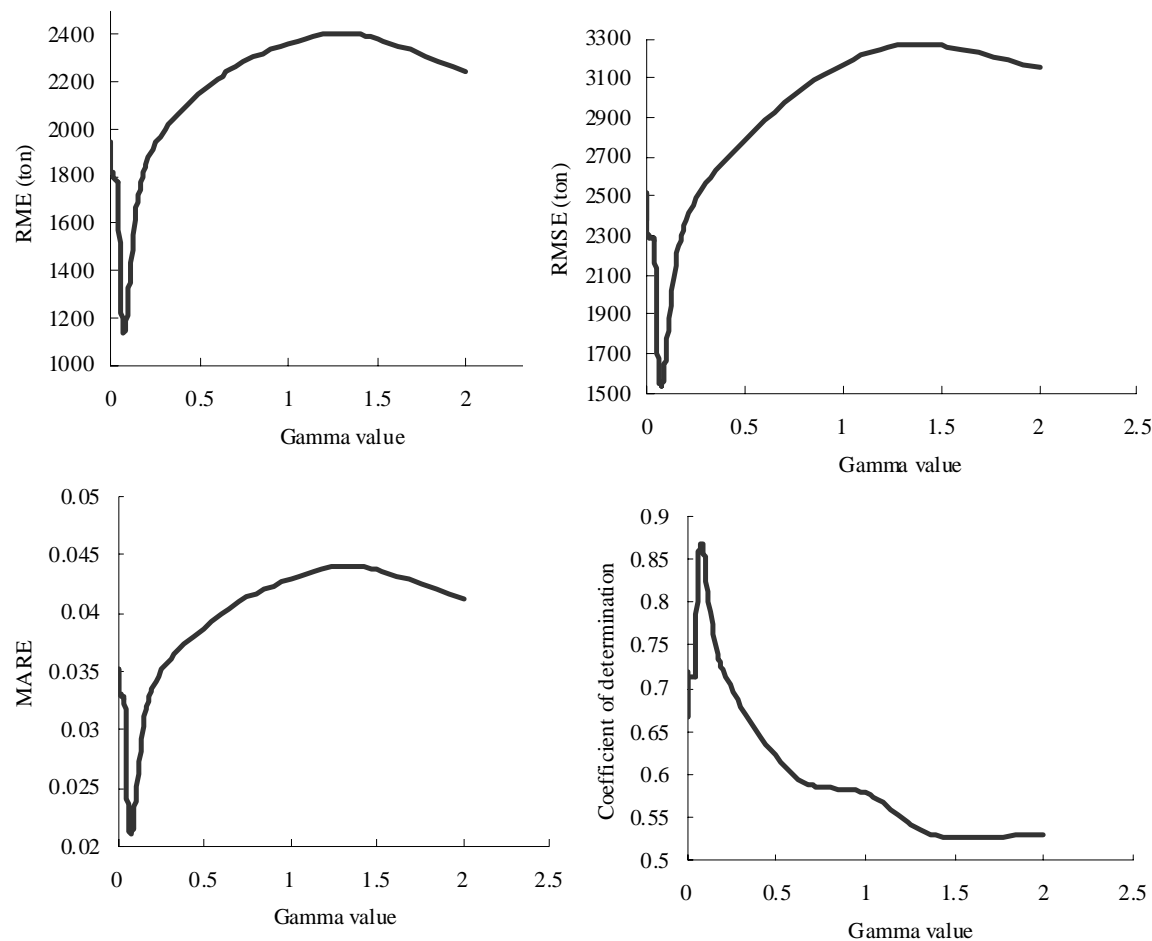


Fig. 4. Statistics indexes via γ values to find PLS-SVM optimums a) RME, b) RMSE, c) MARE and d) R^2

