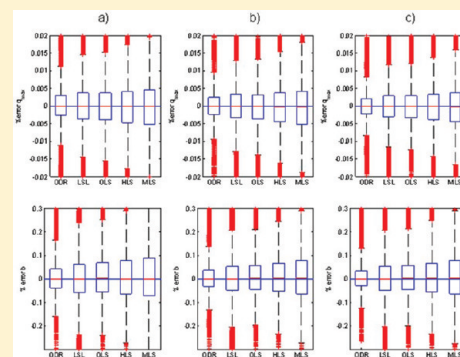


Orthogonal Distance Regression: A Good Alternative to Least Squares for Modeling Sorption Data

Jordi Poch^{*,†} and Isabel Villaescusa^{‡,§}

[†]Department of Applied Mathematics, [‡]Department of Chemical Engineering, Escola Politècnica Superior, Universitat de Girona, Avda. Lluís Santaló, s/n, 17071 Girona, Spain

ABSTRACT: The most used regression methods in sorption studies to estimate the isotherm parameters (least squares linearized, ordinary least squares, Marquardt's least squares (MLS), and hybrid least squares) and orthogonal distance regression (ODR) have been compared. Theoretical Langmuir isotherms were built from different selected values of q_{\max} and b , and from them simulated isotherms were generated by introducing a certain error. With the generated data the corresponding isotherm parameters were estimated by using the different regression methods and their values were compared to the ones of the theoretical isotherms. The results of this study show that ODR gives the most accurate estimates of the isotherm parameters when the theoretical data are perturbed with a fixed error. When the theoretical data are perturbed with an error proportional to concentration, ODR gives also accurate estimates, but they are similar to those obtained with the MLS method.



1. INTRODUCTION

In the past 20 years biosorption processes have been investigated as an alternative method to the conventional water treatment technologies for the removal of metal ions¹ and organic pollutants² from wastewaters. Sorption data provide the capacity and the selectivity of a sorbent for a given component (sorbate). Capacity is defined as the number of moles of a sorbate sorbed per gram of a sorbent and selectivity governs the feasibility of a separation by sorption. Both capacity and selectivity affect sorption equilibrium achievement. Therefore, accuracy in the description of the equilibrium isotherms is of great importance for the simulation and the design of sorption processes.

In order to obtain a sorption isotherm, the remaining sorbate concentration in solution at equilibrium (C_e) is determined experimentally by chemical analysis and the amount of sorbate sorbed (q_e) is determined by the difference from the sorbate initial concentration (C_0) also determined experimentally. Therefore, C_e and q_e are affected by the measurement errors.

In most of the sorption studies reported in literature authors use some of the available isotherm models (Langmuir, Freundlich, and Redlich-Peterson)³ and calculate the isotherm parameters by using regression methods.^{4–8} In general, authors use different function errors and make their decision on the model which best fit the experimental data by the comparison of R^2 (linear regression) or sum square residuals (SSR) (nonlinear regression) or others.

Nevertheless, the use of least-squares regression involves the assumption that the independent variable (C_e) is well-known with no error associated and as said before this variable is determined experimentally and therefore it may contain an error. Therefore, the magnitude of the errors made in the measurement and determination of both variables (C_e and q_e)

and the regression method used affect the isotherm parameters determination.

When the absence of error in x variable cannot be guaranteed, other regression methods such as orthogonal distance regression (ODR) should be used.⁹

The main objective of the present study is to determine whether the orthogonal distance regression is a more suitable method than the other methods to determine the isotherm parameters. For this purpose, the Langmuir isotherm, one of the most used models in biosorption studies, was chosen in order to investigate the effect of experimental error on the isotherm parameters determination with regard to the regression method used and the number of experimental data. For this purpose, theoretical isotherms were built from different selected values of q_{\max} and b and from them different sets of simulated data (isotherms) were generated by introducing a certain error. With the generated data the corresponding isotherm parameters were obtained and compared to the ones of the theoretical isotherms.

2. METHODS

2.1. Theoretical Isotherm and Simulated Data Generation. The used theoretical isotherms (Langmuir-type) have been built from eq 1

$$q_{ei} = \frac{q_{\max} b C_{ei}}{1 + b C_{ei}} \quad (1)$$

Received: October 4, 2011

Accepted: December 14, 2011

Published: January 13, 2012

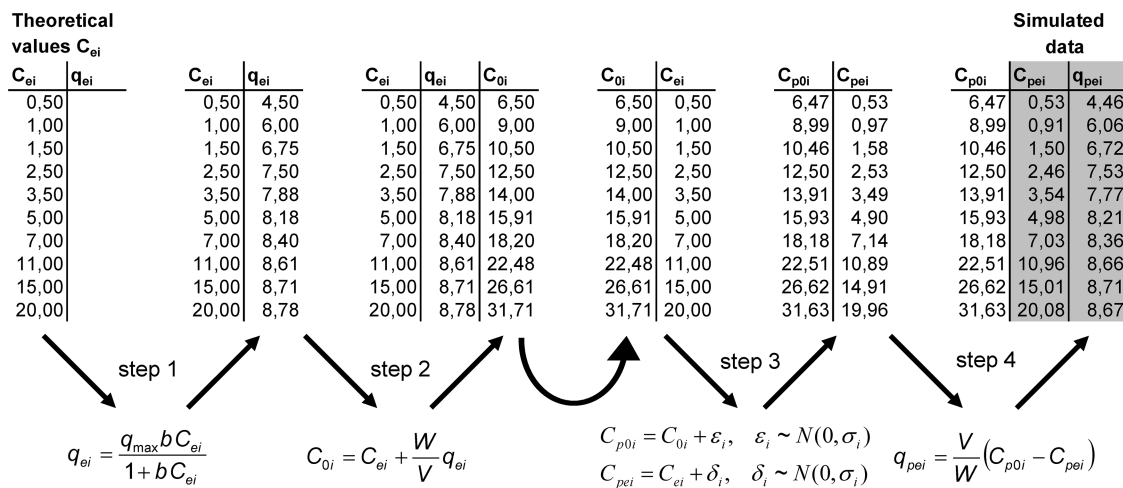


Figure 1. Schema of a set of simulated data generation. Example for a chosen pair of $q_{\max} = 9$ and $b = 2$ values. Perturbation: fixed error.

