



Research Article

**AN MS EXCEL TOOL FOR PARAMETER ESTIMATION BY
MULTIVARIATE NONLINEAR REGRESSION IN ENVIRONMENTAL
ENGINEERING EDUCATION**

**Selami DEMİR*, Aykut KARADENİZ, Hülya CİVELEK YÖRÜKLÜ,
Neslihan MANAV DEMİR**

Department of Environmental Engineering, Yıldız Technical University, Esenler-ISTANBUL

Received: 13.02.2017 Revised: 14.03.2017 Accepted: 17.03.2017

ABSTRACT

This paper presents the implementation of an MS Excel tool for parameter optimization in environmental engineering education. Visual Basic for Applications v7.0 were used for implementation of the program. A number of test cases were also provided to test the performance of the tool including the fields of air pollution control, water treatment, and anaerobic treatment. For each test case, calculated coefficients of determination were 0.98 and above. Results suggested that the MS Excel tool produces satisfactorily fast and reliable results, and can be used confidently for optimization works in environmental studies.

Keywords: Environmental modeling, optimization, multivariate nonlinear regression, MS Excel.

1. INTRODUCTION

Although most formulations in environmental processes can be linearized for handling optimization work easily, sometimes parameter estimation becomes difficult due to nonlinear nature of the environmental process of interest. In such cases, one of the numerical methods that can be used to optimize the parameters required to express the process is the Gauss-Newton algorithm. This algorithm is based on principles similar to the least squares method used in linear regression applications. More precisely, the goal of this method is to minimize the sum of the squares of the differences between the calculated and measured values of the dependent variable.

There are many software packages developed to perform nonlinear regression analysis that explain the nonlinear relationship between variable parameters of many natural phenomena. Softwares such as MATLAB and STATA are used quite frequently for academic purposes among these [1-2]. However, these softwares require high CPU performance. In this regard, Visual Basic for Applications (VBA) draws attention with its simple handling and easy accessibility [3-6]. Visual Basic for Applications (VBA) is an event-driven programming language and is a derivative of Microsoft Visual Basic 6 (VB6). VBA is embedded in Microsoft Office software like Excel and is compiled to packed code (P-code). The P-code is then stored as a separate stream in an Excel file, independent of the spreadsheet stream. This property of VBA simplifies

* Corresponding Author/Sorumlu Yazar: e-mail/e-ileti: seldemir@yildiz.edu.tr, tel: (212) 383 53 72

the problem of compiling the program into an executable file and of preparing a distribution package, if necessary, usually associated with stand-alone computer programming platforms, like VB6. Simple use and easy accessibility make VBA a useful tool in the field of education. Especially in engineering education, VBA is an effective platform in the teaching of numerical methods [7-8].

Due mostly to the licensing fees and other drawbacks of several software packages such as the fact that the code and algorithms are unavailable to the user, professors who teach numerical methods in engineering fields often discourage the use of commercial software packages. Besides, professors usually wish focusing on the numerical methods and algorithms when teaching the regression techniques and other engineering calculations. Thus, open-source software packages are encouraged. Of these, MS Excel and VBA offers a perfect solution. Considering the fact that almost all of engineering students own or have the access to a personal computer, and that a spreadsheet software (usually MS Excel) is already installed on these computers at least with a student license, one concludes that MS Excel can be used efficiently for teaching in engineering fields. Therefore, a great number of papers have been dedicated on the usability of MS Excel and VBA in engineering education as well as in reasearch [9-19].

The pupose of this study is to implement a flexible MS Excel tool incorporating VBA that uses a multivariate nonlinear regression algorithm for parameter estimation in any linear/nonlinear optimization work encountered in environmental studies. The implemented program was tested using experimental data from previous works including cyclone pressure drop, polarization curves for microbial fuel cells, and rate of oxygen uptake of microorganisms in wastewater treatment modeling.

2. MATERIALS AND METHODS

A flexible MS Excel tool using multivariate nonlinear regression for parameter estimation in optimization work for environmental studies was implemented in Visual Basic for Applications (VBA) v7.0 and tested with a number of experimental data.

2.1. Implementation of the Tool

Gauss-Newton method was employed in the implemented code using the flowchart in Fig. 1.

