

## Research Article

# Robust Nonlinear Regression in Enzyme Kinetic Parameters Estimation

Maja Marasović,<sup>1</sup> Tea Marasović,<sup>2</sup> and Mladen Miloš<sup>1</sup>

<sup>1</sup>Faculty of Chemistry and Technology, University of Split, Ruđera Boškovića 35, 21000 Split, Croatia

<sup>2</sup>Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Ruđera Boškovića 32, 21000 Split, Croatia

Correspondence should be addressed to Tea Marasović; [tmarasov@fesb.hr](mailto:tmarasov@fesb.hr)

Received 18 October 2016; Accepted 6 February 2017; Published 5 March 2017

Academic Editor: Murat Senturk

Copyright © 2017 Maja Marasović et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate estimation of essential enzyme kinetic parameters, such as  $K_m$  and  $V_{max}$ , is very important in modern biology. To this date, linearization of kinetic equations is still widely established practice for determining these parameters in chemical and enzyme catalysis. Although simplicity of linear optimization is alluring, these methods have certain pitfalls due to which they more often than not result in misleading estimation of enzyme parameters. In order to obtain more accurate predictions of parameter values, the use of nonlinear least-squares fitting techniques is recommended. However, when there are outliers present in the data, these techniques become unreliable. This paper proposes the use of a robust nonlinear regression estimator based on modified Tukey's biweight function that can provide more resilient results in the presence of outliers and/or influential observations. Real and synthetic kinetic data have been used to test our approach. Monte Carlo simulations are performed to illustrate the efficacy and the robustness of the biweight estimator in comparison with the standard linearization methods and the ordinary least-squares nonlinear regression. We then apply this method to experimental data for the tyrosinase enzyme (EC 1.14.18.1) extracted from *Solanum tuberosum*, *Agaricus bisporus*, and *Pleurotus ostreatus*. The results on both artificial and experimental data clearly show that the proposed robust estimator can be successfully employed to determine accurate values of  $K_m$  and  $V_{max}$ .

## 1. Introduction

Enzymes are molecules that act as biological catalysts and are responsible for maintaining virtually all life processes. Most enzymes are proteins, although a few are catalytic RNA molecules. Like all catalysts, enzymes increase the rate of chemical reactions without themselves undergoing any permanent chemical change in a process. They achieve their effect by temporarily binding to the substrate and, in doing so, lowering the activation energy needed to convert it to a product. The study of the rate at which an enzyme works is called enzyme kinetics and it is often regarded as one of the most fascinating research areas in biochemistry [1].

Mathematically, the relationship between substrate concentration and reaction rate under isothermal conditions for many of enzyme-catalyzed reactions can be modeled by the Michaelis-Menten equation [2]:

$$v = \frac{V_{max}s}{s + K_m}, \quad (1)$$

where  $v$  denotes a reaction rate,  $s$  is a substrate concentration,  $V_{max}$  is the maximum initial velocity, which is theoretically attained when the enzyme has been “saturated” by an infinite concentration of a substrate, and  $K_m$  is the Michaelis constant, representing a measure of affinity of the enzyme-substrate interaction. By definition,  $K_m$  is equal to the concentration of the substrate at half maximum initial velocity. The Michaelis constant,  $K_m$ , is an intrinsic parameter of enzyme-catalyzed reactions and it is significant for its biological function [3].

Three most common methods, available in the literature, for determining the parameters of Michaelis-Menten equation based on a series of measurements of velocity  $v$  as a function of substrate concentration, are Lineweaver-Burk plot, also known as the double reciprocal plot, Eadie-Hofstee plot, and Hanes-Woolf plot. All three of these methods are linearized models that transform the original Michaelis-Menten equation into a form which can be graphed as a straight line.

Lineweaver-Burk [4] (LB) plot, still the most popular and favored plot amongst the researchers, is defined by an equation:

$$\frac{1}{v} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \frac{1}{s}. \quad (2)$$

The  $y$ -intercept in this plot is  $1/V_{\max}$ , the  $x$ -intercept in second quadrant represents  $-1/K_m$ , and the slope of the line is  $K_m/V_{\max}$ .

Eadie-Hofstee [5] (EH) plot is a semireciprocal plot of  $v$  versus  $v/s$ . The linear equation has the following form:

$$v = V_{\max} - K_m \frac{v}{s}, \quad (3)$$

where the  $y$ -intercept is  $V_{\max}$  and the slope is  $K_m$ .

In Hanes-Woolf [6] (HW) plot,  $s/v$  is plotted against  $s$ . The linear equation is given by

$$\frac{s}{v} = \frac{K_m}{V_{\max}} + \frac{1}{V_{\max}} s, \quad (4)$$

where the  $y$ -intercept is  $K_m/V_{\max}$  and the slope is  $1/V_{\max}$ .

In all of the above-described linear transformations, linear regression is used to estimate the slope and intercept of the straight line and afterwards  $K_m$  and  $V_{\max}$  are computed from the straight line parameters. Although these methods are very useful for data visualization and are still widely employed in enzyme kinetic studies, each of them possesses certain deficiencies, which make them prone to errors. For instance, Lineweaver-Burk plot has the disadvantage of compressing the data points at high substrate concentrations into a small region and emphasizing the points at lower substrate concentrations, which are often the least accurate [7]. The  $V$ -intercept in Lineweaver-Burk plot is equivalent to inverse of  $V_{\max}$  due to which any small error in measurement gets magnified. Similarly, the Eadie-Hofstee plot has the disadvantage that  $v$  appears on both axes; thus, any experimental error will also be present in both axes. In addition, experimental errors or uncertainties are propagated unevenly and become larger over the abscissa thereby giving more weight to smaller values of  $v/s$ . Hanes-Woolf plot is the most accurate of the three; however, its major drawback is that again neither ordinate nor abscissa represents independent values: both are dependent on substrate concentration.

In order to reduce the errors due to the linearization of parameters, Wilkinson [8] proposed the use of least-squares nonlinear regression for more accurate estimation of enzyme kinetic parameters. Nonlinear regression allows direct determination of parameter values from untransformed data points. The process starts with initial estimates and then iteratively converges on parameter estimates that provide the best fit of the underlying model to the actual data points [9, 10]. The algorithms used include the Levenberg-Marquardt method, the Gauss-Newton method, the steepest-descent method, and simplex minimization. Numerous software packages, such as Excel, MATLAB, and GraphPrism, nowadays include readily available routines and scripts to perform nonlinear least-squares fitting [11, 12].

Least-squares nonlinear regression has been criticized for its performance in dealing with experimental data. This is mainly due to the fact that implicit assumptions related with nonlinear regression are in general not met in the context of deviations that appear as a result of biological errors (e.g., variations in the enzyme preparations due to oxidation or contaminations) and/or experimental errors (e.g., variations in measured volume of substrates and enzymes, imprecisions of the instrumentation). With the presence of outliers or influential observations in the data, the ordinary least-squares method can result in misleading values for the parameters of the nonlinear regression and estimates may no longer be reliable [13].

In this paper, we propose the use of robust nonlinear regression estimator based on modified Tukey's biweight function for determining the parameters of Michaelis-Menten equation using experimental measurements in enzyme kinetics. The main idea is to fit a model to the data that gives resilient results in the presence of influential observations and/or outliers. To the best of our knowledge, this is the first study that examines the use of this technique for application in Michaelis-Menten enzyme analysis. We employ Monte Carlo simulations to validate the efficacy of the proposed procedure in comparison with the ordinary least-squares method and Eadie-Hofstee, Hanes-Woolf and Lineweaver-Burk plots. In addition, we illustrate the viability of our method by estimating the kinetic parameters of tyrosinase, an important enzyme widely distributed in microorganisms, animals, and plants, responsible for melanin production in mammal and enzymatic browning in plants, extracted from potato and two edible mushrooms.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the robust estimation model. Section 3 describes the experimental setup used in this research and the diagnostics that will be used to evaluate the effectiveness of the proposed procedure in determination of enzyme kinetic parameters. Results are discussed in Section 4. Finally, Section 5 summarizes the paper with a few concluding remarks.