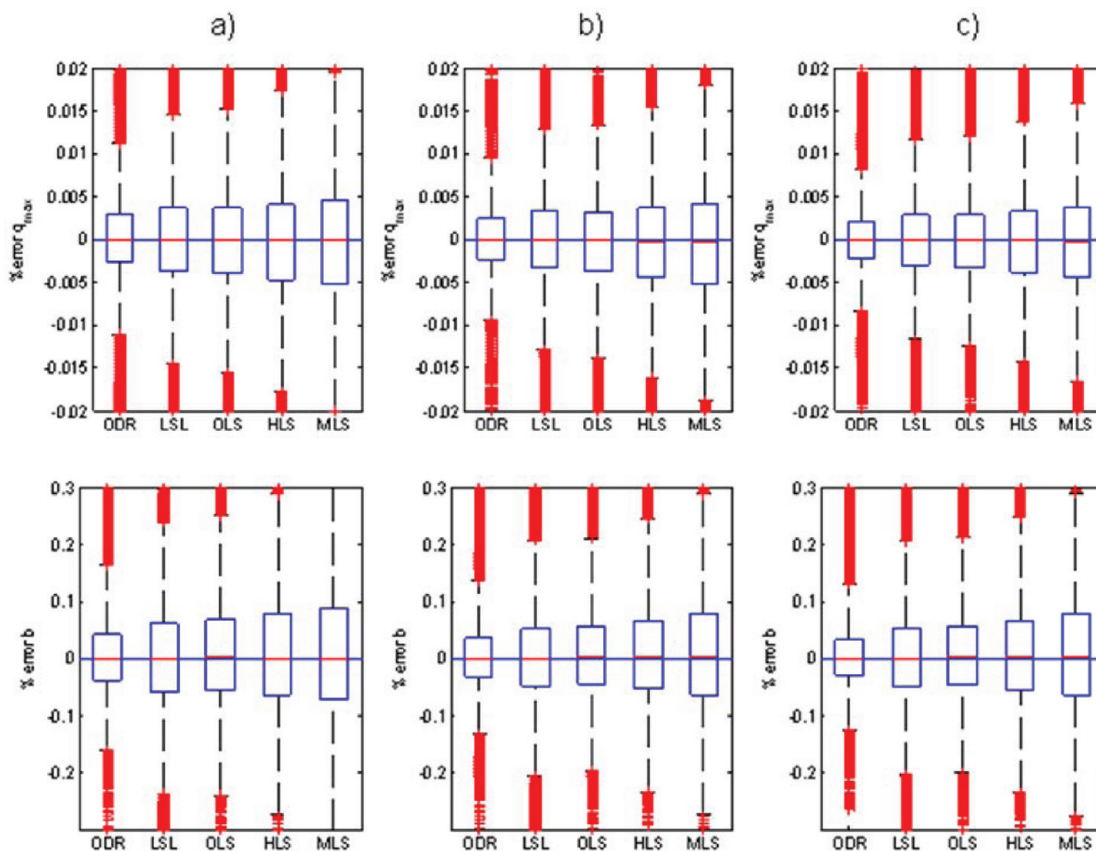


Figure 2. Box plots for the error in estimation of q_{\max} and b by different methods of regression of all simulated adsorption equilibrium data. Isotherm data points: (a) 7, (b) 10, and (c) 13 points. Perturbation: fixed error.

and the following reference values of q_{\max} (5, 7, 9, 11, 13, and 15) and of b values (0.5, 2, 4, 6, and 12).

C_{ei} values within the range (0.5–20) were chosen to generate experimental data sets.

In Figure 1, the different steps to generate a set of simulated data for a chosen pair of q_{\max} and b values are described.

The first step (step 1) consisted on choosing a pair values q_{\max} and b from the above indicated reference values. Then, q_{ei} was calculated for each C_{ei} of the chosen values by using the isotherm eq 1.

In step 2 the initial concentration C_{0i} corresponding to the q_{ei} and C_{ei} were calculated by using the equation

$$C_{0i} = C_{ei} + \frac{W}{V} q_{ei} \quad (2)$$

where W is the sorbent mass and V is the volume of the solution. For this study W and V values were 0.2 and 0.15, respectively.

Step 3 consisted in introducing errors on C_{0i} and C_{ei} data by following normal distributions of mean zero and variance σ to obtain perturbed data of C_{0i} and C_{ei} (C_{p0i} and C_{pei}).

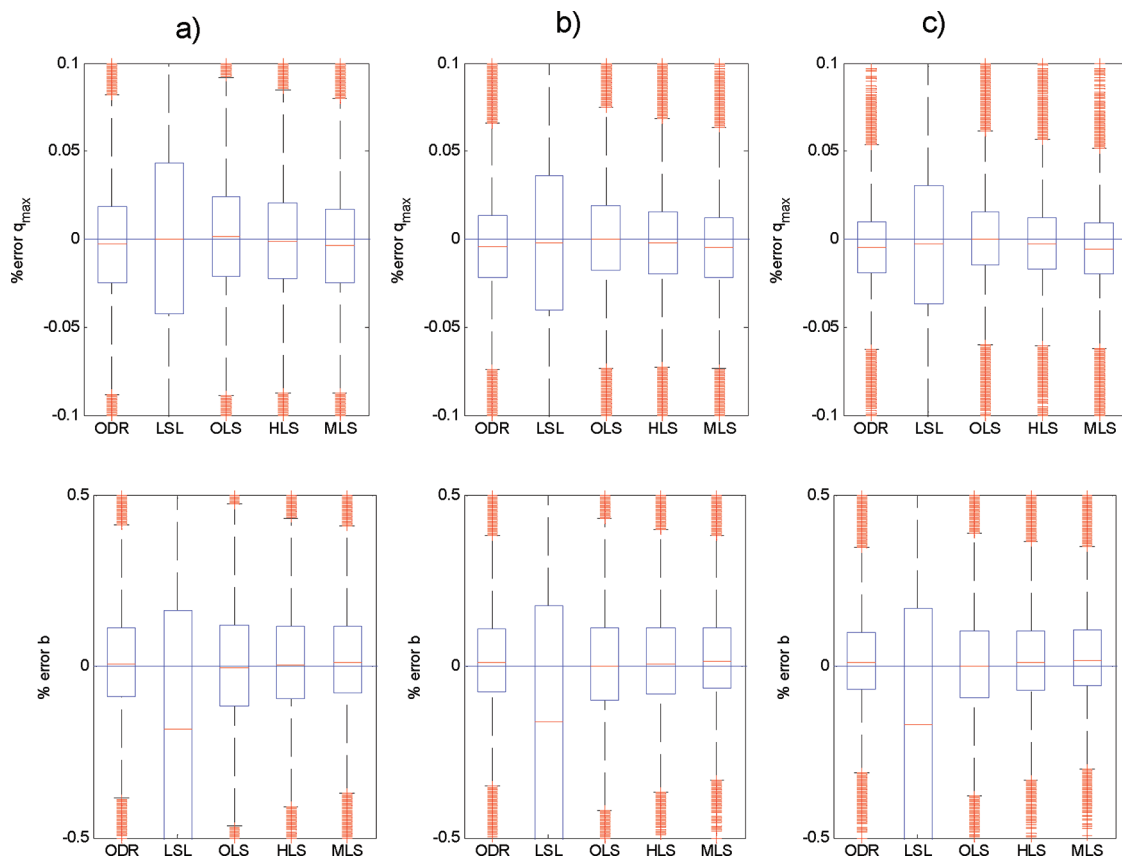


Figure 3. Box plots for the error in estimation of q_{\max} and b by different methods of regression of all simulated adsorption equilibrium data. Isotherm data points: (a) 7, (b) 10, and (c) 13 points. Perturbation: error proportional to concentration.

Table 1. Statistics of Errors in Estimation of q_{\max} and b by Different Methods of Regression^a