Forecasting Municipal Solid waste Generation

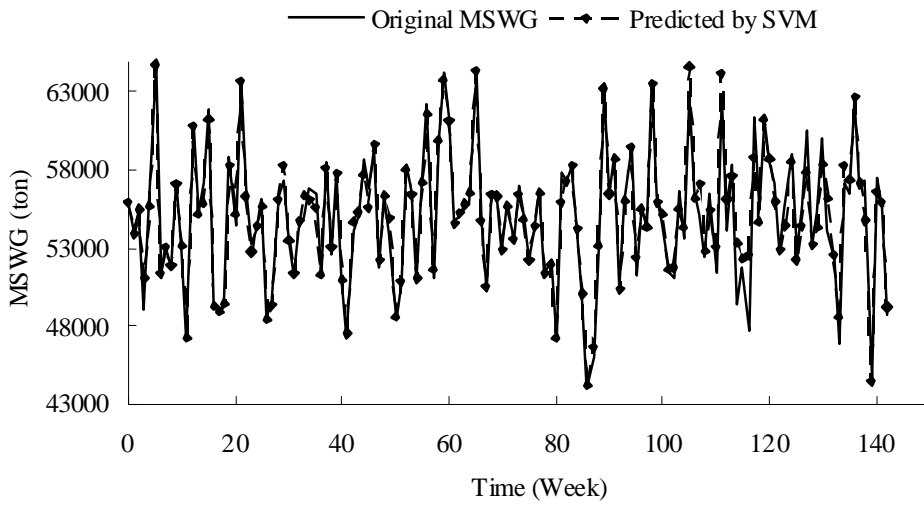


Fig. 5. Forecasting results of MSWGs by PLS-SVM model

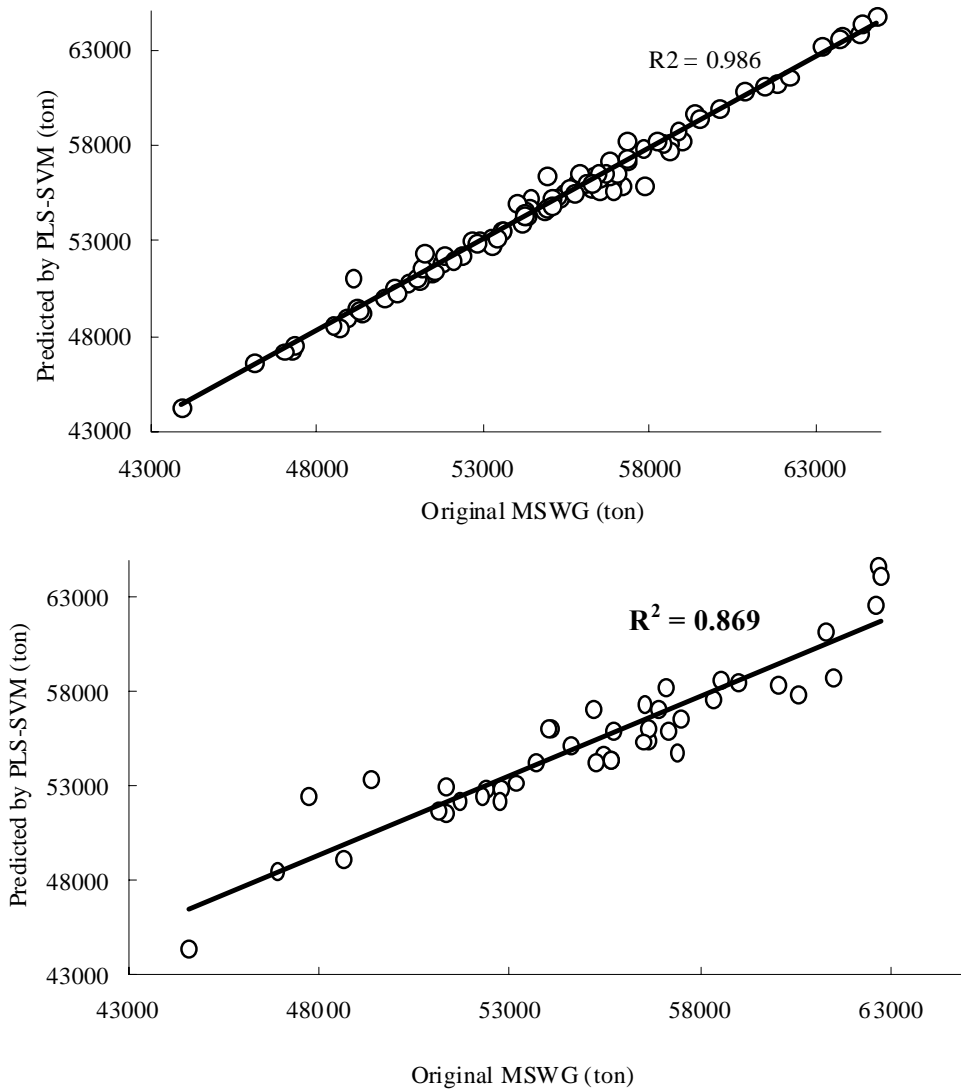


Fig. 6. Observations via predictions by the PLS-SVM model during a) train and b) test stages

To compare results of two SVM and PLS-SVM models, discussed statistics indices were used which illustrated in Table 2. As seen, PLS-SVM resulted in fewer errors and higher coefficient of determination than SVM. Moreover, computation for PLS-SVM took less time than SVM model. The PLS-SVM method produced acceptable results but error of both models still remains considerable. Complexity of the MSW management systems and many factors, which affected MSW generation cause these errors. However, PLS-SVM model achieve simpler model and faster training speed. Obviously, the reduction of the input vector dimensions is resulted in the reduction of SVM size and shortening the SVM training periods. Therefore, PLS-SVM model could be a better predictive model than SVM. Alternatively, the proposed model can be implemented as MSW generation annually or monthly.

The uncertainty in the estimates of the observed and predicted weekly MSWG during the training and test stages has been quantified by estimating the confidence intervals of the simulation results. In this research, the 95 percent prediction uncertainties (95PPU) were calculated for forecasting models. This is calculated by the 2.5th (X_L) and 97.5th (X_U) percentiles of the cumulative distribution of every simulated point. The goodness of fit is, therefore, assessed by the uncertainty measures calculated from the percentage of measured data bracketed by the 95PPU band, and the average distance \bar{d}_x between