Assuming any multivariate function $f(x_1, x_2, \dots, x_m)$ of m independent variables which is defined by an array of constants as $[a_1, a_2, \dots, a_k]^T$, a first-order Taylor expansion of the function is given by the sum of all partial derivatives with respect to each constant as follows:

$$f(x_1, x_2, \dots, x_m)_{j+1} = f(x_1, x_2, \dots, x_m)_j + \frac{\partial f(x_1, x_2, \dots, x_m)_j}{\partial a_1} \Delta a_1 + \frac{\partial f(x_1, x_2, \dots, x_m)_j}{\partial a_2} \Delta a_2 + \dots + \frac{\partial f(x_1, x_2, \dots, x_m)_j}{\partial a_k} \Delta a_k \quad (1)$$

The objective of the tool is to find a set of constants $[a_1, a_2, \dots, a_k]^T$ iteratively that produces the exact value of $f(x_1, x_2, \dots, x_m)_{j+1}$ starting from the point $f(x_1, x_2, \dots, x_m)_j$, through which the error $y_i - f(x_{1i}, x_{2i}, \dots, x_{mi})$ for each data point $i = 1, 2, \dots, n$ is minimized. For a given set of n data points that consists of m independent variables and one dependent variable connected to independent variables by a pre-defined array of k constants, the mean square error is expressed as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_i - f(x_{1i}, x_{2i}, \dots, x_{mi})]^2 \quad (2)$$

Starting with an arbitrary set of constants, this can be accomplished iteratively by updating each constant's value by an update factor that can be expressed as follows:

$$U = (J^T J)^{-1} J^T E \quad (3)$$

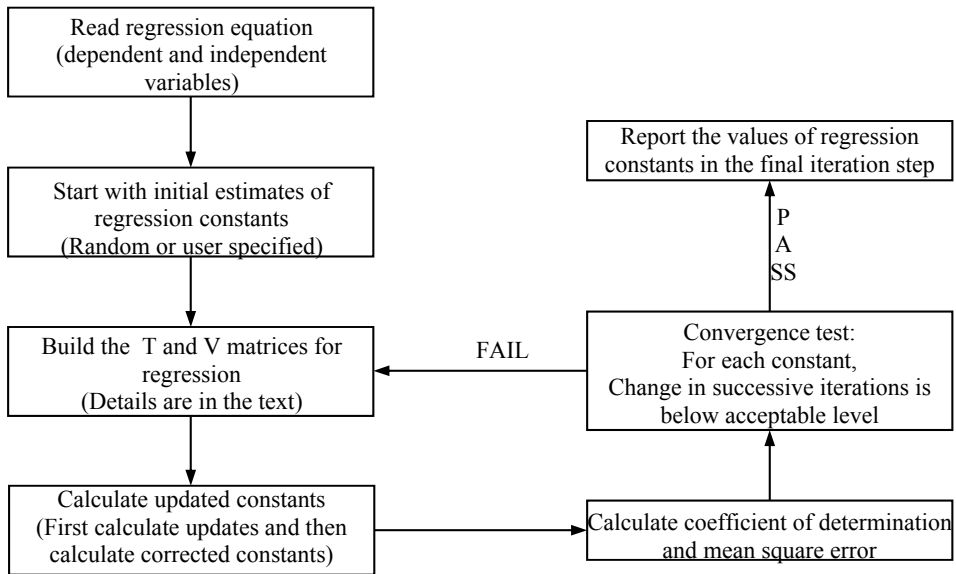


Figure 1. Flowchart of the implemented algorithm

where U is a $k \times 1$ column vector composed of update factors for constants as $[\Delta a_1, \Delta a_2, \dots, \Delta a_k]^T$, J is the $n \times m$ Jacobian matrix composed of partial derivatives of the multivariate function f in an iteration step, E is the $n \times 1$ column vector composed of the error value $y_i - f(x_{1i}, x_{2i}, \dots, x_{mi})$ for each data point $i = 1, 2, \dots, n$. The new set of constants can be calculated by updating the old set of constants as $(a_{new})_i = a_i - \eta u_{i1}$ for each constant $i = 1, 2, \dots, k$, with η being user-defined constant for learning rate which is a constant used for relaxing the solution algorithm. The new set of constants can be used to calculate a still new set of them iteratively until mean square error is minimized, or alternatively until each of the percent changes in constants in two successive iterations is equal to or less than a predefined limit.