	q_{\max} calculated values				b calculated values			
	mean	std	median	iqr	mean	std	median	iqr
5 % constant error (13 equilibrium data)								
ODR	-3.03×10^{-5}	4.03×10^{-3}	-2.81×10^{-5}	4.18×10^{-3}	4.31×10^{-3}	6.37×10^{-2}	6.44×10^{-4}	6.44×10^{-2}
LSL	-3.11×10^{-5}	5.31×10^{-3}	-1.08×10^{-5}	5.80×10^{-3}	1.71×10^{-2}	7.98×10^{-1}	-7.27×10^{-4}	1.02×10^{-1}
OLS	-2.99×10^{-4}	5.29×10^{-3}	-1.62×10^{-4}	6.10×10^{-3}	1.26×10^{-2}	9.63×10^{-2}	2.16×10^{-3}	1.03×10^{-1}
HLS	-3.53×10^{-4}	6.87×10^{-3}	-1.76×10^{-4}	7.05×10^{-3}	1.32×10^{-2}	1.04×10^{-1}	1.71×10^{-3}	1.21×10^{-1}
MLS	-4.47×10^{-4}	1.03×10^{-2}	-2.24×10^{-4}	8.11×10^{-3}	1.43×10^{-2}	1.13×10^{-1}	1.21×10^{-3}	1.42×10^{-1}
5 % constant error (10 equilibrium data)								
ODR	1.02×10^{-6}	4.53×10^{-3}	3.82×10^{-6}	4.76×10^{-3}	4.35×10^{-3}	6.60×10^{-2}	4.66×10^{-4}	6.71×10^{-2}
LSL	1.56×10^{-5}	5.81×10^{-3}	2.65×10^{-7}	6.42×10^{-3}	2.04×10^{-2}	3.59×10^{-1}	-6.57×10^{-4}	1.03×10^{-1}
OLS	-3.21×10^{-4}	5.85×10^{-3}	-2.08×10^{-4}	6.86×10^{-3}	1.30×10^{-2}	9.57×10^{-2}	2.59×10^{-3}	1.02×10^{-1}
HLS	-3.65×10^{-4}	7.54×10^{-3}	-2.31×10^{-4}	7.96×10^{-3}	1.37×10^{-2}	1.03×10^{-1}	2.57×10^{-3}	1.20×10^{-1}
MLS	-4.60×10^{-4}	1.13×10^{-2}	-2.82×10^{-4}	9.23×10^{-3}	1.48×10^{-2}	1.13×10^{-1}	2.70×10^{-3}	1.41×10^{-1}
5 % constant error (7 equilibrium data)								
ODR	4.51×10^{-5}	5.28×10^{-3}	3.47×10^{-5}	5.58×10^{-3}	4.86×10^{-3}	7.58×10^{-2}	7.63×10^{-4}	8.14×10^{-2}
LSL	3.14×10^{-5}	6.71×10^{-3}	2.03×10^{-5}	7.26×10^{-3}	1.15×10^{-2}	1.47	-8.25×10^{-4}	1.20×10^{-1}
OLS	-2.26×10^{-4}	6.81×10^{-3}	-5.89×10^{-5}	7.73×10^{-3}	1.27×10^{-2}	1.08×10^{-1}	1.19×10^{-3}	1.22×10^{-1}
HLS	-3.03×10^{-4}	8.95×10^{-3}	-7.13×10^{-5}	8.78×10^{-3}	1.36×10^{-2}	1.16×10^{-1}	1.04×10^{-3}	1.42×10^{-1}
MLS	-4.45×10^{-4}	1.29×10^{-2}	-9.93×10^{-5}	9.91×10^{-3}	1.50×10^{-2}	1.27×10^{-1}	2.32×10^{-4}	1.62×10^{-1}

^aPerturbation: fixed error.

The perturbed data C_{p0i} and C_{pei} were used to calculate the perturbed q_{pei} through the equation

$$q_{pei} = \frac{V}{W}(C_{p0i} - C_{pei}) \quad (3)$$

and obtain the set of simulated data (step 4).

Two different ways of introducing errors were used in step 3: (i) a fixed error and (ii) an independent error proportional to C_{0i} and C_{ei} . The introduced fixed error follows a normal distribution of mean zero and variance 0.05; the proportional error follows a normal distribution of mean zero and variance 5 % for each C_{0i} and C_{ei} .

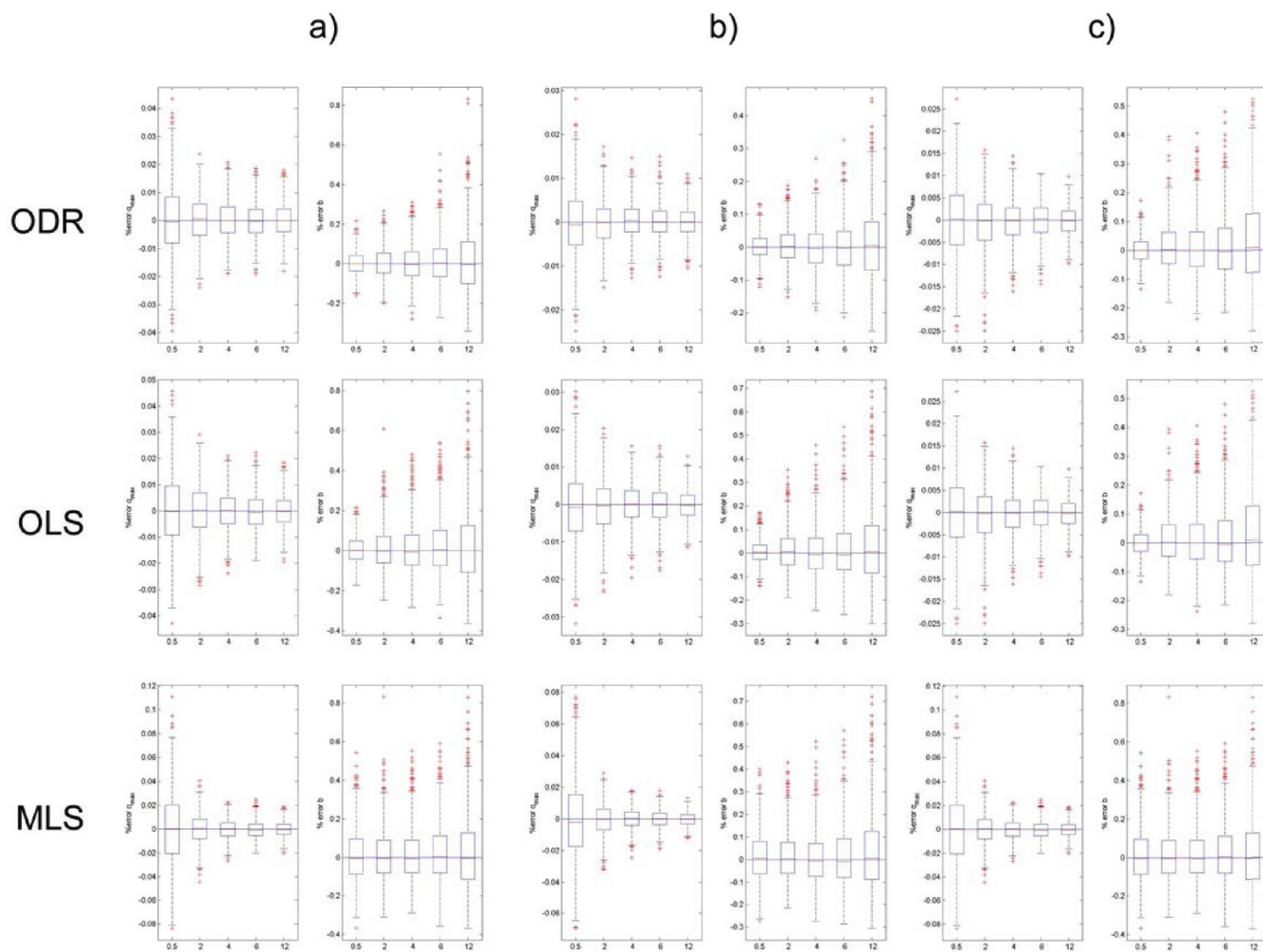


Figure 4. Box plot for the error in estimation of q_{\max} and b as a function of b values for different methods of regression. Perturbation: fixed error; number of isotherm data points: 7; q_{\max} values: (a) 5, (b) 9, and (c) 15.