the upper and the lower 95PPU (or the degree of uncertainty) determined from Eq. (17).

$$\bar{d}_x = \frac{1}{k} \sum_{i=1}^k (X_U - X_L) \tag{16}$$

where k is the number of observed data points. The best outcome is that 100% of the measurements are bracketed by the 95PPU, and \bar{d}_x is close to zero. However, because of model uncertainty, the ideal values will generally not be achieved. A reasonable measure for, is calculated by the d-factor expressed as:

$$d - factor = \frac{\bar{d}_x}{\sigma_x} \tag{17}$$

where σ_x is the standard deviation of the measured variable X. A value of <1 is a desirable measure for the d-factor.

Plot of confidence intervals of SVM and PLS-SVM for Tehran are shown in Fig. 7a and (b) during the training and testing stages. Moreover, Table 3 summarizes result analysis of 1000 simulations for Tehran case study.

As obviously seen in Fig. 7a and (b), both models well predicted the changes of MSWG in Tehran. For SVM model, the magnitudes MSWG, except at peaks,

Table 2. Comparison of SVM and PLS-SVM models during testing and training periods

Estimator	SVM		PLS-SVM	
	Train	Test	Train	Test
RME	677	1467	316	1139
RMSE	935	2070	499	1541
MARE	0.012	0.027	0.006	0.021
R ²	0.954	0.761	0.986	0.869
Computation Time (s)	42		29	

Table 3. Forecasting performance during the training and the testing stages based on averages obtained from 1000 time simulations

Performance index	PLS-SVM		SVM	
	Train	Test	Train	Test
\bar{d}_x	3724	4377	4258	5108
d-factor	0.33	0.46	0.41	0.57