The implemented tool requires maximum number of iterations, predefined limit of change in constants (percent maximum error), percent step size, learning rate, and numbers of independent variables and regression constants. The user is expected to write the regression formula of his choice in a cell as a string. The implemented code recognizes the formula and converts it into a mathematical equation for nonlinear regression. The partial derivatives are calculated numerically as centered finite differences, which requires an input of percent step size for derivation. The tool also allows the input for the number of independent variables through the user interface. Finally, the tool starts iterating by a random or a user-defined set of initial values for constants. An MS Excel view of the tool is shown in Fig. 2.

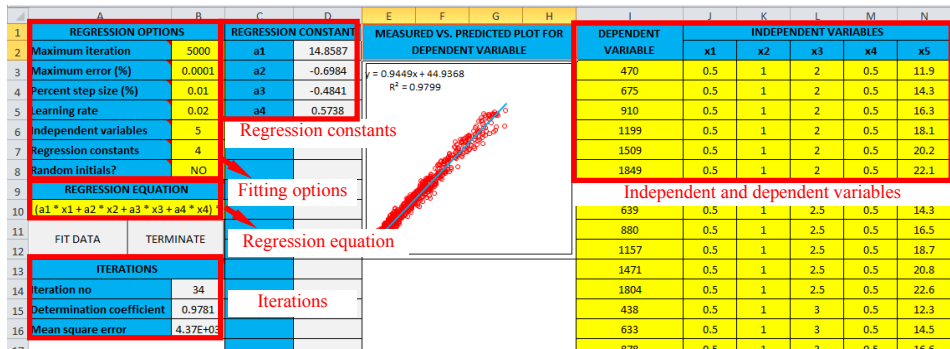


Figure 2. An MS Excel sheet view of the tool

2.2. Test Problem: Cyclone Pressure Drop

Pressure drop in cyclone separators can be estimated by inlet velocity heads as follows:

$$\Delta P = \frac{1}{2} \rho V^2 N_H \quad (4.a)$$

where ΔP is pressure drop in cyclone (N/m^2), ρ is the density of air (kg/m^3), V is the inlet velocity to the cyclone (m/s), and N_H is the number of inlet velocity heads. A great number of models were developed for predicting the number of inlet velocity heads depending on the cyclone geometry. A summary of these models are given in Demir [20], in which a new, practical model was also presented based on experimental data as follows:

$$N_H = k_1 \frac{ab}{D_e^2} + k_2 \frac{h_b}{D} + k_3 \frac{h_c}{D} + k_4 \frac{S}{D} \quad (4.b)$$

where k_1 , k_2 , k_3 , and k_4 are regression constants, a is the inlet height, b is the inlet width, D_e is the diameter of vortex finder, h_b is the cylindrical height, h_c is the conical height, S is the length of vortex finder, and D is the body diameter of the cyclone. Experimental data from Demir [20] and Demir et al. [21, 22] was used for testing the MS Excel tool for obtaining a set of regression constants (k_1 , k_2 , k_3 , and k_4) that fits the experimental data best.

2.3. Test Problem: Aeration

Volumetric mass transfer coefficient ($k_L a$) is of great importance to environmental engineers in designing aeration systems for various water and wastewater treatment plants employing either physical, chemical or biological treatment processes [23]. Therefore, accurate estimation of volumetric mass transfer coefficient from experimental data is vital in the design of cost-effective and efficient aeration systems. The aeration process in a completely mixed reactor is explained by a mass balance equation as follows:

$$\frac{dc}{dt} = k_L a (C_S - C) \quad (5.a)$$

where C is the dissolved oxygen (DO) concentration in liquid phase at time t , and C_S is the saturation concentration of oxygen. Using the boundary conditions $(t, C) = (0, C_0)$ and $(t, C) = (t, C(t))$, this equation can be integrated into a nonlinear formulation for dissolved oxygen concentration in the following form:

$$C(t) = C_0 e^{-k_L a t} + (C_S - C_0)(1 - e^{-k_L a t}) \quad (5.b)$$

where C_0 is the initial DO concentration in liquid phase. Experimental determination of volumetric mass transfer coefficient involves measuring DO in liquid phase at various times t , and

running a nonlinear regression algorithm to estimate experimental $k_L a$ value. Experimental data was collected using a lab-scale aeration equipment and the implemented MS Excel tool was used to estimate $k_L a$ to test the tool's performance.