Table 2. Statistics of Errors in Estimation of q_{\max} and b by Different Methods of Regression^a

	q_{\max} calculated values				calculated values			
	mean	std	median	iqr	mean	std	median	iqr
5 % proportional error (13 equilibrium data)								
ODR	-5.35×10^{-3}	2.50×10^{-2}	-4.62×10^{-3}	2.90×10^{-2}	3.05×10^{-2}	1.78×10^{-1}	1.20×10^{-2}	1.64×10^{-1}
LSL	-3.27×10^{-3}	5.50×10^{-2}	-2.81×10^{-3}	6.77×10^{-2}	-3.42×10^{-1}	4.63×10^1	-1.70×10^{-1}	7.58×10^{-1}
OLS	9.28×10^{-4}	2.80×10^{-2}	1.67×10^{-4}	3.04×10^{-2}	1.87×10^{-2}	1.87×10^{-1}	1.65×10^{-4}	1.92×10^{-1}
HLS	-2.69×10^{-3}	2.57×10^{-2}	-2.70×10^{-3}	2.93×10^{-2}	3.19×10^{-2}	1.86×10^{-1}	1.03×10^{-2}	1.74×10^{-1}
MLS	-6.05×10^{-3}	2.45×10^{-2}	-5.30×10^{-3}	2.85×10^{-2}	4.49×10^{-2}	1.92×10^{-1}	1.86×10^{-2}	1.62×10^{-1}
5 % proportional error (10 equilibrium data)								
ODR	-4.91×10^{-3}	3.05×10^{-2}	-4.28×10^{-3}	3.50×10^{-2}	3.32×10^{-2}	2.01×10^{-1}	1.02×10^{-2}	1.83×10^{-1}
LSL	-1.71×10^{-3}	6.30×10^{-2}	-1.81×10^{-3}	7.67×10^{-2}	-4.32×10^{-1}	5.36×10^1	-1.64×10^{-1}	7.39×10^{-1}
OLS	1.15×10^{-3}	3.43×10^{-2}	1.20×10^{-4}	3.70×10^{-2}	2.20×10^{-2}	2.08×10^{-1}	2.92×10^{-4}	2.13×10^{-1}
HLS	-2.39×10^{-3}	3.15×10^{-2}	-2.40×10^{-3}	3.55×10^{-2}	3.37×10^{-2}	2.05×10^{-1}	8.34×10^{-3}	1.92×10^{-1}
MLS	-5.69×10^{-3}	2.98×10^{-2}	-4.90×10^{-3}	3.44×10^{-2}	4.57×10^{-2}	2.11×10^{-1}	1.59×10^{-2}	1.79×10^{-1}
5 % proportional error (7 equilibrium data)								
ODR	-3.60×10^{-3}	3.65×10^{-2}	-3.00×10^{-3}	4.25×10^{-2}	3.36×10^{-2}	2.19×10^{-1}	5.91×10^{-3}	2.00×10^{-1}
LSL	5.92×10^{-4}	7.02×10^{-2}	-2.33×10^{-4}	8.56×10^{-2}	-4.30×10^{-1}	5.35×10^1	-1.84×10^{-1}	7.50×10^{-1}
OLS	2.63×10^{-3}	4.16×10^{-2}	1.27×10^{-3}	4.53×10^{-2}	2.23×10^{-2}	2.31×10^{-1}	-3.32×10^{-3}	2.36×10^{-1}
HLS	-9.62×10^{-4}	3.80×10^{-2}	-1.21×10^{-3}	4.32×10^{-2}	3.29×10^{-2}	2.27×10^{-1}	4.24×10^{-3}	2.11×10^{-1}
MLS	-4.35×10^{-3}	3.57×10^{-2}	-3.74×10^{-3}	4.18×10^{-2}	4.45×10^{-2}	2.34×10^{-1}	1.07×10^{-2}	1.96×10^{-1}

^aPerturbation: error proportional to concentration.

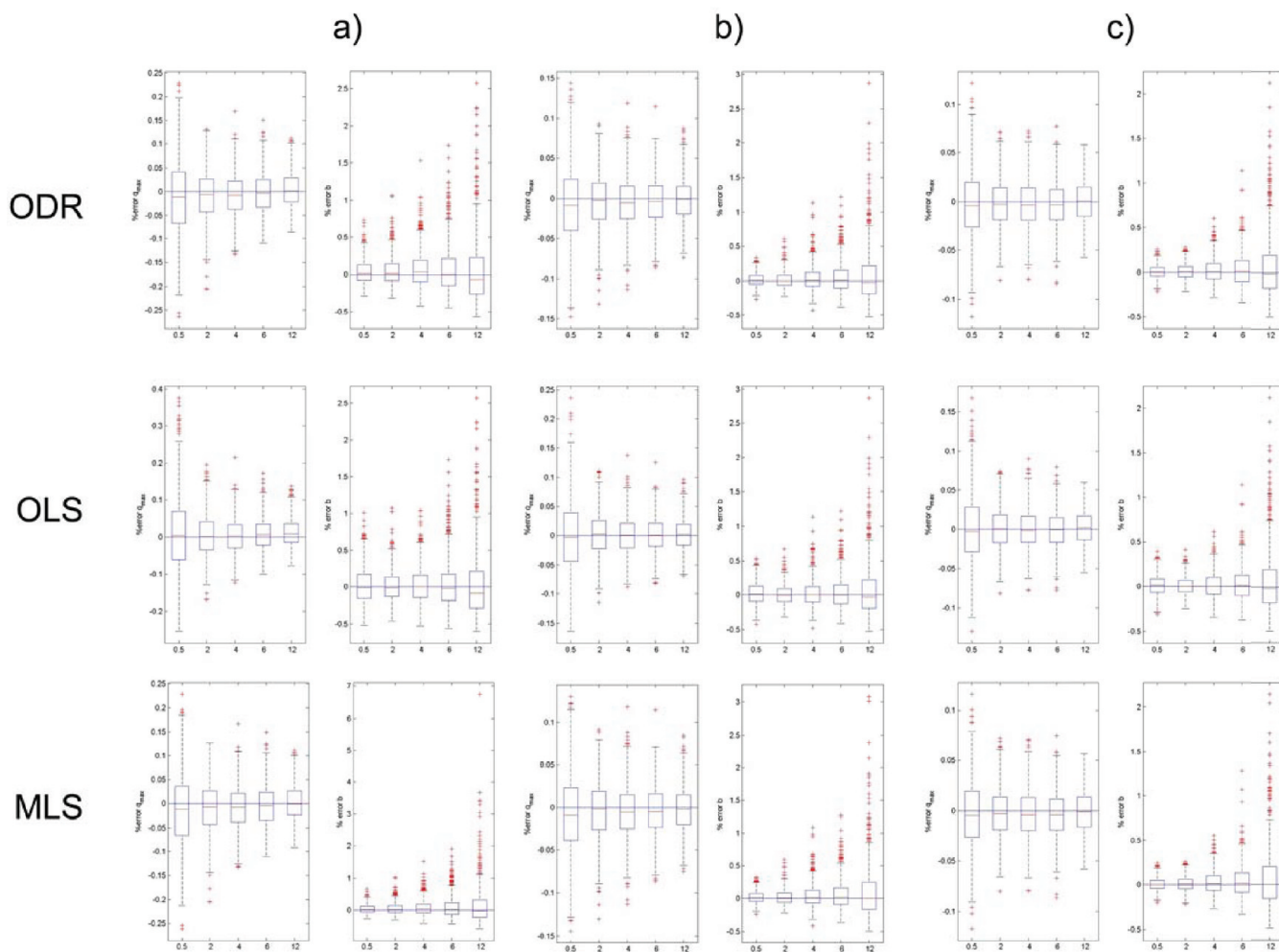


Figure 5. Box plot for the error in estimation of q_{\max} and b as a function of b values for different methods of regression. Perturbation: error proportional to concentration; number of isotherm data points: 7; q_{\max} values: (a) 5, (b) 9, and (c) 15.