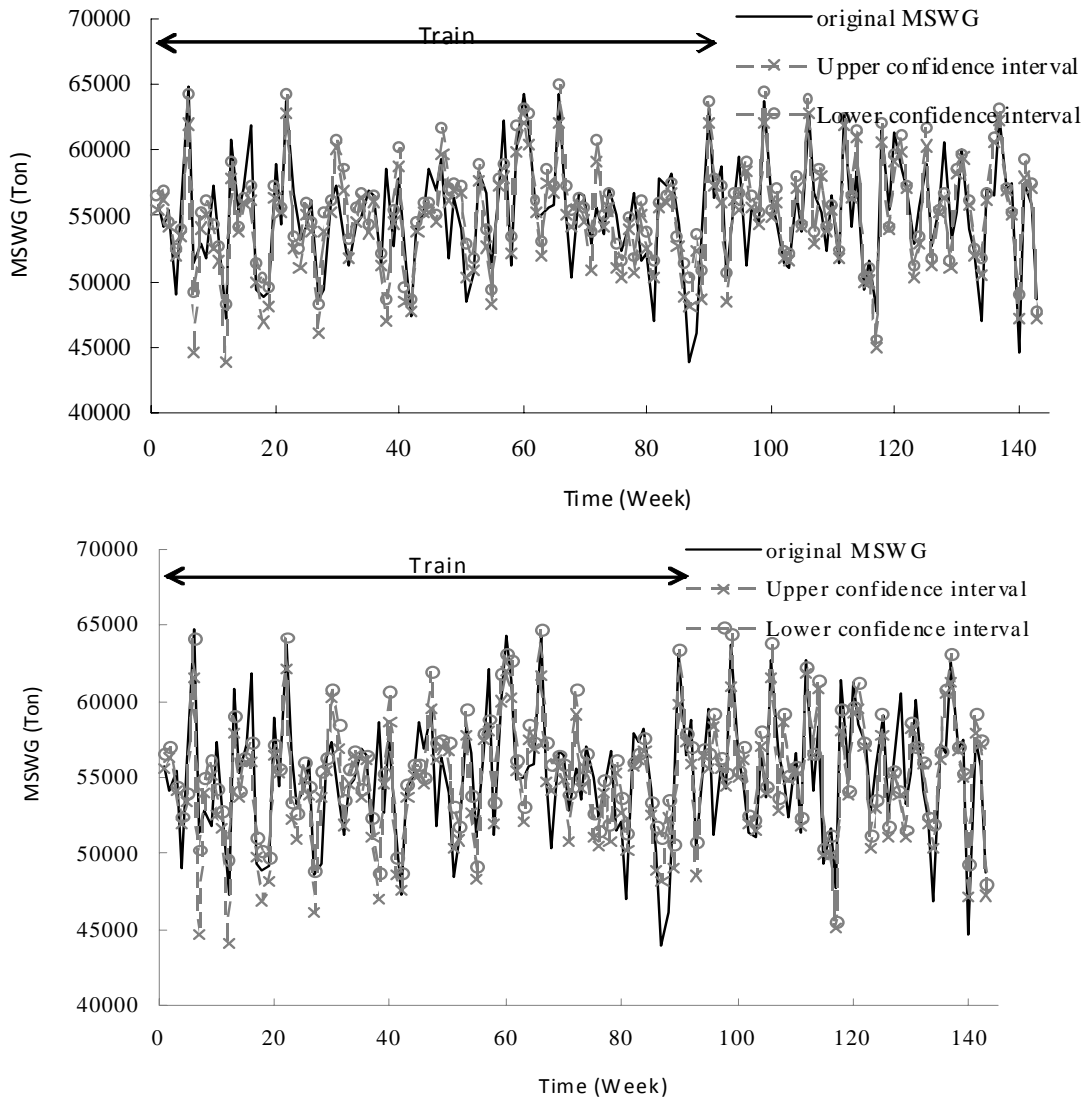


Fig. 7. the estimates of MSWG during the training and test stages for 1000 simulations by a) SVM and b) PLS-SVM

were estimated closer to the observed data. However PLS-SVM estimations were closer to the observed data even at peaks. A similar trend was also found at the testing stage although the magnitude of uncertainty in the training stage was lower than that of the testing stage and the magnitudes of lower and upper confidence bounds are estimated closer to the observed data in the training stage.