## 2. Robust Nonlinear Regression

Nonlinear regression, same as linear regression, relies heavily on the assumption that the scatter of data around the ideal curve follows, at least approximately, a Gaussian or normal distribution. This assumption leads to the well-known regression goal: to minimize the sum of the squares of the vertical distances (a.k.a residuals) between the points and the curve. In practice, however, this assumption does not always hold true. The analytical data often contains outliers that can play havoc with standard regression methods based on the normality assumption, causing them to produce more or less strongly biased results, depending on the magnitude of deviation and/or sensitivity of the procedure. It is not unusual to find an average of 10% of outlying observations in data set of some processes [14].

Outliers are most commonly thought to be extreme values which are a result of measurement or experimental errors. Barnett and Lewis [15] provide a more cautious definition of

the term outlier, describing it as the observation (or subset of observations) that appears to be inconsistent with the remainder of the dataset. This definition also includes the observations that do not follow the majority of the data, such as values that have been measured correctly but are, for one reason or another, far away from other data values, while the formulation “appears to be inconsistent” reflecting the subjective judgement of the observer whether or not an observation is declared to be outlying.

The ordinary least-squares (OLS) estimate  $\hat{\beta}_{LS}$  of the parameter vector  $\beta = [\beta_1, \beta_2, \dots, \beta_n]$  is obtained as the solution of the problem:

$$\begin{aligned}\hat{\beta}_{LS} &= \arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n r_i(x, \beta)^2 \\ &= \arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i, \beta))^2,\end{aligned}\quad (5)$$

where  $n$  denotes the number of observations,  $x = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$  is a  $n \times p$  matrix, whose rows are  $p$ -dimensional vectors of predictor variables (or regressors),  $y = [y_1, y_2, \dots, y_n]^T$  is a  $n \times 1$  vector of responses, and  $f(x_i, \beta)$  is model function. Since all data points are attributed the same weights, OLS implicitly favors the observations with very large residuals and, consequently, the estimated parameters end up distorted if outliers are present.

In order to achieve robustness in coping with the problem of outliers, Huber [16] introduced a class of so-called  $M$ -estimators, for which the sum of function  $\rho$  of the residuals is minimized. The resulting vector of parameters  $\hat{\beta}_M$  estimated by an  $M$ -estimator is then

$$\hat{\beta}_M = \arg \min_{\beta} \sum_{i=1}^n \rho\left(\frac{r_i}{\sigma}\right). \quad (6)$$

The residuals are standardized by a measure of dispersion  $\sigma$  to guarantee scale equivariance (i.e., independence with respect to the measurement units of the dependent variable). Function  $\rho(\cdot)$  must be even, nondecreasing for positive values, and less increasing than the square.

The minimization in (6) can always be done directly. However, often it is simpler to differentiate  $\rho$  function with respect to  $\beta$  and solve for the root of the derivative. When this differentiation is possible, the  $M$ -estimator is said to be of  $\psi$ -type. Otherwise, the  $M$ -estimator is said to be of  $\rho$ -type.

Let  $\psi = \rho'$  be the derivative of  $\rho$ . Assuming  $\sigma$  is known and defining weights  $w_i = \psi(r_i/\sigma)/r_i$ , the estimates  $\hat{\beta}_M$  can be obtained by solving the system of equations:

$$\sum_{i=1}^n w_i^2 r_i^2 = 0. \quad (7)$$

The weights are dependent upon the residuals, the residuals are dependent upon the estimated coefficients, and the estimated coefficients are dependent upon the weights. Hence, to solve for  $M$ -estimators, an iteratively reweighted least-squares (IRLS) algorithm is employed. Starting from some

initial estimates  $\hat{\beta}^{(0)}$ , at each iteration  $t$  until it converges, this algorithm computes the residuals  $r_i^{(t-1)}$  and the associated weights  $w_i^{(t-1)} = w[r_i^{(t-1)}]$  from the previous iteration and yields new weighted least-squares estimates.

**2.1. Objective Function.** Several  $\rho$  functions can be used. Here we opted for Tukey's biweight [17] or bisquare function defined as

$$\rho(z_i) = \begin{cases} \frac{c^2}{6} \left( 1 - \left[ 1 - \left( \frac{z_i}{c} \right)^2 \right]^3 \right) & \text{if } |z_i| \leq c \\ \frac{c^2}{6} & \text{if } |z_i| > c \end{cases}, \quad (8)$$

where  $c$  is a tuning constant and  $z_i = r_i/\sigma$ .

The corresponding  $\psi(z_i)$  function is

$$\psi(z_i) = \begin{cases} z_i \left[ 1 - \left( \frac{z_i}{c} \right)^2 \right]^2 & \text{if } |z_i| \leq c \\ 0 & \text{if } |z_i| > c \end{cases}. \quad (9)$$

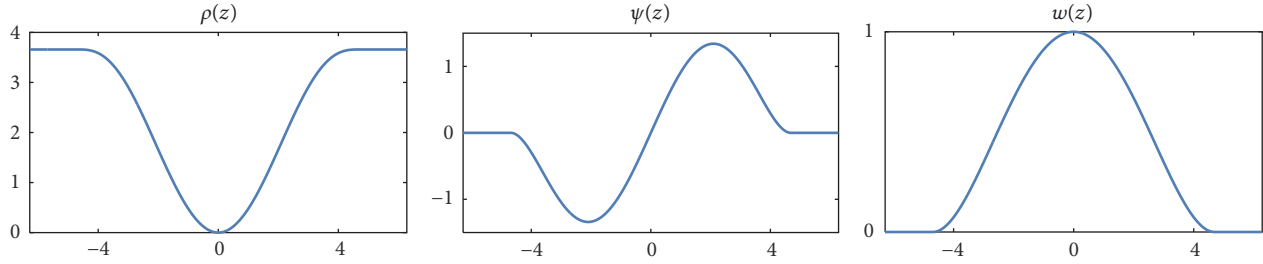
Tukey's biweight estimator has a smoothly redescending  $\psi$  function that prevents extreme outliers to affect the calculation of the biweight estimates by assigning them a zero weighting. As can be seen in Figure 1, the weights for the biweight estimator decline as soon as  $z$  departs from 0 and are 0 for  $|z| > c$ . Smaller values of  $c$  produce more resistance to outliers, but at the expense of lower efficiency when the errors are normally distributed. The tuning constant is generally picked to give reasonably high efficiency in normal case; in particular  $c = 4.685$  produces a 95% efficiency when the errors are normal, while guaranteeing resistance to contamination of up to 10% of outliers.

In an application, an estimate of the standard deviation of the errors is needed in order to use these results. Usually a robust measure of spread is used in preference to the standard deviation of the residuals. A common approach is to take  $\hat{\sigma} = \text{MAD}/0.6745$ , where MAD is the median absolute deviation. Despite having the best possible breakout point of 50%, the MAD is not without its weaknesses. It exhibits superior statistical efficacy for the contaminated data (i.e., the data that contains extreme scores); however, when the data approaches a normal distribution, the MAD is only 37% efficient. Furthermore, it is ill-suited for asymmetrical distributions, since it attaches equal importance to positive and negative deviations from location estimate.

Hence, the scale parameter  $\sigma$  is computed using Rousseeuw-Croux estimator  $Q_n$  [18]:

$$Q_n = d \left\{ |x_i - x_j|; i < j \right\}_{(k)}, \quad (10)$$

where  $d$  is a calibration factor and  $k = \left( \frac{h}{2} \right) \approx \left( \frac{n}{2} \right) / 4$ , where  $h = n/2 + 1$  is roughly half the number of observations. The estimator  $Q_n$  has the optimal 50% breakdown point; it is equally suitable for both symmetrical and asymmetrical distributions and considerably more efficient (about 82%) than the MAD under a Gaussian distribution.