Table 3. Interquartile Ranges for % Error in Estimation of q_{\max} and b for a Fixed q_{\max} ^a

model	interquartile ranges			
	% error in q_{\max}		% error in b	
	q_{\max} 9, b 0.5	q_{\max} 9, b 12	q_{\max} 9, b 0.5	q_{\max} 9, b 12
ODR	0.0099	0.0044	0.0491	0.1462
OLS	0.0127	0.0054	0.0605	0.2003
MLS	0.0328	0.0057	0.1413	0.2125

^aPerturbation: fixed error.

Table 4. Interquartile Ranges for % Error in Estimation of q_{\max} and b for a Fixed q_{\max} ^a

model	interquartile ranges			
	% error in q_{\max}		% error in b	
	q_{\max} 9, b 0.5	q_{\max} 9, b 12	q_{\max} 9, b 0.5	q_{\max} 9, b 12
ODR	0.0643	0.0345	0.2030	0.3986
OLS	0.0830	0.0354	0.3280	0.4045
MLS	0.0615	0.0344	0.1900	0.4111

^aPerturbation: error proportional to concentration.

In the example of Figure 1, the selected isotherm parameters were $q_{\max} = 9$ and $b = 2$ and a fixed error was applied.

By following the described procedure in Figure 1, 1000 sets of simulated data were generated for each pair of q_{\max} and b .

The influence of the value of q_{\max} and b on each other has been investigated by fixing a q_{\max} value and varying the value of b and vice versa. The influence of number of data points on the regression model fit was also investigated. For this purpose, isotherms simulations were done by using 7 (0.5, 1, 3, 4, 7, 15, and 20), 10 (0.5, 1, 1.5, 2.5, 3.5, 5, 7, 11, 15, and 20), and 13 (0.5, 1, 1.5, 2.5, 3.5, 4.25, 5, 7, 9, 11, 13, 15, and 20) equilibrium points.

2.2. Regression Analysis. In the single-component isotherm studies, the objective consists of adjusting the parameters of a model function to best fit a data set. A simple data set consists of n points (x_i, y_i) , $i = 1, \dots, n$, where $x_i(C_e)$ is an independent variable and $Y_i(q_e)$ is a dependent variable whose value is found by observation. The model function has the form

$$Y_i = f(x_i, \beta) + \varepsilon_i \quad (4)$$

where the adjustable parameters (q_{\max} and b) are held in the vector β and ε is the experimental error.

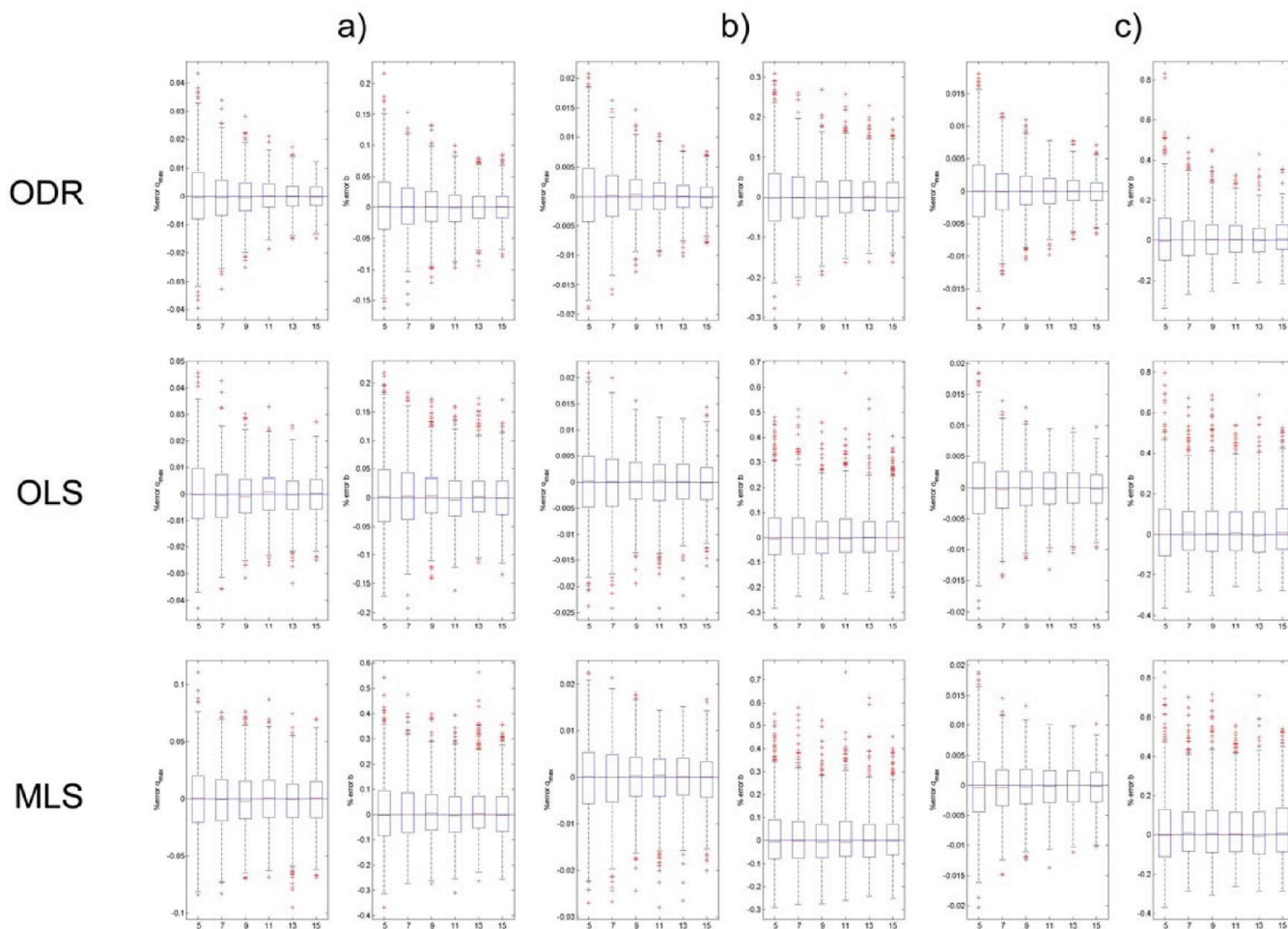


Figure 6. Box plots for the error in estimation of q_{\max} and b as a function of q_{\max} values for different methods of regression. Perturbation: fixed error; number of isotherm data points: 7; b values: (a) 0.5, (b) 4.0, and (c) 12.0.

Table 5. Interquartile Ranges for % Error in Estimation of q_{\max} and b for a Fixed b^a

model	interquartile ranges			
	% error in q_{\max}		% error in b	
	b 4, q_{\max} 5	b 4, q_{\max} 15	b 4, q_{\max} 5	b 4, q_{\max} 15
ODR	0.0091	0.0035	0.1189	0.0730
OLS	0.0097	0.0061	0.1499	0.1195
MLS	0.0110	0.0076	0.1693	0.1341

^aPerturbation: fixed error.

The most used method to determine the isotherm parameters is the least-squares method. The least-squares regression makes various assumptions about errors in a regression model. The basic assumptions are⁹

- The error, ε , is uncorrelated with x , the independence assumption.
- The error has the same variance (σ^2) across the different levels of x , i.e., the variance of ε is homoskedastic and not heteroskedastic.
- The values of ε are independent of each other, i.e., not autocorrelated or serially correlated.
- The error is normally distributed.
- The independent variable x is fixed, i.e., there is not measured error in x .

Table 6. Interquartile Ranges for % Error in Estimation of q_{\max} and b for a Fixed b^a

model	interquartile ranges			
	% error in q_{\max}		% error in b	
	b 4, q_{\max} 5	b 4, q_{\max} 15	b 4, q_{\max} 5	b 4, q_{\max} 15
ODR	0.0596	0.0328	0.2798	0.1716
OLS	0.0629	0.0329	0.3044	0.1841
MLS	0.0591	0.0325	0.2765	0.1673

^aPerturbation: error proportional to concentration.