In addition, it was found from Table 3 the PLS-SVM had lower uncertainty during training and test stages. Higher value of d -factor of SVM shows that this model is more sensible than PLS-SVM model to training data. Consequently, the wide of 95PPU bond, e.g. \bar{d}_x , for PLS-SVM is smaller than for SVM. For both models, 50% of the measurements were bracketed

by the 95PPU. Because PLS-SVM creates lower bound, numbers of observed data located on confidence bound have no change. However, uncertainty of this model is more reasonable on 95PPU than SVM. This shows that PLS-SVM had more robustness, and had a lower sensitivity to change of input variables than SVM.

CONCLUSION

Support vector machine and support vector regression have demonstrated their success in time-series analysis and statistical learning. However, little work has been done in waste management also in forecasting waste generation. In this paper, the feasibility of applying support vector regression and data reduction were examined for waste generation time series prediction. After numerous experiments, a set of SVM parameters

proposed that can predict MSW generation time series very well. The results show that the SVM predictor significantly outperforms the other baseline predictors. This evidences the applicability of support vector regression for forecasting MSW generation. Also, PLS minimized input data and decreased error of final modeling. Consequently, training finished very quickly with acceptable correlation. Combination of SVM with PLS produced a suitable model for MSW generation in large cities like Tehran.

Meanwhile, the uncertainty associated with the estimation of MSWG was estimated by Monte Carlo simulations. The simulation results using 95% confidence intervals indicated that estimations of the SVM and PLS-SVM models were closer to the observed data except in peak points. However PLS-SVM had more robustness, and had a lower sensitivity to change of input variables than SVM.