FIGURE 1: Tukey's biweight estimator objective,  $\psi$ , and weight functions for  $c = 4.685$ .

### 3. Experimental Setup

To illustrate the efficacy of the proposed approach we use both artificial data, generated from the Monte Carlo simulations, and the experimental data for the tyrosinase enzyme (EC 1.14.18.1) extracted from three different sources.

**3.1. Monte Carlo Simulations.** Simulation studies are useful for gaining insight into the examined algorithms strengths and weaknesses, such as robustness, against number of variable factors. There are three outlier scenarios and a total of 18 different situations considered in this research. The data sets are generated from the model:

$$y_i = \frac{(\theta_1 x_i)}{(\theta_2 + x_i)} + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (11)$$

where regression coefficients are fixed  $\theta = [80 \ 240]$  for each  $i = 1, 2, \dots, n$ . The explanatory variables  $x_i$  are set to  $100 \cdot i$  and a zero mean unit variance random number with Gaussian density  $\epsilon_i$  is added as measurement error.

The factors considered in this simulation are (1) level of outlier contamination: 10% or 20%, (2) sample size: small ( $n = 10$ ), medium ( $n = 20$ ), or large ( $n = 50$ ), and (3) distances of outliers from clean observations: 10 standard deviations, 50 standard deviations, or 100 standard deviations. There are 1200 replications for each scenario and all simulations are carried out in MATLAB. The 3 scenarios and the 18 situations considered in this research are summarized in Table 1.

These simulated data are then used to estimate the values of  $K_m$  and  $V_{\max}$  using different fitting techniques. The mean estimated values of  $K_m$  and  $V_{\max}$  for a particular scenario and fitting technique are subsequently calculated by averaging  $K_m$  and  $V_{\max}$  values obtained in each of the 1200 trials. The estimator efficacy is assessed in terms of its bias, precision, and accuracy. Bias is defined as an absolute difference between mean estimated parameter values and known parameter values:

$$\text{Bias} = \left| \frac{\sum_{j=1}^N \theta_{ij}}{N} - \hat{\theta}_i \right|, \quad (12)$$

where  $N$  is the total number of replications in the simulated scenario. The term precision refers to the absence of random errors or variability. It is measured by the coefficient of

TABLE 1: The 18 situations considered in the simulations.

Scenario	Situation	Sample size	Outliers	$\sigma$
1	1	10	10%	10
	2		20%	10
	3		10%	50
	4		20%	50
	5		10%	100
	6		20%	100
2	7	20	10%	10
	8		20%	10
	9		10%	50
	10		20%	50
	11		10%	100
	12		20%	100
3	13	50	10%	10
	14		20%	10
	15		10%	50
	16		20%	50
	17		10%	100
	18		20%	100

variation ( $C_v$ ), that is, the standard deviation expressed as a percentage of the mean:

$$C_v = \frac{\sigma}{\theta_{ij}} \cdot 100. \quad (13)$$

The prediction accuracy is defined as the overall distance between estimated values and true values. The accuracy is measured by a normalized mean squared error (NMSE), that is, the mean of the squared differences between the estimated and the known parameter values normalized by a mean of estimated data:

$$\text{NMSE} = \frac{1}{N} \sum_{j=1}^N \left( \frac{\theta_{ij} - \hat{\theta}_i}{\theta_{ij}} \right)^2, \quad (14)$$

where again  $N$  is the total number of replications in the simulated scenario.

**3.2. Enzyme Data Sets.** Tyrosinase (EC 1.14.18.1) is a ubiquitous enzyme responsible for melanization in animals and plants [19, 20]. In the presence of molecular oxygen,



TABLE 2: Input experimental kinetic data sets.

Concentration [ $\mu\text{mol}$ ]	Reaction rate [ $\mu\text{mol}/\text{min}$ ]		
	Ab	Po	Potato
25	0.0243	$1.1 \times 10^{-4}$	0.0004
50	0.0292	$2.6 \times 10^{-4}$	0.0011
100	0.0546	$2.9 \times 10^{-4}$	0.0018
250	0.1388	$3.7 \times 10^{-4}$	0.0023
500	0.1726	$6.9 \times 10^{-4}$	0.0030
1000	0.2374	$13.1 \times 10^{-4}$	0.0101
2500	0.3023	$26.9 \times 10^{-4}$	0.0124
5000	0.3395	$49.6 \times 10^{-4}$	0.0375
7500	0.3515	$50.3 \times 10^{-4}$	0.0485
10000	0.3652	$60.7 \times 10^{-4}$	0.0569

this enzyme catalyzes hydroxylation of monophenols to *o*-diphenols (cresolase activity) and their subsequent oxidation to *o*-quinones (catecholase activity). The latter products are unstable in aqueous solution, further polymerizing to undesirable brown, red, and black pigments. Tyrosinase has attracted a lot of attention with respect to its biotechnological applications [21], due to its attractive catalytic ability, as the catechol products are useful as drugs or drug synthons.

For the purposes of present study, tyrosinase was extracted from potato (*Solanum tuberosum*) and two species of common edible mushrooms: *Agaricus bisporus* (Ab) and *Pleurotus ostreatus* (Po). All the source materials were purchased from the local green market in Split, Croatia. Enzyme extraction was prepared with 100 mL of cold 50 mM phosphate buffer (pH 6.0) per 50 g of a source material. The homogenates were centrifuged at 5000 rpm for 30 min and supernatant was collected. The sediments were mixed with cold phosphate buffer and were allowed to sit in cold condition with occasional shaking. Then the sediments containing buffer were centrifuged once again to collect supernatant. These supernatants were subsequently used as sources of enzyme.

The tyrosinase activity was determined spectrophotometrically at room temperature ( $t = 25^\circ\text{C}$ ) and  $\lambda = 475\text{ nm}$ , measuring the conversion of L-DOPA to red coloured oxidation product dopachrome [22]. The reaction mixture—obtained after adding a  $50\text{ }\mu\text{L}$  of enzyme extract to a cuvette containing  $1.2\text{ mL}$  of  $50\text{ mM}$  phosphate buffer (pH 6.0) and  $0.8\text{ mL}$  of  $10\text{ mM}$  L-DOPA—was immediately shaken and the increase in absorbance was measured for 3 minutes. The change in the absorbance was proportional to the enzyme concentration. The initial rate was calculated from the linear part of the recorded progress curve. One unit of enzyme was defined as the amount which catalyzed the transformation of  $1\text{ }\mu\text{mol}$  of L-DOPA to dopachrome per minute under the above conditions. The dopachrome extinction coefficient at  $475\text{ nm}$  was  $3600\text{ M}^{-1}\text{ cm}^{-1}$ .

To determine the values of  $K_m$  and  $V_{\max}$  for tyrosinase, experimental kinetic data, summarized in Table 2, was gathered by measuring enzyme activity in a cuvette where  $50\text{ }\mu\text{L}$  of enzyme solution was added to  $2\text{ mL}$  of  $50\text{ mM}$  phosphate

buffer (pH 6.0) containing various concentrations of L-DOPA ( $0\text{--}10\text{ mM}$ ). In this case, the estimator performance is evaluated by computing mean absolute error (MAE), that is, the mean of the absolute differences between the observed reaction rate,  $v_i$  and the expected reaction rate, calculated using estimates of  $K_m$  and  $V_{\max}$ , at a concentration,  $s_i$ :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left( v_i - \frac{V_{\max} \cdot s_i}{s_i + K_m} \right), \quad (15)$$

where  $n$  is the number of experimental data points. Mean absolute error is regularly employed quality measure that provides an objective assessment of how well the various estimated values of  $K_m$  and  $V_{\max}$  fit the untransformed experimental data.