If these assumptions are met, then the estimates of the regression constant and the regression coefficients are unbiased and efficient. Violation of one of more of these assumptions may lead to biased and/or inefficient estimates.^{9,10}

Assumption (v) cannot be met; x is determined experimentally by chemical analysis which involves a series of operations like pipetting, dilution, standards preparation, calibration of the instruments, etc. All of these operations result in sources of error. Taking into account that Y_i is calculated by using eq 3, the variance of Y_i is dependent on X_{ij} ; therefore, assumption (ii) is not met either.

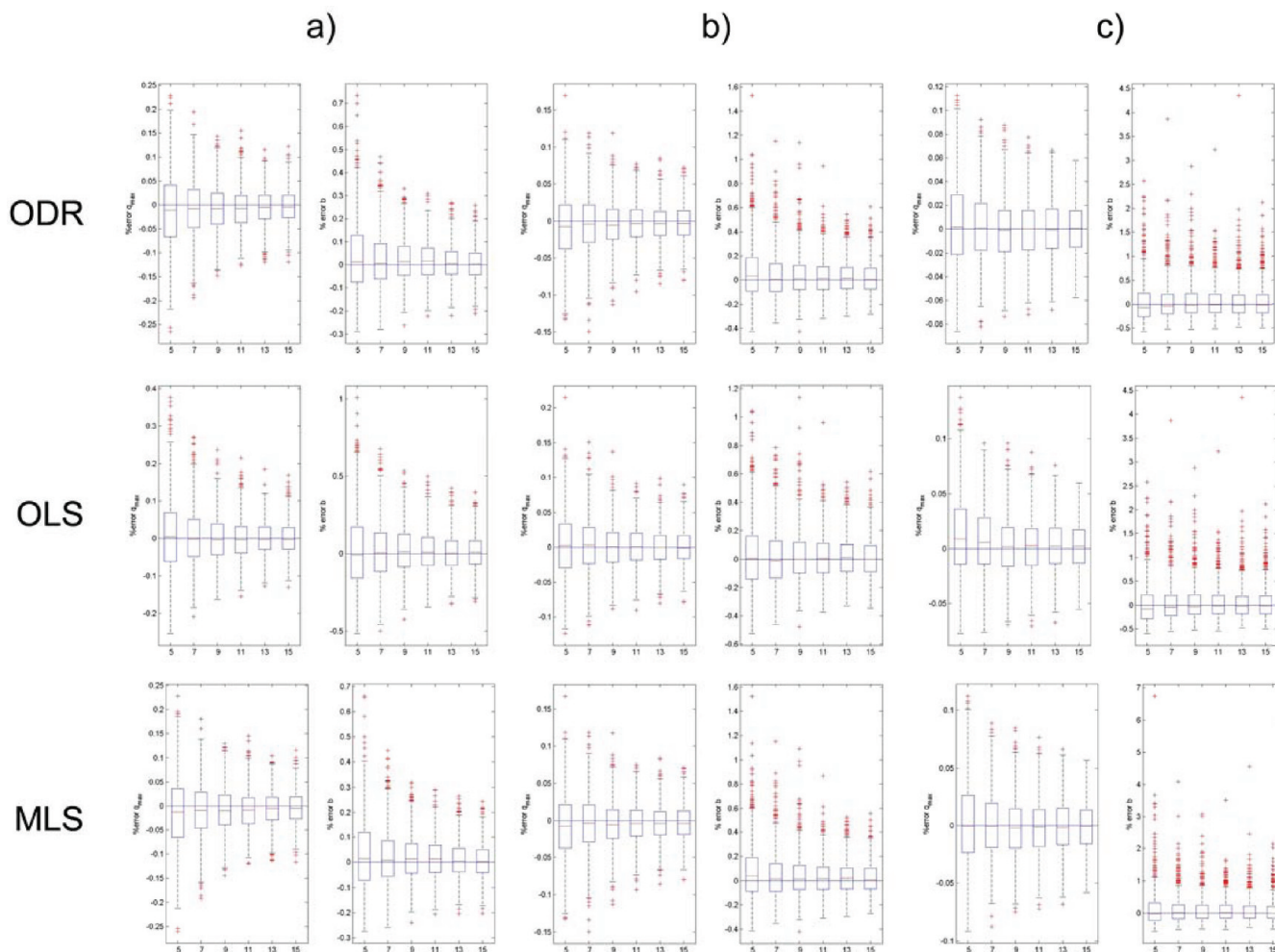


Figure 7. Box plots for the error in estimation of q_{\max} and b as a function of q_{\max} values for different methods of regression. Perturbation: error proportional to concentration; number of isotherm data points: 7; b values: (a) 0.5, (b) 4.0, and (c) 12.0.

When assumption (v) is not met, the problem is

$$\begin{aligned} Y_i &= f(x_i, \beta) + \varepsilon_i \\ X_i &= x_i + \delta_i \end{aligned} \quad (5)$$

where δ_i is the experimental error in x_i determination. Therefore, in order to obtain a better estimation of the parameters errors in both X and Y must be considered.

When assumption (ii) is not met some corrections must be made by giving some weights to correct the differences in the variances.

This fact leads the authors to use different error functions to estimate the isotherm parameters. The most used error functions are described in the following sections.

2.2.1. Least-Squares Linearized (LSL). One procedure usually used for obtaining isotherm parameters is the linearization of the isotherm equation and the application of a method of linear regression. Four of the linearized forms of the Langmuir isotherm are^{11–13}

$$\frac{q_e}{C_e} = bq_{\max} - bq_e \quad (6)$$

where q_e and C_e/q_e are not independent.

$$\frac{1}{q_e} = \left(\frac{1}{bq_{\max}} \right) \frac{1}{C_e} + \frac{1}{q_{\max}} \quad (7)$$

where q_e and q_e/C_e are not independent. In eqs 6 and 7, the abscissa is not free of error therefore they violate assumption (v) in the method of least-squares. The presence of q_e in both dependent and independent variables leads to spurious correlation.

$$\frac{1}{q_e} = \left(\frac{1}{bq_{\max}} \right) \frac{1}{C_e} + \frac{1}{q_{\max}} \quad (8)$$

This linearized form leads to clumping of data points near origin. It is very sensitive to low values of C_e and/or q_e .

$$\frac{C_e}{q_e} = \frac{1}{q_{\max}} C_e + \frac{1}{bq_{\max}} \quad (9)$$

C_e/q_e and C_e are not independent. The presence of C_e in both dependent and independent variables leads to spurious correlation. In all the above-described linearized equations, hypotheses (ii), (iv), and (v) are usually not true.

In this work eq 9 has been chosen because it is one of the most used in literature. Therefore the function error to be minimized is

$$\sum_{i=1}^n \left(\left(\frac{C_{ei}}{q_{ei}} \right)_{\text{exp}} - \left(\frac{C_{ei}}{q_{ei}} \right)_{\text{calc}} \right)^2 \quad (10)$$

The use of the linearization procedure provokes the alteration of the original data, and therefore, the obtained fit with eq 10 could not be the optimum for the original experimental data.^{13,14} At present, the use of this method is not justified as adequate software to fit experimental data to a nonlinear model is easily available.

2.2.2. Ordinary Least-Squares (OLS). When the isotherm equation is not linearized, a widely used error function is the ordinary least-squares (OLS) method. This method consists on minimizing the sum of square errors.

$$\sum_{i=1}^n (q_{ei,\text{calc}} - q_{ei,\text{exp}})^2 \quad (11)$$

In general, the parameters obtained with this error function fit better the data corresponding to the high values of the concentration range.⁶ This error function will provide the best fit when all the assumptions described above are met.