REFERENCES

- Abdoli, M. A., Karbassi, A. R., Samiee-Zafarghandi, R., Rashidi, Zh., Gitipour, S. and Pazoki, M. (2012). Electricity Generation from Leachate Treatment Plant, *Int. J. Environ. Res.*, **6** (2), 493-498.
- Arshad, A., Hashmi, H. N. and Qureashi, I. A. (2011). Anaerobic Digestion of CHLORphenolic Wastes, *Int. J. Environ. Res.*, **5** (1), 149-158.
- Bayar, S., Demir, I. and Engin, G. O. (2009). Modeling leaching behavior of solidified wastes using back-propagation neural networks. *Ecotoxicology and Environmental Safety*, **72** (3), 843-850.
- Beigl, P. and Lebersorger, S. (2009). Forecasting Municipal Solid Waste Generation for urban and rural regions. In R. Cossu and R. Stegmann (eds.), *twelfth International Waste Management Landfill Symposium* (pp. 27-38). Sardinia, Italy, CISA Environmental Sanitary Engineering Center.
- Blasco, J. A., Fueyo, N., Dopazo, C. and Ballester, J. (1998). Modeling the temporal evolution of a reduced combustion chemical system with an artificial neural network. *Combustion and Flame*, **113**, 38-52.
- Broomhead, D. and Lowe, D. (1998). Multivariable functional interpolation and adaptive networks. *Complex Systems*, **2**, 321-355.
- Chang, N. B. and Lin, Y. T. (1997). An analysis of recycling impacts on solid waste generation by time series intervention modeling. *Resources Conservation and Recycling*, **19** (3), 165-186.
- Chang, N. B., Pan, Y. C. and Haung, S. D. (1993). Time series forecasting of solid waste generation. *J Resour Manage Technol*, **21**, 1-10.
- Chen, H. W. and Chang, N. B. (2000). analysis of solid waste generation based on grey fuzzy dynamic modeling *Resources, Conservation and Recycling*, **29**, 1-18.
- Chen, J. Huang, W., Han, J. and Cao, Sh. (2011). The Characterization and Application of Biological Remediation Technology for Organic Contaminants, *Int. J. Environ. Res.*, **5** (2), 515-530.
- Choy, K. Y. and Chan, C. W. (2003). Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science*, **34** (14-15), 763-773.
- Corcoran, J., Wilson, I. and Ware, J. (2003). Sparse support vector regression based on orthogonal forward selection for the generalised kernel model. *International Journal of Forecasting*, **19**, 623-634.
- Damghani, A. M., Savarypour, G., Zand, E. and Deihimfard, R. (2008). Municipal solid waste management in Tehran: Current practices, opportunities and challenges. *Waste Management*, **28** (5), 929-934.
- Dibike, Y. B., Velickov, S., Solomatine, D. and Abbott, M. B. (2001). Model induction with support vector machines, Introduction and applications. *Journal of Computing in Civil Engineering*, **15** (3), 208-216.
- Dong, C. Q., Jin, B. S. and Li, D. J. (2003). Predicting the heating value of MSW with a feed forward neural network. *Waste Management*, **23** (2), 103-106.
- Geladi, P. and Bruce, K. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, **185**, 1-17.
- Grossman, D, H. J. and Mark, D. H. (1974). Waste generation methods for solid waste collection. *J. Environ. Eng.*, **6**, 1219-1230.
- Han, D. and Cluckie, I (2004). Support vector machines identification for runoff modeling. In S. Y. Liong, Phoon, K.K., Babovic, V (ed.), *Proceedings of the Sixth International Conference on Hydroinformatics*, Singapore, 21-24 June 2004 (pp. 511-520). Singapore, *Proceedings of the Sixth International Conference on Hydroinformatics*.
- Hyun, I, Borinara, P. and Hong, K. D. (2011). Geotechnical Considerations for End-Use of Old Municipal Solid Waste Landfills, *Int. J. Environ. Res.*, **5** (3), 573-584.
- Jalili, G. Z. M. and Noori, R. (2008). Prediction of municipal solid waste generation by use of artificial neural network: A case study of Mashhad. *International Journal of Environmental Research*, **2** (1), 13-22.
- Karaca, F. and Ozkaya, B. (2006). NN-LEAP, A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. *Environmental Modelling & Software*, **21** (8), 1190-1197.
- Liong, S. Y. and Sivapragasam, C. (2002). Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association*, **38** (1), 173-186.
- Lorber, A., Kowalski, B. R. and Chemometrics, J. (1987). A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*, **1**, 19-31.

