Parameter Estimation Using Spreadsheet Optimization: A Review of Applications in Civil and Environmental Engineering Education and Research

Lizette R. Chevalier[1], Shashi Marikunte[2], Jemil Yesuf[3]
Southern Illinois University Carbondale
Carbondale, IL 62901

Abstract - **The use of empirical models is an integral part of the design and analysis of civil and environmental engineering systems. One of the most commonly used methods for parameter estimation is linear regression. This technique is also applied to nonlinear models with two unknown parameters when the equation can be linearized through transformation. Equations with more than two unknown parameters can be solved using matrices that expand on the basic form of the solution for linear equations. The use of spreadsheets in engineering education and research is a common tool used for basic curve fitting (two unknown parameters) as well as smaller matrix solutions. This paper presents a different approach for the use of spreadsheets, namely, parameter estimation using optimization. Spreadsheet optimization is a practical platform that easily determines the parameters of equations. The number of unknown variables is not constrained. In addition, the analysis can be expanded to include minimizing a range of nonlinear error functions, consequently allowing for a broader mathematical approach to parameter estimation. This paper reviews the use of spreadsheet optimization on several civil and environmental engineering applications.**

Index Terms - Spreadsheets, optimization, sorption isotherms, weir equation, strength of concrete

INTRODUCTION

In the undergraduate engineering curriculum, the traditional approach to parameter estimation of nonlinear equations is to first linearize the data through transformation, followed by linear regression. This technique is also used in industry in numerous applications. Linear regression assumes a Gaussian distribution of the error at each point. However, linearization of the data implicitly alters this error structure and may also violate the error variance and normality assumption [1]-[3]. Furthermore, this technique is limited to the estimation of two variables.

An alternative method for parameter estimation is optimization based on minimizing the error. This approach provides two significant advantages. First, it is applicable to equations with more than two parameters. Secondly, the method allows for a choice of an error function and consequently a broader mathematical approach to parameter estimation. In this study, optimization was performed using the solver add-in of Microsoft Excel®.

Our investigation is focused on curriculum commonly taught in the civil and environmental engineering courses: calibrating flow over a weir, dye sorption using the Langmuir isotherm model, dye sorption using the Redlich-Peterson isotherm and estimating the strength of concrete (materials). The Redlich-Peterson isotherm and the equation for estimating the strength of concrete represent equations with more than two unknown parameters.

MATHEMATICAL BACKGROUND

The basic mathematical equations evaluated were:

$$y = a_o x^{a_1} \tag{1a}$$

$$y = \frac{a_o x}{1 + a_1 x} \tag{2a}$$

$$y = \frac{a_o x}{1 + a_1 x^{a_2}} \tag{3}$$

$$y = a_o + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + .. \tag{4a}$$

where y is the dependent variable, x is the independent variable(s) and a_n represent the unknown coefficients or fitting parameters. Equations 1 and 2 can be linearized using the following transformations:

$$\ln y = \ln a_o + a_1 x \tag{1b}$$

$$\frac{1}{y} = \left(\frac{1}{a_o}\right)\frac{1}{x} + \frac{a_1}{a_0} \tag{2b}$$

Because Eqn. 3 has three parameters, it cannot be solved using linearization techniques. Equation 4 can be solved using a solution technique based on the basic principles of linear

---

[1] Lizette R. Chevalier, Chair, Dept. of Civil and Environmental Engineering, cheval@engr.siu.edu
[2] Shashi Marikunte, Asst. Professor, Dept. of Civil and Environmental Engineering, marikunt@engr.siu.edu
[3] Jemil Yesuf, Graduate Student, Dept. of Civil and Environmental Engineering, jemiln74@yahoo.com

regression, namely, minimizing the error by setting the partial derivative of the coefficients equal to zero. A generalized solution, which can be expanded to include additional terms, is:

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{Bmatrix} a_o \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{Bmatrix} \quad (4b)$$