2.2.3. Marquardt's Least-Squares (MLS). This error function minimizes the sum of the relative error between experimental and calculated data. It has been modified to include the number of degrees of freedom of the system.

$$100 \sqrt{\frac{1}{n-p} \sum_{i=1}^n \left(\frac{q_{ei,\text{exp}} - q_{ei,\text{calc}}}{q_{ei,\text{exp}}} \right)^2} \quad (12)$$

where n is the number of data points and p is the number of parameters.¹⁵

2.2.4. Hybrid Least-Squares (HLS). This error function was derived by Porter et al.¹⁵ with the aim to improve the fit of OLS at low concentrations. In this function, each sum of the square of errors is divided by the experimental solid-phase concentration $q_{ei,\text{exp}}$. It also includes a divisor as a term of the function for the number of degrees of freedom of the system.

$$\frac{100}{n-p} \sum_{i=1}^n \frac{(q_{ei,\text{exp}} - q_{ei,\text{calc}})^2}{q_{ei,\text{exp}}} \quad (13)$$

This error function and the previously described (MLS) are two forms of introducing weights to compensate the effects of the nonhomogeneity of the variance (σ^2).

The above-described error functions (LSL, OLS, MLS, and HLS) have been used in sorption studies to evaluate and compare the fit of the experimental data to different isotherms models.^{4-8,15-18}

2.2.5. Orthogonal Distance Regression (ODR). The previously described error functions assume that the values of C_{ei} are exactly known and q_{ei} are observed with error. And it is under this assumption that those error functions provide the optimum solution.

Nevertheless, the real situation is that C_{ei} and C_{oi} are determined experimentally by chemical analysis, therefore should be observed with error. On the other hand q_{ei} is calculated by eq 3

and thus, both C_{ei} and C_{oi} errors contribute to q_{ei} error. Therefore, errors in both C_e and q_e must be taken into account.

The ODR error function takes into account this aspect and provides the best fit by minimization of the sum of square relative errors with regard to both C_e and q_e .

$$\sum_{i=1}^n \left(\frac{C_{ei,\text{exp}} - C_{ei,\text{calc}}}{C_{ei,\text{exp}}} \right)^2 + \left(\frac{q_{ei,\text{exp}} - q_{ei,\text{calc}}}{q_{ei,\text{exp}}} \right)^2 \quad (14)$$

This error function was used by El-Khaiary⁹ when modeling sorption data.

2.3. Numerical Calculations. Linear, nonlinear, and orthogonal distance regression computations were carried out using Matlab R2008b. Specifically the optimization has been performed by applying the Generalized Reduced Gradient method and using for this purpose the library GRG2.¹⁹ Statistics Toolbox of Matlab has been used to generate the perturbations and the statistics analysis.

3. RESULTS

The values obtained from the simulated experimental data with a fixed error ($N(0, 0.05)$) and with $\pm 5\%$ error proportional to concentration were compared to the q_{max} and b theoretical values and the % of error in q_{max} and b was calculated

$$\% \text{ error in } q_{\text{max}} = \frac{q_{\text{max,theo}} - q_{\text{max,calc}}}{q_{\text{max,theo}}} \quad (15)$$

$$\% \text{ error in } b = \frac{b_{\text{theo}} - b_{\text{calc}}}{b_{\text{theo}}} \quad (16)$$

The statistics of the calculated errors by using different error functions (ODR, LSL, OLS, HLS, and MLS) are summarized in Tables 1 and 2. In the same tables, results corresponding to three different numbers of considered equilibrium data (13, 10, and 7) are also presented.

To describe the distribution of the errors, box plots were used. A box plot provides an excellent visual summary of many important aspects of a values distribution. The line across the box represents the median, whereas the bottom and top of the box show the location of the first and third quartiles (Q1 and Q3). The whiskers are the lines that extend from the bottom and top of the box to the lowest and highest observations inside the region defined by $Q1 - 1.5(Q3 - Q1)$ and $Q3 + 1.5(Q3 - Q1)$.²⁰ The box itself contains the middle 50 % of the data. If the median is not equidistant from the top and the bottom of the box, then the data are skewed. In the literature box plots were applied to assess the robustness of a model to describe Cr(VI) kinetics biosorption²¹ and to test the sensitivity of different regression methods to experimental errors.⁹

Box plots for the error in estimation of q_{max} and b obtained by different methods of regression for all simulated data from perturbed data with a fixed error and proportional error are depicted in Figures 2 and 3, respectively. In both cases, different number of isotherm data points has been considered: 7, 10, and 13.

As seen in Figure 2, the ODR method presents the lowest variability. This result was expected because ODR method is the only one who considers there is error in X (it does not meet assumption (v)) and therefore this method provides efficient and not biased estimated parameters. When comparing results obtained by using the other error functions, OLS presents less

variability than the others because in this case it is assumed that variance is constant with independence of the different X values.

As expected, the variability decreases with the increase of the number of data points independently of the used error function.

When looking at Table 1, it can be observed that in all the cases the ODR method present the lowest mean, standard deviation and interquartile range. Also, the standard deviation (std) and the interquartile range (iqr) increase when the number of points decreases.

The distribution of relative error in the determination of q_{\max} and b is centered as observed in Figure 2. This fact can be confirmed by the mean ($<5 \times 10^{-4}$) and median ($<10^{-2}$) values shown in Table 1. A major variability in b than in q_{\max} determination can be observed in Figure 2; indeed in b determination the standard deviation and interquartile range values are higher (Table 1).

The LSL method presents similar variability to OLS when looking at Figure 2; nevertheless, LSL method shows a greater standard deviation than the other methods in b determination as shown in Table 1.

Figure 3 shows the box plot corresponding to the error in estimation of q_{\max} and b obtained by different methods of regression for all simulated data from perturbed data with a proportional error. As seen, ODR and MLS present the lowest variability but it is not clear which is the most efficient method. As said before, ODR method considers there is error in X axis nevertheless, it does not take into account that in the case of perturbed data with a proportional error assumption (ii) is not met. Therefore, in this case, it cannot be guaranteed that ODR method is the optimum one. MLS cannot also be considered the optimum one because even though it takes into account that the data are heteroskedastic it does not consider the errors in X axis. The variability of these two methods is closely followed by HLS, then by OLS and more distantly by LSL.

The trend of variability as a function of number of data points is the same that the observed in Figure 2.

When comparing results shown in Figures 2 and 3, it must be pointed out that the error in estimation of q_{\max} and b is always greater when the data were perturbed with a proportional error.

In Table 2, it can be observed that MLS and ODR methods present similar standard deviation and interquartile range in q_{\max} determination. In b determination, the lowest standard deviation is given by ODR and the narrowest interquartile range by MLS.

Taking into account the results obtained with these two methods, it can be stated that none of them is outstanding over the other. Furthermore, variability values provided by HLS and OLS are not so different.

In order to analyze the effect of the theoretical q_{\max} and b values chosen to achieve the simulated data and the estimation of q_{\max} and b obtained by the different methods, as an example, box-plots for the error in estimation of q_{\max} and b as a function of b values from perturbed data with a fixed error and proportional error are shown in Figures 4 and 5, respectively. In these examples, q_{\max} takes three fixed values (5, 9, and 15) and the number of data points is 7.

As can be seen in Figure 4 in general, with independence of the theoretical q_{\max} and the method used, as higher is b the lower variability is in q_{\max} determination. Conversely, the variability in b determination increases when increasing the theoretical b value.