- Mahmoudkhani, R., Hassani, A. H., Torabian, A. and Borghai, S. M. (2012). Study on High-strength Anaerobic Landfill Leachate Treatability By Membrane Bioreactor Coupled with Reverse Osmosis, *Int. J. Environ. Res.*, **6 (1)**, 129-138.
- Maqbool, F., Bhatti, Z. A., Malik, A. H., Pervez, A. and Mahmood, Q. (2011). Effect of Landfill Leachate on the Stream water Quality, *Int. J. Environ. Res.*, **5 (2)**, 491-500.
- Mukherjee, S., Osuna, E. and Girosi, F. (1997). Nonlinear prediction of chaotic time series using a support vector machine. *IEEE Workshop on Neural Networks and Signal Processing*. Amelia Island, FL.
- Müller, K. R., Smola, A. J., Raßtsch, G., Schölkopf, B., Kohlmorgen, J. and Vapnik, V. (1997). Predicting Time Series with Support Vector Machines. *Proceedings of the 7th International Conference on Artificial Neural Networks* pp. 999-1004, Springer-Verlag.
- Nada, W. M., Van Rensburg, L., Claassens, S., Blumenstein, O. and Friedrich, A. (2012). Evaluation of Organic Matter Stability in Wood Compost by Chemical and Thermogravimetric Analysis, *Int. J. Environ. Res.*, **6 (2)**, 425-434.
- Niessen, W. and Alsobrook, A. (1972). Municipal and industrial refuse, composition and rates. *Proceedings of National Waste Processing Conference*, pp. 112-117.
- Noori, R., Abdoli, M. A., Farokhnia, A. and Abbasi, M. (2009). Results uncertainty of solid waste generation forecasting by hybrid of wavelet transform-ANFIS and wavelet transform-neural network. *Expert Systems with Applications*, **36 (6)**, 9991-9999.
- Nouri, N., Poorhashemi, S. A., Monavari, S.M., Dabiri, F. and Hassani, A. H. (2011). Legal Criteria and Executive Standards of Solid Waste Disposal Subjected to Solid Waste Management Act, *Int. J. Environ. Res.*, **5 (4)**, 971-980.
- Rashidi, Zh., Karbassi, A. R., Ataei, A., Ifaei, P., Samiee-Zafarhandi, R. and Mohammadzadeh, M. J. (2012). Power Plant Design Using Gas Produced By Waste Leachate Treatment Plant, *Int. J. Environ. Res.*, **6 (4)**, 875-882.
- Safari, E., Jalili Ghazizade, M., Shokouh, A. and Nabi Bidhendi, Gh. R. (2011). Anaerobic Removal of COD from High Strength Fresh and Partially Stabilized Leachates and Application of Multi stage Kinetic Model, *Int. J. Environ. Res.*, **5 (2)**, 255-270.
- Saikat, M. and Jun, Y. (2008). Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. *Casualty Actuarial Society*. Arlington, Virginia, 79-90.
- Shafieiyoun, S., Ebadi, T. and Nikazar M. (2012). Treatment of Landfill Leachate by Fenton Process with Nano sized Zero Valent Iron particles, *Int. J. Environ. Res.*, **6 (1)**, 119-128.
- Stone, M. and Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society*, **52**, 237-269.
- Svozil, D., Kvasnicka, V. and Pospichal, J. (1997). Introduction to multilayer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, **39**, 43-62.
- Vapnik, V. (1995). *Nature of Statistical Learning Theory*. Springer. New York.
- Tchobanoglous, G., Eliaseen, R. and Theisen, H. (1977). *Solid Waste: Engineering principles and Management*. Tokyo, McGraw Hill.
- Thanh, N. P. and Matsui, Y. (2011). Municipal Solid Waste Management in Vietnam: Status and the Strategic Actions, *Int. J. Environ. Res.*, **5 (2)**, 285-296.
- Vapnik, V. (1998). *Statistical Learning Theory*. Wiley. New York.
- Vapnik, V., Golowich, S. and Smola, A. (1997a). Support method for function approximation regression estimation, and signal processing Report, MIT Press, Cambridge, MA.
- Wang, X. X., Chen, S., Lowe, D. and Harris, C. J. (2006). Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemom. Intell. Lab. Syst.*, **82**, 165-175.
- Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (ed.), *Multivariate analysis* pp. 391-420. New York, Academic Press.
- Wold, S., Ruhe, A., Wold, H. and Dunn, W. (1984). The collinearity problem in linear regression, the partial least squares (PLS) approach to generalized inverses. *Journal of Statistics Computation*, **5**, 735-743.
- Yu, X., Liang, S.-Y. and Babovic, V. (2004). EC-SVM approach for realtime hydrologic forecasting. *Journal of Hydroinformatics*, **6 (3)**, 209-223.
- Zhang, Y. X., Li, H., Hou, A. X. and Havel, J. (2006). Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemometrics and Intelligent Laboratory Systems*, **82 (1-2)**, 165-175.
- Zhang, Y. X. (2007). Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis. *Talanta*, **73 (1)**, 68-75.
- Zhu, Z. and ReVelle, C. (1993). A cost allocation method for facilities siting with fixed-charge cost functions. *Civil Engineering Systems*, **7**, 29-35.