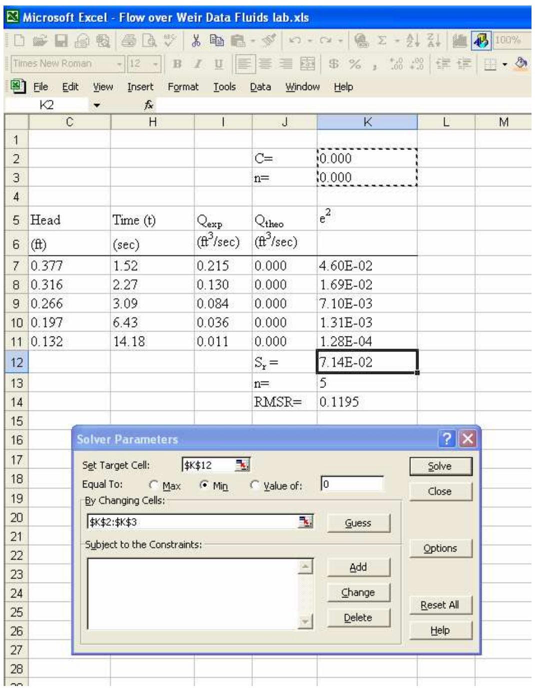

FIGURE 1
EXAMPLE OF USING MICROSOFT EXCEL® SOLVER FOR PARAMETER ESTIMATION

FLOW OVER A WEIR

In civil engineering practice, it is necessary to determine the flow rate of streams and similar open channels for water resource management or to forecast floods. One method of measuring flowrate is the weir structure. A popular weir is the V-notch weir, named for it's V-shaped crest.

The basic principle of a weir is that flowrate (Q) is directly related to the water depth (H) above the weir crest. Mathematically, this relationship is written as:

$$Q = CH^n \quad (5)$$

where C is the coefficient of discharge which is a function of the notch angle θ, and n is a weir constant. The basic form of this equation and the transformation is (1a) and (1b).

An alternative to this approach is parameter estimation using the solver add-in of Microsoft Excel®. An example worksheet for this method is presented in Figure 1. Cell K12 is the sum of the residual error, S_r. The residual error (e^2) is the square of the difference between Q_{exp} and Q based on (5). This value is minimized by changing the values of C and n. A comparison of parameter estimation based on both techniques is presented in Figure 2 and Table 1. As evident by the graph, parameter estimation based on spreadsheet optimization is a better fit model for the data. When the data was linearized, the R^2 value of the linear regression was 0.9996. This linearized R^2 value is often reported as evidence of a good fit. As evident by Figure 2, the parameters from this method do not provide a close fitting model.

An appropriate quantitative measure of a good fit for non-linear equations is the root mean squared residual (RMSR):

$$RMSR = \sqrt{\frac{\sum \left(y_i - y_{i,model}\right)^2}{N}} \quad (6)$$

where N is the number of data points. Values close to zero indicate a good fit. In this case, the linear transform model (LTRM) resulted in a higher RMSR than the optimization model (Table 1), which is in agreement with Figure 2.

TABLE 1.
COMPARISON OF THE PARAMETER ESTIMATION METHODS FOR A V-NOTH WEIR

	C	n	R^2	RMSR
LTRM	3.322	2.797	0.9996	0.0069
Optimization	3.101	2.741	n/a	0.0013

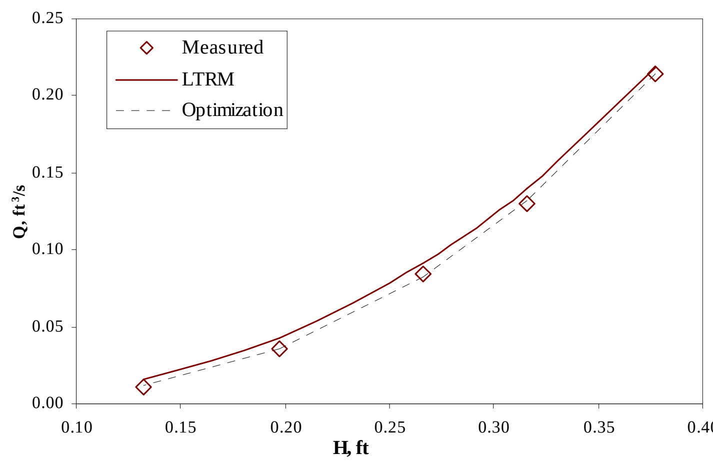

FIGURE 2
COMPARISON OF EXPERIMENTAL AND MODELED DATA FOR FLOW OVER A WEIR.

LANGMUIR ISOTHERM

The design and operation of adsorption processes require equilibrium adsorption data for use in mass transfer models which can then be used to predict the performance of the adsorption contact processes under a range of operating conditions [4]. Mostly, these adsorption studies are performed by equilibrating predetermined quantities of adsorbent with solutions of adsorbate(s) at fixed pH and temperature. Graphs of resulting equilibrium data correlating the variation of solid-phase concentration (q_e), or the amount of solute adsorbed per unit mass of solid, to the variation of equilibrium solution-phase concentration (C_e) are termed adsorption isotherms. Langmuir [5] developed a theoretical equilibrium isotherm model that is widely used in industry and research:

$$q_e = \frac{K_L C_e}{1 + a_L C_e} \qquad (7)$$

where K_L is given as:

$$K_L = Q_o a_L \qquad (8)$$

Q_o ($mg \cdot g^{-1}$) is the solid-phase concentration corresponding to a condition in which all sites are filled or maximum adsorption capacity attained, and a_L ($L \cdot mg^{-1}$) is a constant related to the net enthalpy of adsorption. This equation is in the form of (2a), and can be transformed into a linear equation in the from of (2b)

As with the previous example, we estimated the coefficients using both linear regression the solver add-in of Microsoft Excel® by minimizing the sum of the residual errors. The data used was from a batch study evaluing the sorption of methylene blue onto activated carbon. The initial concentration of adsorbate (methylene blue) ranged from 50 – 1400 $mg \cdot L^{-1}$ while a fixed mass of adsorbent (0.25g of filtrasorb 400) was used to set up this experiment.