Table 3 shows the interquartile-ranges corresponding to q_{\max} and b determination for a $q_{\max} = 9$ and two different values of b (0.5 and 12), and a fixed error when using ODR, OLS, and MLS methods. For q_{\max} determination, the interquartile ranges of ODR are narrower compared to OLS and MLS. From the results presented in the table it can be pointed out that the difference between the interquartile ranges values presented by ODR and the other methods is bigger when the theoretical values of b are low. In the case of b determination this difference is not so remarkable.

In Figure 5, the trend in q_{\max} determination is the same as observed when a fixed error was applied (see Figure 4). That is, as higher is b the lower variability is found for q_{\max} . Nevertheless, it must be pointed out that in that case (proportional error) q_{\max} values are skewed for the lowest b value (0.5) and become centered as b increases. When comparing the three methods, OLS presents the highest variability in both q_{\max} and b determination.

In the case of b determination, in general, it appears that a great percentage of the values are outliers with a great variation range (Figure 5a–c). Comparing the three methods, OLS presents the highest variability for the lowest b value (0.5) nevertheless the three methods tend to show the same variability when increasing the theoretical b value. This observation is corroborated by the interquartile-ranges values presented in Table 4.

In Figures 6 and 7, box-plots for the error in estimation of q_{\max} and b as a function of q_{\max} values from perturbed data with a fixed error and proportional error are shown. As observed in both figures, the trend of the variability in both q_{\max} and b determination is to decrease when increasing the q_{\max} value.

Figure 6 shows that in general, the errors distribution is centered with independence of the q_{\max} and b values and the method used. The corresponding interquartile ranges are presented in Table 5 where it can be observed that the narrowest interquartile-range values are provided by ODR method. It can also be observed that difference of the interquartile-ranges between ODR and the other methods is more evident when q_{\max} is 15.

In Figure 7, it can be observed that for low q_{\max} the values of errors distribution for the estimation of q_{\max} is skewed toward low values (Figure 7, panels a and b). This is not the case of Figure 7 c where the errors distribution is quite centered for ODR and MLS and skewed to high values for OLS. For b estimation, in most of the cases the distribution is quite centered but a general trend is not observed. In the case of b 12, outliers with great positive values are observed for all the methods. The corresponding interquartile ranges are presented in Table 6. As observed in the table, when q_{\max} is high the interquartile range of OLS are similar to the ones of ODR and MLS.

4. CONCLUSIONS

The results of this study show that orthogonal distance regression (ODR) gives the most accurate estimates of the isotherm parameters among the different methods when the experimental data have a fixed error. When the experimental data have a proportional error ODR gives also accurate estimates but they are similar to the obtained with MLS method.

As expected, an increase of number of data enhances accuracy in the estimation of isotherm parameters with independence of the regression method used.

ODR method provides the most accurate estimates of q_{\max} and b as higher is q_{\max} . Conversely, accuracy in both parameters

estimation is not favored with the increase of b value. High b values provide accuracy in q_{\max} estimates while low b values in b estimates.

As a final conclusion, it is advisable to make an appropriate scaling of the data in order to obtain the highest accuracy when using the ODR method.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jordi.poch@udg.edu.

Funding

This work has been supported by Ministerio de Ciencia e Innovación, Spain, Project CTM2008-06776-C02-01.

Notes

§E-mail: isabel.villaescusa@udg.edu.

REFERENCES

- (1) Nurchi, V. M.; Villaescusa, I. Biomasses as sorbents of some trace metals. *Coord. Chem. Rev.* **2008**, *25*, 1178–1188.
- (2) Mubeena, A.; Hasany, S. M.; Bhanger, M. I.; Iqbal, S. Sorption potential of Moringa oleifera pods for the removal of organic pollutants from aqueous solutions. *J. Hazard. Mater.* **2007**, *141*, 546–556.
- (3) Limousin, G.; Gaudet, J. P.; Charlet, L.; Szenknect, S.; Barthès, V.; Krimissa, M. Sorption Isotherms: A review on physical bases, modelling and measurement. *Appl. Geochem.* **2007**, *22*, 249–275.
- (4) Allen, S. J.; Gan, Q.; Matthews, R.; Johnson, A. Comparison of optimized isotherm models for basic dye adsorption by kudzu. *Bioresour. Technol.* **2003**, *88*, 143–152.
- (5) Demirbas, E.; Kobya, M.; Konukman, A. E. S. Error analysis of equilibrium studies for the almond shell activated carbon adsorption of Cr(VI) from aqueous solutions. *J. Hazard. Mater.* **2008**, *154*, 787–794.
- (6) Ng, J. C. Y.; Cheung, W. H.; W.H. McKay, G. Equilibrium studies for the sorption of lead from effluents using chitosan. *Chemosphere* **2003**, *52*, 1021–1030.
- (7) Rivas, F. J.; Beltrán, F. J.; Gimeno, O.; Frades, J.; Carvalho, F. Adsorption of landfill leachates onto activated carbon. Equilibrium and kinetics. *J. Hazard. Mater.* **2006**, *B131*, 170–178.
- (8) Wong, Y. C.; Szeto, Y. S.; Cheung, W. H.; McKay, G. Adsorption of acid dyes on chitosan – equilibrium isotherm analysis. *Process Biochem.* **2004**, *39*, 693–702.
- (9) El-Khaiary, M. I. Least-squares regression of adsorption equilibrium data: comparing the options. *J. Hazard. Mater.* **2008**, *158*, 73–87.
- (10) Rawlings, J. O.; Pantula, S. G.; Dickey, D. A. *Applied Regression Analysis. A research tool*; Springer-Verlag: New York, 2001; p 657.
- (11) Bolster, C. H.; Hornberger, G. M. On the use of linearized Langmuir Equations. *Nutrient Manag. Soils Plant Anal.* **2007**, *71*, 1796–1806.
- (12) Kumar, K. V.; Sivanesan, S. Prediction of optimum sorption isotherm: Comparison of linear and non-linear method. *J. Hazard. Mater.* **2005**, *B126*, 198–201.
- (13) Barrow, N. J. The description of sorption curves. *Eur. J. Soil Sci.* **2008**, *59*, 900–910.
- (14) Badertscher, M.; Pretsch, E. Bad results from good data. *Trends Anal. Chem.* **2006**, *25*, 1131–1138.
- (15) Porter, J. F.; McKay, G.; Choy, K. H. The prediction of sorption from a binary mixture of acid dyes using simple single-and mixed-isotherm variance of the ideal adsorption solute theory. *Chem. Eng. Sci.* **1999**, *54*, 5863–5885.
- (16) Allen, S. J.; McKay, G.; Porter, J. F. Adsorption isotherm models for basic dye adsorption by peat in single and binary component systems. *J. Colloid Interface Sci.* **2004**, *280*, 322–333.
- (17) Gunay, A. Application of nonlinear regression analysis for ammonium exchange by natural (Bigadiç) clinoptilolite. *J. Hazard. Mater.* **2007**, *148*, 708–713.
- (18) Ncibi, M. C. Applicability of some statistical tools to predict optimum adsorption isotherm after linear and non-linear regression analysis. *J. Hazard. Mater.* **2008**, *153*, 207–212.
- (19) Lasdon, L. S.; Warren, A. D. Generalized reduced gradient software for linearly and nonlinearly constrained problems. In *Design and Implementation Optimization Software*; Alphen aa den Rijn: The Netherlands, 1979; pp 363–396.
- (20) Vega, M.; Pardo, R.; Barrado, E.; Debán, L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* **1998**, *32*, 3581–3592.
- (21) Poch, J.; Villaescusa, I. A model to describe Cr(VI) kinetics biosorption. *J. Hazard. Mater.* **2010**, *175*, 770–778.