## 4. Results and Discussion

**4.1. Parameter Estimation Using Simulated Data.** Figures 2–5 provide the summary of our simulation outcomes for different sample sizes, different contamination levels, and different outlier distances. By examining the simulation results, it is evident that modified robust Tukey's biweight estimator outperforms all other four alternative fitting techniques with respect to bias, coefficient of variation, and normalized mean square error, yielding both accurate and precise estimates of  $K_m$  and  $V_{\max}$  at all test conditions. For example, looking at the set of values obtained for a small sample size ( $n = 10$ ) with a minimal level of contamination present in the data and minimal outlier scatter (Situation 1, Table 1), we observe that as per the RNR estimator the average estimated values of  $K_m$  and  $V_{\max}$  are  $240.08 \pm 14.39$  and  $80 \pm 1.5$ . When EH, HW, and LB plots were used, the data produced  $224.38 \pm 38.01$ ,  $241.02 \pm 44.08$ , and  $237.47 \pm 52.05$ , respectively, as average estimates of  $K_m$  and  $78.18 \pm 4.46$ ,  $79.8 \pm 4.85$  and  $79.42 \pm 6.15$ , respectively, as  $V_{\max}$ . When OLS estimator was applied, the corresponding average values of  $K_m$  and  $V_{\max}$  were  $238.71 \pm 40.04$  and  $79.86 \pm 4.39$ , respectively. If the reported standard deviations are scaled by dividing them with an appropriate mean, the resulting coefficients of variation of  $K_m$  and  $V_{\max}$  are 16.9 and 5.7% (EH), 18.3% and 6.1% (HW), 21.9% and 7.7% (LB), 16.8 and 5.5% (OLS), and 6% and 1.9% (RNR), respectively. Thus, it is revealed that, though all three Hanes-Woolf, Lineweaver-Burke, and ordinary least-squares methods have a low bias (Figures 2 and 4) and produce the results that are in a close proximity of the values obtained by robust nonlinear regression method, their estimates are much more imprecise and as such are of a limited utility. Figure 6 shows the plots of fitted reaction curves for the randomly selected replications of situations with small sample size (Situations 1–6).

For a medium sample size ( $n = 20$ ) with the same levels of contamination and outlier scatter (Situation 7, Table 1), the analysis for the RNR approach yielded almost identical results, that is, average  $K_m$  and  $V_{\max}$  estimates as  $239.68 \pm 9.3$  and  $80 \pm 0.68$ , respectively. The EH, HW, and LB plots estimated the average  $K_m$  as  $228.83 \pm 26.39$ ,  $238.01 \pm 38.12$ , and  $238.52 \pm 49.21$ , respectively, and  $V_{\max}$  as  $79.09 \pm 2.12$ ,  $79.6 \pm 2.73$ , and  $79.6 \pm 4.07$ , respectively. The average kinetic parameters, that is,  $K_m$  and  $V_{\max}$ , obtained by the OLS method were

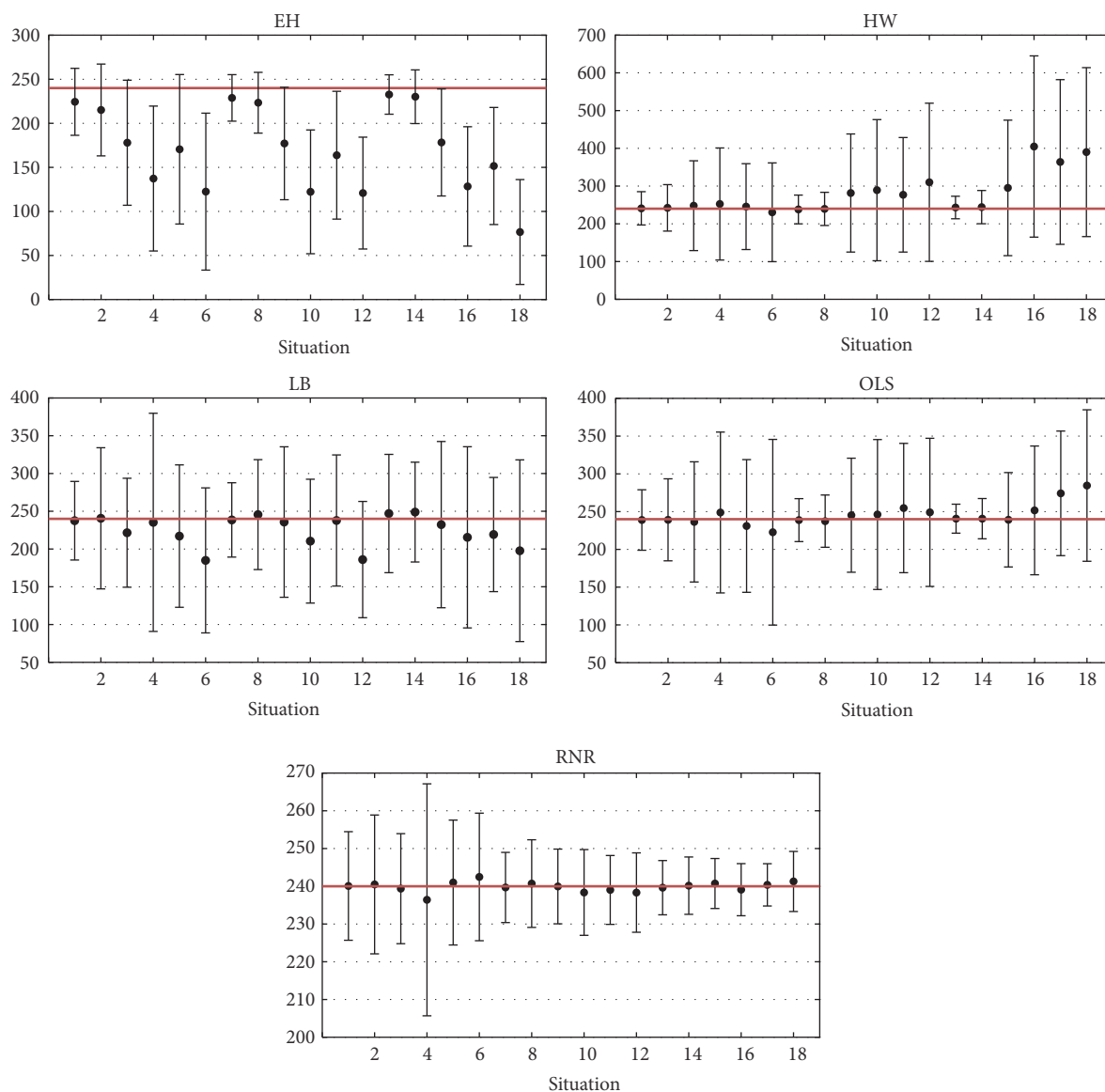


FIGURE 2: Mean estimated values (black dots) and standard deviations of Michaelis constant,  $K_m$ , for different simulated scenarios. Red lines denote true parameter value.