The results are presented in Table 2 and Figure 3. Visual examination of the graph shows that parameter estimation based on spreadsheet optimization is a better fit for the data. When the data was linearized, the R^2 value of the linear regression was 0.9995, but RMSR based on the actual data was 22.272. The RMSR for the spreadsheet optimization method is 17.187, which indicates a better fit. This is consistent with the results of the previous example. We are not suggesting that this level of improvement would be evident in all systems, however, this does support the use of teaching this method in the undergraduate curriculum and utilizing it for model development in industry and research.

TABLE 2
COMPARISON OF THE PARAMETER ESTIMATION METHODS FOR THE LANGMUIR ISOTHERM

	K_L	a_L	R^2	RMSR
LTRM	34.01	0.126	0.9995	22.272
Optimization	21.77	0.081	n/a	17.187

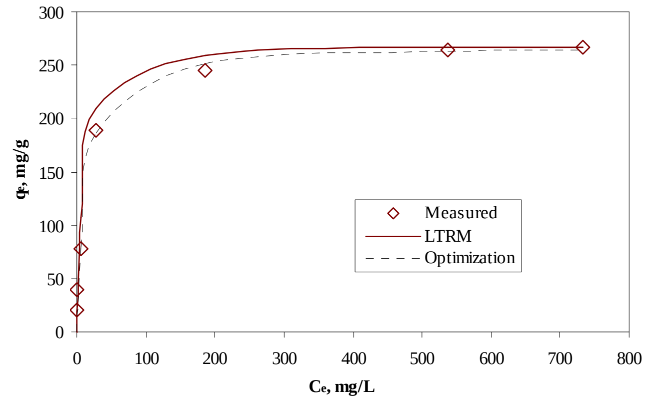

FIGURE 3
COMPARISON OF EXPERIMENTAL AND MODELED DATA FOR LANGMUIR ISOTHERM

REDLICH-PETERSON ISOTHERM

Redlich and Peterson [6] proposed an empirical equation, designated the "three parameter equation," which may be used to represent adsorption equilibria over a wide concentration range:

$$q_e = \frac{K_R C_e}{1 + a_R C_e^{b_R}} \qquad (9)$$

where K_R, a_R, and b_R are Redlich-Peterson parameters. This equation reduces to a linear isotherm at low surface coverage, and to the Langmuir isotherm when $b_R = 1$.

The two models considered so far have all been two parameter equations that employed linear transformation in order to determine and evaluate the quality of the model constants. However, Equation (3c) cannot be linearized to produce the values of the three parameters. Instead, parameter estimation was only determined using spreadsheet optimization. The experimental data used to evaluate this model is identical to that used for the Langmuir isotherm evaluation.

The graphical results of parameter estimation of the Redlich-Peterson isotherm are presented in Figure 4. The RMSR of the fit is 17.147, the resulting parameters for K_R, a_R and b_R were 22.852, 0.091 and 0.989.

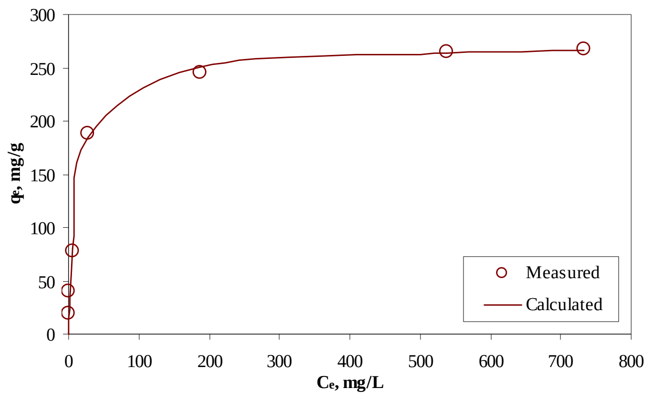

FIGURE 4
COMPARISON OF EXPERIMENTAL AND MODELED DATA FOR THE REDLICH-PETERSON ISOTHERM

STRENGTH OF CONCRETE

Concrete has been extensively used in the construction industry since the beginning of the 20th century. Concrete is a highly complex heterogeneous material. The properties of concrete can be widely varied depending on the appropriate selection and proportioning of constituent materials. The response of the material to stress depends on the mix ingredients and their interaction. The response to compressive forces is considered to be the most important property of concrete. Two major factors that are critical for concrete performance are the material matrix and water content [6]. Supplementary cementious materials, such as silica fume, fly ash and metakaolin are often partially substituted with cement in concrete to produce a high-performance concrete by improving the bond between the aggregate and the cement paste [7]. A decrease in the water to cement (w/c) ratio results in increased strength and reduced porosity in cement paste and hence the concrete becomes stronger and impermeable.