2.4. Test Problem: Anaerobic Digestion

As a result of anaerobic digestion techniques for producing energy, in the form of biogas, from organic wastes and wastewater treatment plant sludge becoming more and more widespread in recent years [24], an increased need for new mathematical models has emerged rapidly to establish methods to assess effectiveness of anaerobic process and to compare results from various research papers using kinetic constants [25]. A number of models were used for explaining the S-shaped curve of biogas production (as methane) depending on several kinetic parameters [25, 26]. These models include the modified Gompertz equation, the logistic function, and the transference function, all of which considers cumulative bio-methane production as a nonlinear function of several kinetic parameters and time.

In order to test the performance of the implemented MS Excel Tool, a bio-methane potential (BMP) test was conducted using a sample of macroalgae as substrate, and the experimental data was used in MS Excel tool to estimate kinetic parameters of bio-methane potential for transference function which is given as:

$$V_g = V_{max} \left[1 - \exp \left(-\frac{R}{V_{max}} (L - t) \right) \right] \quad (6)$$

where V_g is the cumulative biogas produced at the end of time t , V_{max} is the biogas potential, R is the rate of biogas production (1/time), and L is a model parameter (time) representing the lag phase of the process.

3. RESULTS

An MS Excel tool was implemented in Visual Basic for Applications for estimating parameters in nonlinear equations commonly encountered in undergraduate/graduate environmental studies. The tool employs Gauss-Newton algorithm and is able to estimate parameters in multivariate nonlinear functions. The most distinguishing property of the tool is that it allows the user to define a linear/nonlinear function on MS Excel sheet as a string (text) variable, recognizes the given function and converts it into a mathematical expression for use in parameter estimation. A number of sample problems most commonly encountered in environmental engineer's daily work was provided and the tool was tested with experimental data on these problems.

3.1. Cyclone Pressure Drop

The experimental data for cyclone problem was obtained from Demir [20] in which a new model was proposed based on pressure drop measurements on 162 cyclones at 6 different inlet velocities (972 data points). The data set was composed of five independent variables and four regression constants. Definition of the regression equation and results are shown in Table 1. The formulation in Eqn. 4.a was typed as the regression equation and global settings including maximum number of iterations, maximum error, percent step size, and learning rate were defined as given in Table 1. The tool estimated the parameters in 996 iterations (less than 10 seconds) starting from random initial values between 0 and 1. The estimated values of parameters were almost identical to those previously published in Demir [20], indicating that the implemented tool proves a valuable tool in cyclone pressure drop calculation.

Table 1. Performance data of the implemented tool for test problem: cyclones

<i>Definition of equation and global settings</i>		
Parameter	Value	Remarks
Regression equation	$0.61 \cdot x_5^2 \cdot (a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4)$	$x_1 = ab/D_c^2$ (Eqn. 4.b); $x_2 = b_y/D$ (Eqn. 4.b); $x_3 = b_z/D$ (Eqn. 4.b); $x_4 = S/D$ (Eqn. 4.b); $x_5 = V$ (Eqn. 4.a); $a_1 = k_1$ (Eqn. 4.b); $a_2 = k_2$ (Eqn. 4.b); $a_3 = k_3$ (Eqn. 4.b); $a_4 = k_4$ (Eqn. 4.b) Gas density (ρ) is assumed to be 1.22 kg/m ³ Dependent variable was cyclone pressure drop
Maximum iteration		5000
Maximum error		0.0001%
Percent step size		0.01%
Learning rate		0.01
<i>Estimated parameters and statistics</i>		
Parameter	Value (this study)	Value [9]
k_1	14.8588	14.8587
k_2	-0.6984	-0.6984
k_3	-0.4841	-0.4841
k_4	0.5738	0.5738
Coefficient of determination (R^2)	0.980	0.976
Mean square error (MSE)	4370	Not specified