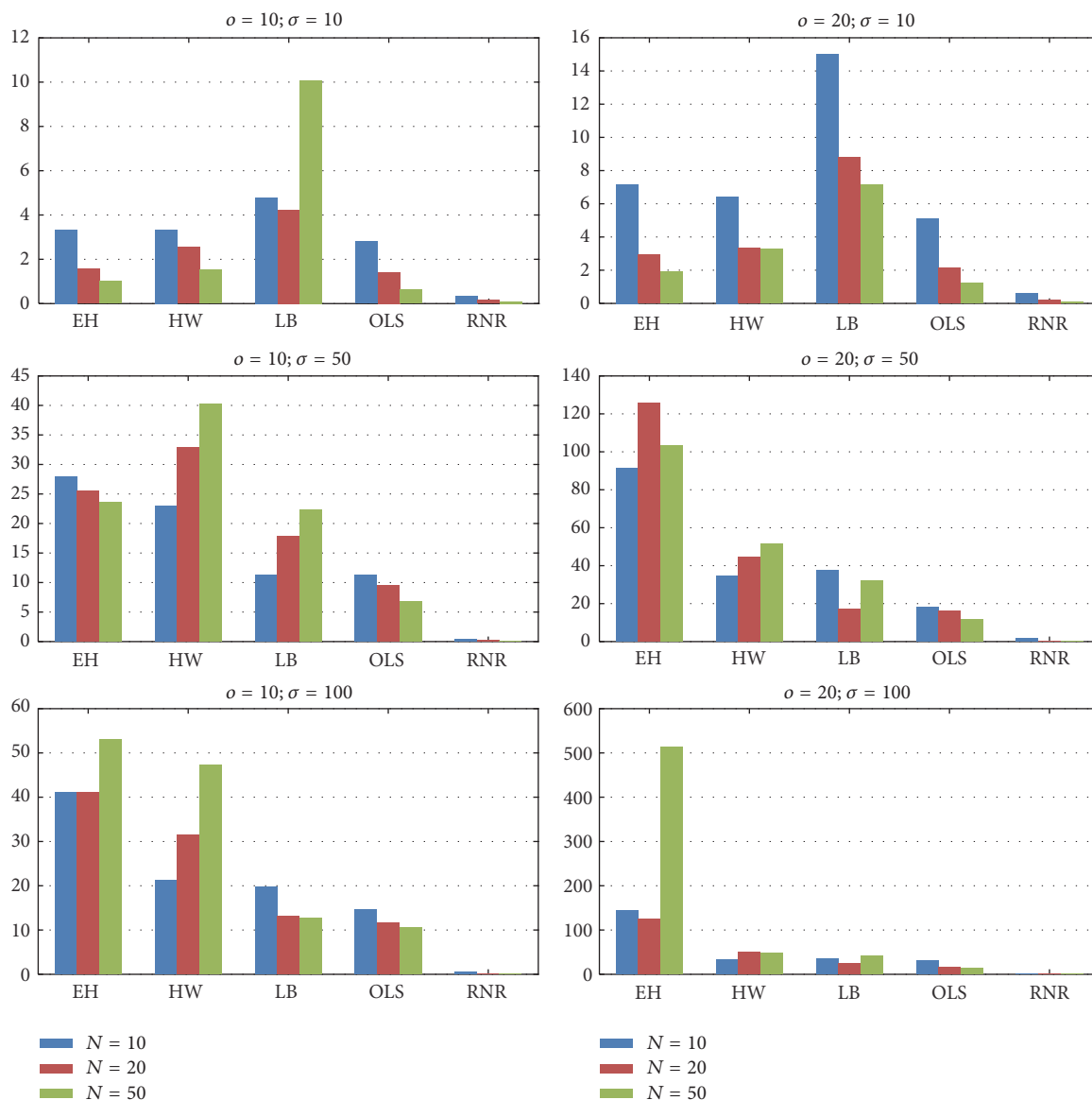
$238.75 \pm 28.43$  and  $79.88 \pm 2.05$ , respectively. Again, if the standard deviations are scaled by an appropriate mean, the resulting coefficients of variation of  $K_m$  and  $V_{\max}$  are 11.5% and 2.7% (EH), 16% and 3.4% (HW), 20.6% and 5.1% (LB), 11.9% and 2.6% (OLS), and 3.9% and 0.8% (RNR), respectively.

Similarly, for a large sample size ( $n = 50$ ) with the same levels of contamination and outlier scatter (Situation 13, Table 1), the values of  $K_m$  and  $V_{\max}$  are estimated as  $232.67 \pm 22.38$  and  $79.71 \pm 0.98$  (EH),  $243.2 \pm 29.88$  and  $79.99 \pm 1.11$  (HW),  $246.97 \pm 78.33$  and  $80.15 \pm 3.43$  (LB),  $240.68 \pm 19.19$  and  $80.08 \pm 0.86$  (OLS), and  $239.62 \pm 7.19$  and  $80.01 \pm 0.31$  (RNR), respectively. In this case, the resulting coefficients of variation of  $K_m$  and  $V_{\max}$  are 9.6% and 1.2% (EH), 12.3% and 1.4% (HW), 31.7% and 4.3% (LB), 8% and 1.1% (OLS), and 3% and 0.4% (RNR), respectively. It should be noted that, all the while in all of the aforementioned cases, the Eadie-Hofstee method

has coefficients of variation that are highly comparable to those based on ordinary least-squares method; the EH estimated  $K_m$  and  $V_{\max}$  values are much further away from the true values than the estimates obtained by Hanes-Woolf, ordinary least-squares, and robust nonlinear regression methods.

With the increase of the contamination level and the outlier scatter, the average estimates of  $K_m$  and  $V_{\max}$  values as per linear plots and the OLS method begin to deviate significantly. However, the modified robust Tukey's biweight estimator is able to keep the errors in check and produce the results that are highly comparable and much closer to the true parameter values.

Numerically, by looking at the estimated values, it is hard to tell which of the selected estimators has the overall best performance; nevertheless, with the help of the normalized mean square error method we can see the values of parameters for

FIGURE 3: Normalized mean square error of  $K_m$  for different scenarios.

which the error is minimum. Thus, from Figures 3 and 5, we may say that the best (most accurate) values of  $K_m$  and  $V_{\max}$  are obtained in situation 17, for which the minimum error values are 0.0541 and  $8.5 \times 10^{-4}$ , respectively (as obtained by RNR). This proves that the estimated values are more or less similar to the true values. The worst error values of 1.6982 and 0.1517 for RNR method are obtained in situation 4 (Figures 3 and 5). In all other situations, the normalized mean square errors for RNR method are less than 1 and 0.1, respectively (Figures 3 and 5). This shows the credibility and the robustness of the proposed modified Tukey's biweight estimator relative to other methods when outliers or influential observations are present in the data. If we compare the robust nonlinear regression method with ordinary least-squares method, we find that the RNR method normalized mean square errors

are on average more than 10 times lower than the normalized mean square errors produced by the OLS method.

**4.2. Parameter Estimation Using Experimental Data.** The viability of the proposed robust estimator was also tested by using the experimental kinetic data for tyrosinase enzyme. The corresponding  $K_m$  and  $V_{\max}$  values, produced by different estimation models, are given in Table 3. Upon closer inspection and analysis of these values, it can be observed that, in case of Ab mushroom and potato tyrosinase, the kinetic values, that is,  $K_m$  and  $V_{\max}$ , yielded by the RNR method (599  $\mu\text{mol}$  and 0.38  $\mu\text{mol}/\text{min}$  for Ab mushroom and 10740  $\mu\text{mol}$  and 0.118  $\mu\text{mol}/\text{min}$  for potato, resp.) are much closer to the values yielded by HW plot (555  $\mu\text{mol}$  and 0.381  $\mu\text{mol}/\text{min}$  for Ab mushroom and