While, relatively small number of data points may be sufficient to arrive at general conclusions, a comprehensive experimental design with sufficient data points and strong statistical tools are needed to develop accurate models to predict the effect of individual response on the strength of concrete. Previous research [8] determined that the model that best fits the analyzed data is presented by the following quadratic equation:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_2^2 \qquad (10)$$

where y is the compressive strength in MPa, x_1 is the water to binder ratio and x_2 is the silica fume content. Table 3 summarizes the results of the parameter estimation of (10) using regression as well as spreadsheet optimization. Figure 5 presents a comparison of the raw data with the two models. As found with the other models evaluated, the fit using spreadsheet optimization provided a better fit, as evidence from the RMSR of 17.71 compared to 35.72, and visual inspection of the figure.

TABLE 3
COMPARISON OF MODELS FOR THE STRENGTH OF CONCRETE

	a_0	a_1	a_2	a_3	RMSR
Regression	180.20	-378.24	2.79	-0.09	35.72
Optimization	182.65	-354.90	2.79	-0.12	17.71

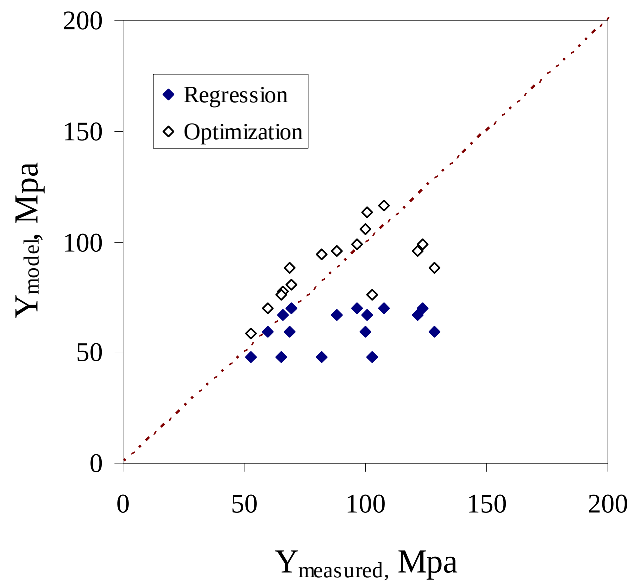

FIGURE 5
COMPARISON OF SPREADSHEET OPTIMIZATION MODEL TO MEASURED DATA.

SUMMARY

This paper presents several studies in civil and environmental engineering where spreadsheet optimization was used for parameter optimization. In three of the four cases, this method performed better than regression techniques. The comparisons included an example of multiple regression with two independent variables. In the fourth case, we demonstrated the use of this method for a three parameter equation. As stated previously, we are not suggesting that parameter estimation using spreadsheet optimization will always provide a better fit to the data. However, the results do support the use of teaching this method in the undergraduate curriculum and utilizing it for model development in industry and research.

REFERENCES

[1] Myers, R.H., "Classical and Modern Regression With Applications", *PWS-Kent*, 1990.

[2] Ratkowski, D.A., "Handbook Of Nonlinear Regression Models", *Marcel Dekker*, 1990.

[3] Porter, J.F., Mckay, G., and Choy, K.H., "The prediction of sorption from a binary mixture of acidic dyes using single and mixed-isotherm variants of the ideal adsorbed solute theory". *Chemical Engineering Science,* 54**,** 1999, 5863-5885.

[4] Weber, W. J. Jr. and DiGiano, F.A. Process dynamics in environmental systems. John Wiley and Sons. 1996.

[5] Langmuir, I., "The adsorption of gases on plane surface of glass, mica and platinum", *J. Am. Chem. Soc*. 40, 1916,1361-1368.

[6] Redlich, O and Peterson, D.L., "A useful adsorption isotherm", *J. Phys. Chem*, 63, 1959, 1024.

[7] S. Mindess, J.F. Young, D.C. Darwin, "Concrete," 2nd edition, Prentice Hall, New Jersey, 2002.

[8] S.P. Shah, S.H. Ahmad, "High Performance Concretes and Applications," Edward Arnold, London, 1994.

[9] Samir Nacer, "Optimization of Silica Fume Content and Water to Enhance Performance of Concrete," Master's Thesis, Dept. of Civil and Environmental Engineering, Southern Illinois University Carbondale, Dec. 2005.