3.2. Aeration

A simple aeration equipment was used to obtain experimental data on the process of aeration in a complete-mix reactor. Experiments were conducted at constant temperature and pressure at a constant, low mixing intensity for a period of 230 minutes. In the experiment, water was purged with pure nitrogen gas to reduce initial concentration of dissolved oxygen to zero. A total of 29 data points were used in the regression. Table 2 shows the regression settings and results. The formulation in Eqn. 5.b was typed as the regression equation. The tool estimated the parameters in 955 iterations (less than 10 seconds) starting from random initial values between 0 and 1. The estimated value of final dissolved oxygen concentration, that's the solubility of oxygen, was 8.45 mg/L with an estimated value of volumetric mass transfer coefficient of 0.0143/min. The mean square error (MSE) was calculated as $3.26 \cdot 10^{-2}$ and the coefficient of determination between experimental and modeled DO concentration (R^2) was calculated as 0.996. The results suggested that the implemented tool can be used for predicting experimental $k_L a$ value for a given aeration system. Fig. 3 shows the good agreement between experimental and model dissolved oxygen concentrations, verifying that the estimated value of $k_L a$ represents experimental data very well.

Table 2. Performance data of the implemented tool for test problem: aeration

<i>Definition of equation and global settings</i>		
Parameter	Value	Remarks
Regression equation	$a_1 \cdot \exp(-a_2 \cdot x_1) + (a_3 - a_1) \cdot (1 - \exp(-a_2 \cdot x_1))$	$x_1 = t$ (Eqn. 5.b); $a_1 = C_0$ (Eqn. 5.b); $a_2 = k_L a$ (Eqn. 5.b); $a_3 = C_S$ (Eqn. 5.b)
Maximum iteration		5000
Maximum error		0.0001%
Percent step size		0.01%
Learning rate		0.01
<i>Estimated parameters and statistics</i>		
Parameter	Value (this study)	
a_1	0.00	
a_2	0.0143 min ⁻¹	
a_3	8.45 mg.L ⁻¹	
Coefficient of determination (R^2)	0.996	
Mean square error (MSE)	$3.26 \cdot 10^{-2}$	

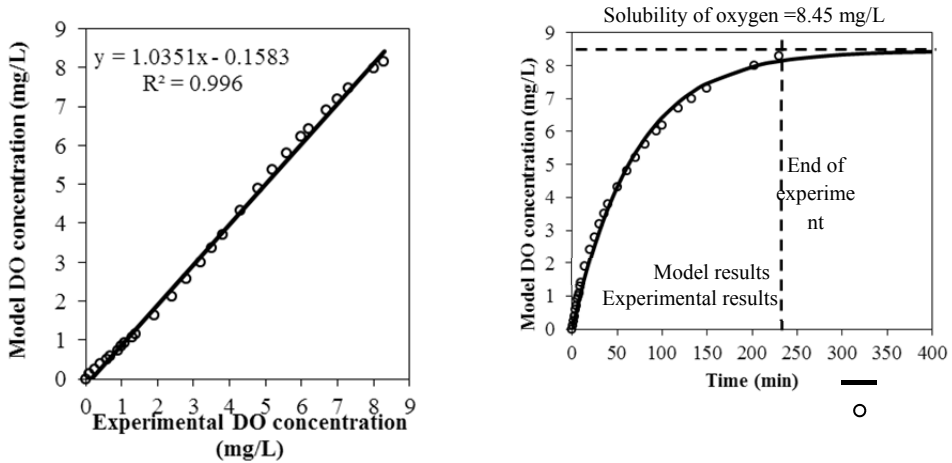


Figure 3. Experimental and model DO concentrations in water. a. correlation plot, b. time series data

3.3. Anaerobic Digestion

Bio-methane potential assay was performed for 20 days and produced biogas was measured at certain intervals. The rate of biogas production started with a low value and rapidly increased. The highest rate of production was observed in the second day of the experiment as 90 L/day under experimental conditions. Cumulative biogas production rate was 670 L at the end of 20 days.

A total of 13 data points each consisting of a set of assay time (t) and cumulative biogas production (V_g) was used for regression to estimate the parameters V_{max} , R , and L . Iterations were made by random initials. Table 3 shows the regression settings and results. The formulation in Eqn. 6 was typed as the regression equation. The tool found a feasible solution to the BMP problem in 257 iterations at a learning rate of 0.1. Calculated values were $V_{max} = 822.46$ L, $R = 67.74$ day⁻¹ and $L = 0.103$ day with $R^2 = 0.998$ and $MSE = 91.9$. Fig. 4 shows the correlation plots of experimental and model results for cumulative biogas produced. Also, time series of experimental and model results are shown. The results indicated that the tool is able to handle nonlinear functions such as the transference function used in this work.