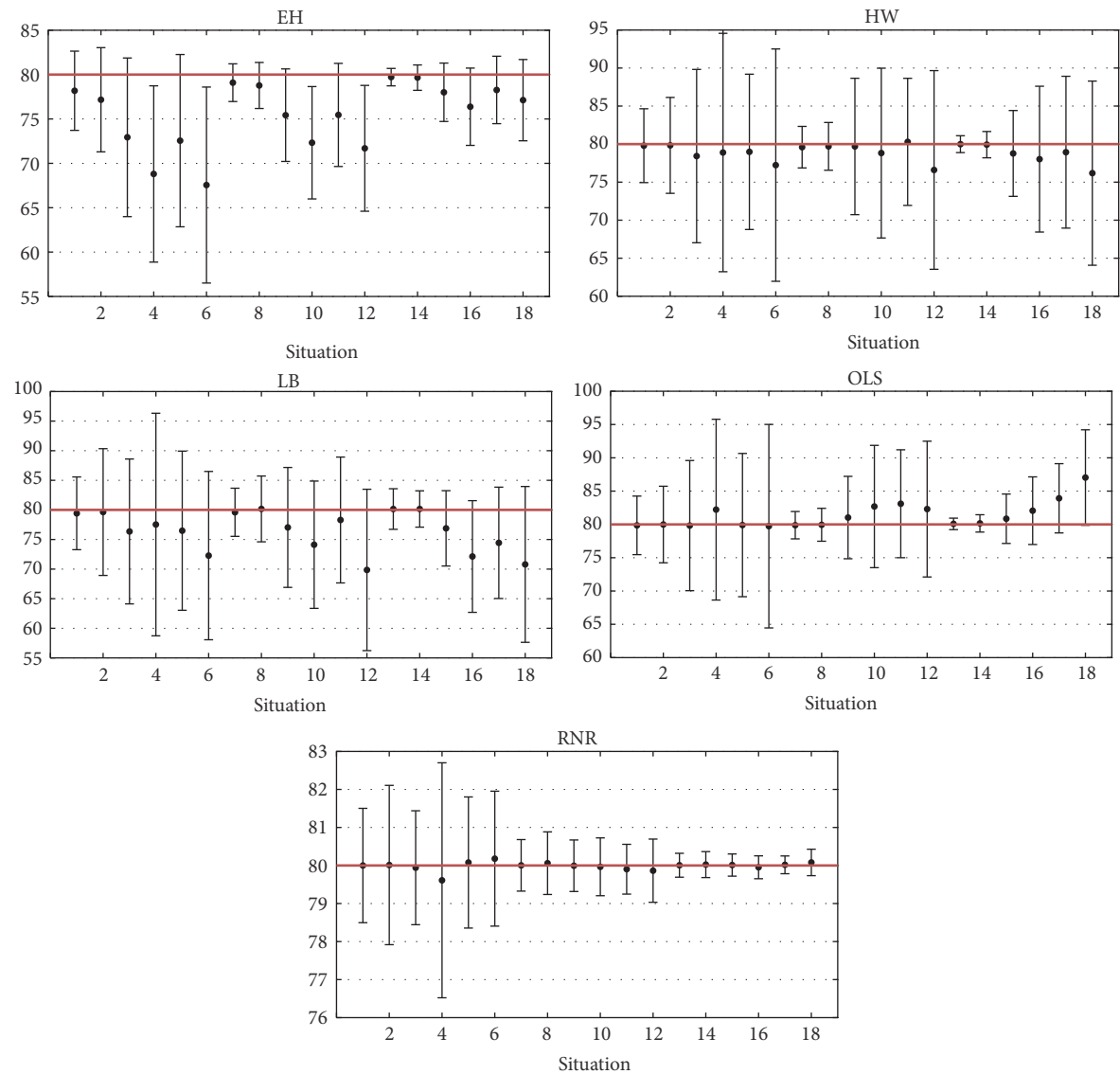
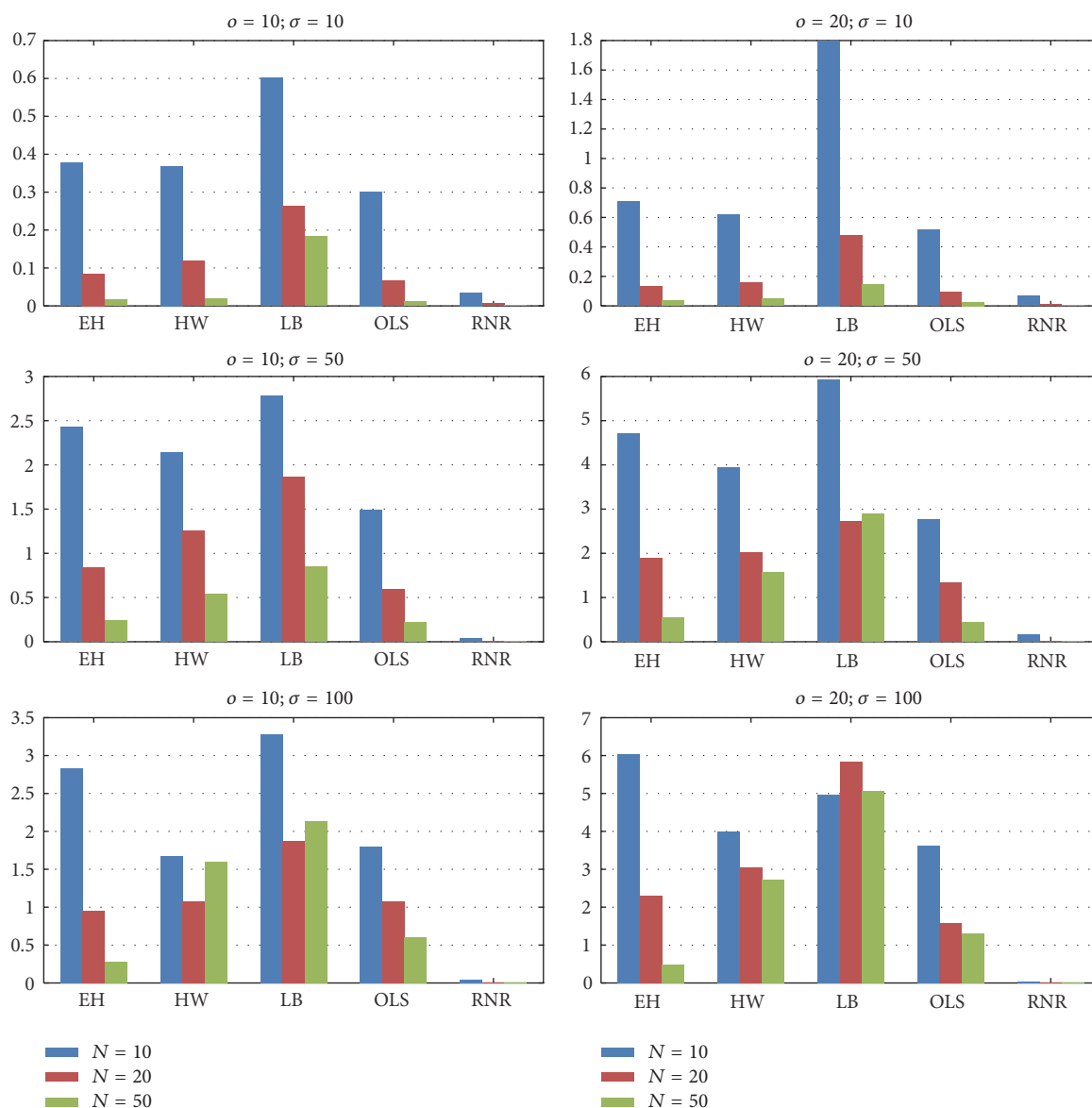


FIGURE 4: Mean estimated values (black dots) and standard deviations of maximum initial velocity,  $V_{\max}$  for different simulated scenarios. Red lines denote true parameter value.

TABLE 3: Kinetic parameters,  $K_m$  and  $V_{\max}$ , values, and mean absolute errors for tyrosinase extracted from different sources, estimated using different methods.

Source	Parameter	Method				
		LB	HW	EH	OLS	RNR
Ab	$K_m$	263.09	555.18	414.09	544.79	599.41
	$V_{\max}$	0.2484	0.3812	0.3470	0.3767	0.3798
	MAE	0.0511	0.0071	0.0137	0.0064	0.0048
Po	$K_m$	359.91	3592.1	992.27	5700.3	6630.5
	$V_{\max}$	0.0017	0.0079	0.0042	0.0095	0.0099
	MAE	0.0013	$2.6 \times 10^{-4}$	$7.5 \times 10^{-4}$	$1.7 \times 10^{-4}$	$1.5 \times 10^{-4}$
Potato	$K_m$	1216.0	10659	2093.4	25720	10740
	$V_{\max}$	0.0208	0.1104	0.0394	0.2086	0.1180
	MAE	0.0139	0.0022	0.0087	0.0018	0.0014



FIGURE 5: Normalized mean square error of  $V_{\max}$  for different scenarios.

10659  $\mu\text{mol}$  and 0.11  $\mu\text{mol}/\text{min}$  for potato, resp.) than by OLS method (545  $\mu\text{mol}$  and 0.376  $\mu\text{mol}/\text{min}$  for Ab mushroom and 25720  $\mu\text{mol}$  and 0.209  $\mu\text{mol}/\text{min}$  for potato, resp.). Furthermore, it is interesting to note that the parameter values yielded by LB plot ( $K_m$  263  $\mu\text{mol}$  and  $V_{\max}$  0.248  $\mu\text{mol}/\text{min}$  for Ab mushroom,  $K_m$  360  $\mu\text{mol}$  and  $V_{\max}$  0.002  $\mu\text{mol}/\text{min}$  for Po mushroom, and  $K_m$  1216  $\mu\text{mol}$  and  $V_{\max}$  0.021  $\mu\text{mol}/\text{min}$  for potato) for all three tyrosinase source are very far from the values yielded by other four estimation methods. Figures 7(a), 7(b), and 8(a) show the curves fitted to the experimental data using the modified Tukey's biweight estimator in comparison with the standard linearization methods and the ordinary least-squares non-linear regression. The mean absolute errors between the predicted reaction rates and the actual data are plotted in the right graph in Figure 8. Particularly, the mean errors for

RNR method are 0.0048,  $1.5 \times 10^{-4}$ , and 0.0014, respectively, which shows a good fit of the achieved model.

## 5. Conclusion

When an enzymatic reaction follows Michaelis-Menten kinetics, the equation for the initial velocity of reaction as a function of the substrate concentration is characterized by two parameters, the Michaelis constant,  $K_m$ , and the maximum velocity of reaction,  $V_{\max}$ . Up to this day, these parameters are routinely estimated using one of these different linearization models: Lineweaver-Burke plot ( $1/v$  versus  $1/s$ ), Eadie-Hofstee plot ( $v$  versus  $v/s$ ), and Hanes-Wolfe plot ( $s/v$  versus  $v$ ). Although the linear plots obtained by these methods are very illustrative and useful in analyzing the behavior of enzymes, the common problem they all share

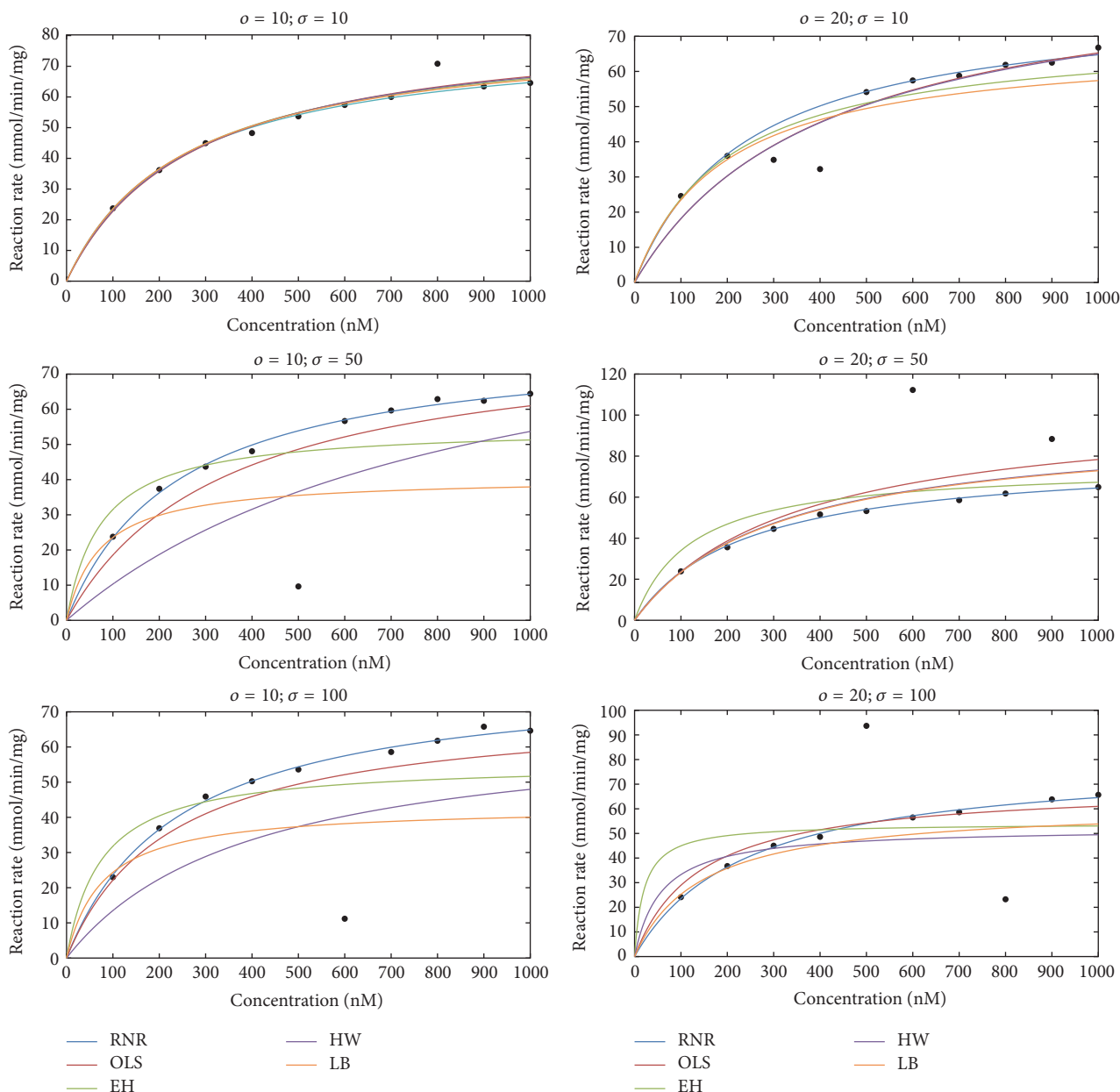


FIGURE 6: Curves fitted using different fitting techniques for random replications of situations 1–6.

is the fact that transformed data usually do not satisfy the assumptions of linear regression, namely, that the scatter of data around the straight line follows Gaussian distribution, and that the standard deviation is equal for every value of independent variable.

More accurate approximation of Michaelis-Menten parameters can be achieved through use of nonlinear least-squares fitting techniques. However, these techniques require good initial guess and offer no guarantee of convergence to the global minimum. On top of that, they are very sensitive to the presence of outliers and influential observations in the data, in which case they are likely to produce biased, inaccurate, and imprecise parameter estimates.