Table 3. Performance data of the implemented tool for test problem: anaerobic digestion

Definition of equation and global settings		
Parameter	Value	Remarks
Regression equation	$a1*(1-\exp(-a2*(a3-x1)/a1))$	$x1 = t$ (Eqn. 6); $a1 = V_{max}$ (Eqn. 6); $a2 = R$ (Eqn. 6); $a3 = L$ (Eqn. 6)
Maximum iteration	5000	
Maximum error	0.0001%	
Percent step size	0.01%	
Learning rate	0.1	
Estimated parameters and statistics		
Parameter	Value (this study)	
a_1	822.46 L	
a_2	67.74 day ⁻¹	
a_3	0.103 day	
Coefficient of determination (R^2)	0.998	
Mean square error (MSE)	91.9	

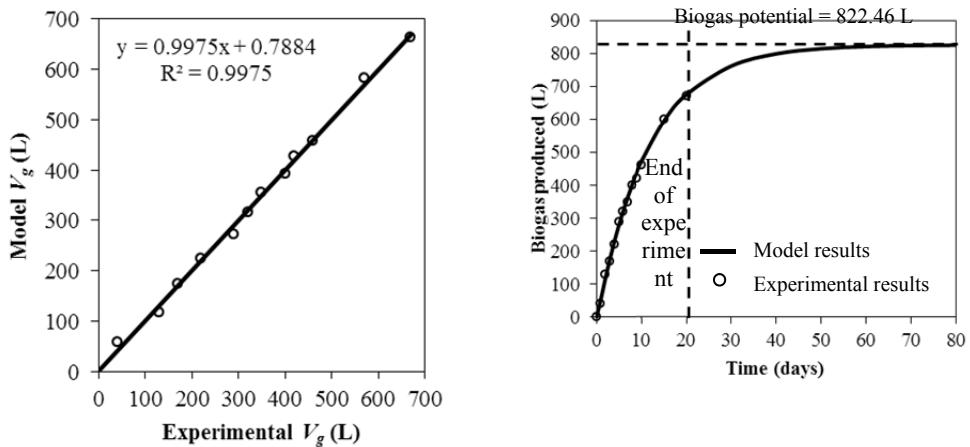


Figure 4. Experimental biogas production and results obtained by the implemented tool. **a.** correlation plot, **b.** time series data

4. CONCLUSIONS

An MS Excel VBA tool for general purpose multivariate nonlinear regression in environmental studies is presented in this study. The tool employs standard Gauss-Newton algorithm to minimize the cost function which is defined as the sum of squares of differences between target and predicted values of a dependent variable. The tool's performance is tested with a number of optimization problems that any environmental engineer encounter in his/her daily works. For each test problem, the tool completed calculations in less than 10 seconds. Calculated coefficients of determination were 0.98 and above and mean square errors were satisfactory. The results suggested that the MS Excel tool produces satisfactorily fast and reliable results, and can be used confidently for optimization works in environmental studies as well as for teaching purposes. The tool is an open-source, user-friendly VBA code and it is available on academic web pages of the authors.

REFERENCES

- [1] Li X., Huang Z., (2017) An inverted classroom approach to educate MATLAB in chemical process control, *Educ. Chem. Eng.* 19(1), 1-12.
- [2] Dehghanimohammadabadi M., Keyser T.K., (2017) Intelligent simulation: Integration of SIMIO and MATLAB to deploy decision support systems to simulation environment, *Simul. Modell. Pract. Theory* 71(45), 45-60.
- [3] Ku H., Fulcher R., (2012) Using computer software packages to enhance the teaching in Engineering Management Science – Part 2: Programming Techniques, *Comput. Appl. Eng. Educ.* 20(1), 114-123.
- [4] Orosa J.A., Oliveira A.C., (2012) Improvement in quality control for applications used by marine engineers, *Comput. Appl. Eng. Educ.* 20(1), 187-192.
- [5] Norton T., Tiwari B., (2013) Aiding the understanding of novel freezing technology through numerical modelling with visual basic for applications (VBA), *Comput. Appl. Eng. Educ.* 21(3), 530-538.
- [6] Niazkari M., Afzali S.H., (2016) Streamline performance of Excel in stewise implementation of numerical solutions, *Comput. Appl. Eng. Educ.* 24(4), 555-566.