In this paper, a robust estimator of nonlinear regression parameters based on a modification of Tukey's biweight function is introduced. Robust regression techniques have received considerable attention in mathematical statistics literature, but they are yet to receive similar amount of attention by practitioners performing data analysis. Robust nonlinear regression aims to fit a model to the data so that the results are more resilient to the extreme values and are relatively consistent when the errors come from the high-tailed distribution. The experimental comparisons, using both real and synthetic kinetic data, show that the proposed robust nonlinear estimator based on modified Tukey's biweight function outperforms the standard linearization models and

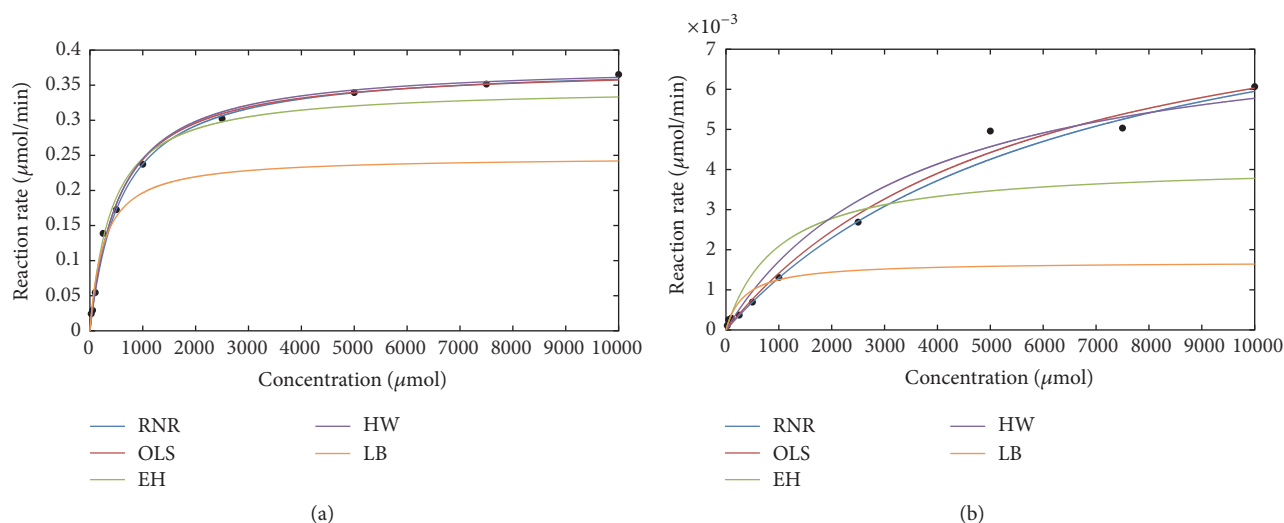


FIGURE 7: Curves fitted using different fitting techniques for tyrosinase extracted from *Agaricus bisporus* (a) and *Pleurotus ostreatus* (b) mushrooms.

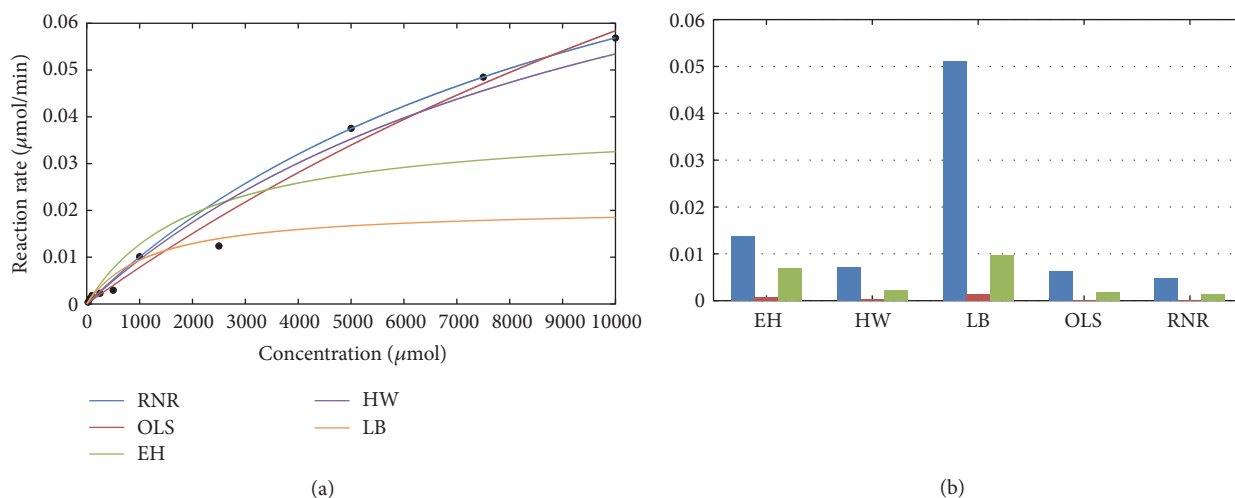


FIGURE 8: Curves fitted using different fitting techniques for tyrosinase extracted from potato (a). Mean absolute errors for different regression models used (b).

ordinary least-squares method and yields superior results with respect to bias, accuracy, and consistency, when there are outliers or influential observations present in the data.

## Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The work reported in this paper was supported by Croatian Science Foundation Research Project “Investigation of Bioactive Compounds from Dalmatian Plants: Their Antioxidant, Enzyme Inhibition and Health Properties” (IP 2014096897).

## References

- [1] H. J. Fromm and M. Hargrove, *Essentials of Biochemistry*, Springer, Berlin, Germany, 2012.
- [2] K. A. Johnson and R. S. Goody, “The original Michaelis constant: translation of the 1913 Michaelis-Menten paper,” *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, 2011.
- [3] A. Cornish-Bowden, “The origins of enzyme kinetics,” *FEBS Letters*, vol. 587, no. 17, pp. 2725–2730, 2013.
- [4] H. Lineweaver and D. Burk, “The determination of enzyme dissociation constants,” *Journal of the American Chemical Society*, vol. 56, no. 3, pp. 658–666, 1934.
- [5] G. S. Eadie, “The inhibition of cholinesterase by physostigmine and prostigmine,” *The Journal of Biological Chemistry*, vol. 146, pp. 85–93, 1942.
- [6] C. S. Hanes, “Studies on plant amylases,” *Biochemical Journal*, vol. 26, no. 5, pp. 1406–1421, 1932.

- [7] A. Ferst, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman, 2000.
- [8] G. N. Wilkinson, "Statistical estimations in enzyme kinetics," *The Biochemical journal*, vol. 80, pp. 324–332, 1961.
- [9] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression*, GraphPad Software Inc, 2003.
- [10] C. Cobelli and E. Carson, *Introduction to Modeling in Physiology and Medicine*, Academic Press, 2008.
- [11] S. R. Nelatury, C. F. Nelatury, and M. C. Vagula, "Parameter estimation in different enzyme reactions," *Advances in Enzyme Research*, vol. 2, no. 1, pp. 14–26, 2014.
- [12] G. Kemmer and S. Keller, "Nonlinear least-squares data fitting in Excel spreadsheets," *Nature Protocols*, vol. 5, no. 2, pp. 267–281, 2010.
- [13] C. Lim, P. K. Sen, and S. D. Peddada, "Robust nonlinear regression in applications," *Journal of the Indian Society of Agricultural Statistics*, vol. 67, no. 2, pp. 215–234, 291, 2013.
- [14] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2006.
- [15] V. Barnett and T. Lewis, *Outliers in statistical data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, England, Third edition, 1994.
- [16] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [17] F. Mosteller and J. W. Tukey, *Data Analysis and Regression*, Addison-Wesley, 1977.
- [18] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.
- [19] M. R. Loizzo, R. Tundis, and F. Menichini, "Natural and synthetic tyrosinase inhibitors as antibrowning agents: an update," *Comprehensive Reviews in Food Science and Food Safety*, vol. 11, no. 4, pp. 378–398, 2012.
- [20] S. Y. Lee, N. Baek, and T.-G. Nam, "Natural, semisynthetic and synthetic tyrosinase inhibitors," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 31, no. 1, pp. 1–13, 2016.
- [21] K. U. Zaidi, A. S. Ali, S. A. Ali, and I. Naaz, "Microbial tyrosinases: promising enzymes for pharmaceutical, food bioprocessing, and environmental industry," *Biochemistry Research International*, vol. 2014, Article ID 854687, 16 pages, 2014.
- [22] Z. Yang and F. Wu, "Catalytic properties of tyrosinase from potato and edible fungi," *Biotechnology*, vol. 5, no. 3, pp. 344–348, 2006.