- [7] Orosa Garcia J.A., (2011) Programming Languages for Marine Engineers, *Comput. Appl. Eng. Educ.* 19(3), 591-597.
- [8] Fernandes S., Orosa J.A., Galan J.J., (2012) A new methodology to teach numerical methods with MS Excel, *J. Mar. Res.* 9(2), 29-32.
- [9] Leon S.B., Ulfig R., Blanchard J., (1998) Teaching the numerical solution of ordinary differential equations using Excel 5.0, *Comput. Appl. Eng. Educ.* 4(2), 117-125.
- [10] Ferreira E.C., Salced, R., (2001) Can spreadsheet solvers solve demanding optimization problems?, *Comput. Appl. Eng. Educ.* 9(1), 49-56.
- [11] Rivas A., Gomez-Acebo T., Ramos J.C., (2006) The application of spreadsheets to the analysis and optimization of systems and processes in the teaching of hydraulic and thermal engineering, *Comput. Appl. Eng. Educ.* 14(4): 256-268.
- [12] Demir S., Yetilmezsoy K., Manav N., (2008) Development of a modified Hardy-Cross algorithm for time-dependent simulations of water distribution networks, *Fresen. Environ. Bull.* 17(8a), 1045-1053.
- [13] Orosa Garcia J.A., (2009) Programming languages for marine engineers, *Comput. Appl. Eng. Educ.* 19(3), 591-597.
- [14] Orosa J.A., Oliveira A.C., (2009) Improvement in Quality control for applications used by marine engineers, *Comput. Appl. Eng. Educ.* 20(1), 187-192.
- [15] Demir S., Saral A., Ertürk F., Kuzu L., (2010) Combined use of principal component analysis (PCA) and chemical mass balance (CMB) for source identification and source apportionment in air pollution modeling studies, *Water Air Soil Pollut.* 212, 429-439.
- [16] Demir S., Saral A., (2011) A new modification to the chemical mass balance receptor model for volatile organic compound source apportionment, *Clean – Soil Air Water*, 39(10), 891-899.
- [17] Fernandez S., Orosa J.A., Galan J.J., (2012) A new methodology to teach numerical methods with MS Excel, *Journal of Maritime Research* IX(2), 29-32.
- [18] Chebbi R., (2014) Chemical reactors sequencing. *Comput. Appl. Eng. Educ.* 22(2), 195-199.
- [19] Niazkar M., Afzali S.H., (2017) Analysis of water distribution networks using matlab and Excel spreadsheet: h-based methods, *Comput. Appl. Eng. Educ.* 25(1), 129-141.
- [20] Demir S., (2014) A practical model for estimating pressure drop in cyclone separators: An experimental study, *Powder Technol.* 268, 329-338.
- [21] Demir S., Karadeniz A., Manav-Demir N., (2016) Artificial neural network simulation of cyclone pressure drop: Selection of the best activation function, *Pol. J. Environ. Stud.* 25(5), 1891-1899.
- [22] Demir S., Karadeniz A., Manav-Demir N., (2016) Using steepness coefficient to improve artificial neural network performance for environmental modeling, *Pol. J. Environ. Stud.* 25(4), 1467-1477.
- [23] Painmanakul P., Wachirasak J., Jammongwong M., Hebrard G., (2009) Theoretical prediction of volumetric mass transfer coefficient (k_1a) for designing an aeration tank, *Engineering Journal* 13(3), 13-28.
- [24] Esposito G., Frunzo L., Liotta F., Panico A., Pirozzi F., (2012) Bio-methane potential tests to measure the biogas production from the digestion and co-digestion of complex organic substrates, *The Open Environmental Engineering Journal* 5, 1-8.
- [25] Donoso-Bravo A., Perez-Elvira S.I., Fdz-Polanco F., (2010) Application of simplified models for anaerobic biodegradability tests. Evaluation of pre-treatment processes, *Chem. Eng. J.* 160(2), 607-614.
- [26] Montecchio D., Gallipoli A., Gianico A., Mininni G., Pagliaccia P., Braguglia C.M., (2016) Biomethane potential of food waste: Modeling the effects of mild thermal pretreatment and digestion temperature, *Environ. Technol.* DOI: 10.1080/09593330.2016.1233293.