bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

On the estimation errors of K_M and V from time-course experiments using the Michaelis–Menten equation

Wylie Stroberg[a], Santiago Schnell[a,b,c,*]

[a] *Department of Molecular & Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA*
[b] *Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA*
[c] *Brehm Center for Diabetes Research, University of Michigan Medical School, Ann Arbor, MI 48105, USA*

Abstract

The conditions under which the Michaelis–Menten equation accurately captures the steady-state kinetics of a simple enzyme-catalyzed reaction is contrasted with the conditions under which the same equation can be used to estimate parameters, K_M and V, from progress curve data. The conditions for the validity of the underlying assumptions leading to the Michaelis–Menten equation are shown to be necessary, but not sufficient to guarantee accurate estimation of K_M and V. Detailed error analysis and numerical "experiments" are used to show the required experimental conditions for the independent estimation of both K_M and V from progress curve experiments. It is found that, if the initial substrate concentration is of the same order of magnitude as K_M, the estimated values of the K_M and V will correspond to their true values calculated from the microscopic rate constants of the corresponding mass-action system.

Keywords: experimental design, parameter estimation, reproducibility, inverse problem.

*Corresponding author.
Email addresses: stroberg@umich.edu (Wylie Stroberg), schnells@umich.edu (Santiago Schnell)

August 5, 2016

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1. Introduction

The fundamental equation of enzyme kinetics is the Michaelis–Menten (MM) equation, which relates the rate of an enzyme-catalyzed reaction to the concentration of substrate [1, 2]. The MM equation is nowadays derived using the steady-state assumption as proposed by Briggs and Haldane [3]. It is characterized by two parameters: the Michaelis constant, K_M, which acts as an apparent dissociation constant under the assumption of steady-state, and the limiting rate, V (or the catalytic constant, k_{cat} if the enzyme concentration is known) [4]. These parameters are often viewed as thermodynamic properties of an enzyme–substrate pair, and hence depend on conditions such as pH and temperature, but not on time-dependent enzyme nor substrate concentrations [5]. As a result, measuring K_M and V are essential to characterizing enzymatic reactions [6]. However, the treatment of K_M and V as constants with respect to enzyme and substrate concentrations relies on simplifying assumptions relating to the quasi-steady-state of the intermediate complex formed by the enzyme and substrate [7]. If conditions for the reaction lie outside the range for which the simplifying assumptions are valid, K_M becomes dependent on the concentration of the substrate, and hence, on time. Experiments to estimate K_M must be conducted under conditions for which the MM equation is valid [7, 8]. This can be problematic since it is generally necessary to know K_M a priori in order to insure the experimental conditions meet the requirements for the using MM equation. Additionally, values of K_M and V measured under valid experimental conditions can only be transferred to cases that also meet the requirements. Since this is often not the case in vivo, using values of K_M and V measured in vitro to predict the activity of an enzyme in living organisms can often be seriously unreliable [9].

The range of substrate and enzyme concentrations over which the MM equation can be applied has long history of theoretical investigation [see 8, for a recent review], and requires two assumptions be valid. The first, called the steady-state assumption, implies that the timescale for the formation of the intermediate complex is much faster than that of the conversion of the substrate into product [10]. The second, called the reactant-stationary assumption, implies that the fast, transient period in which the steady-state population of intermediate complex first forms, depletes only a negligible amount of substrate [11]. It has been shown that the reactant-stationary assumption is more restrictive and, if valid, the reaction velocity (after the ini-

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

tial transient period) will follow the MM equation and be well-characterized by the parameters K_M and V [10, 12, 8].

At first sight, it is tempting to assume that, when the MM equation is valid, experimental data should also yield accurate estimates of K_M and V [13, 14]. However, the conditions for the validity of the steady-state and reactant-stationary assumptions are based on a forward problem, i.e. one in which the parameters are known. Estimating parameters from experimental data, on the other hand, is an inverse problem [15]. Extracting true values of parameters from data requires a stable and unique inverse mapping that is not guaranteed by the existence of a solution to the forward problem [see 16, for example]). Hence, even in cases where the assumptions underlying the MM equation are valid, and the MM equation accurately fits an experimental progress curve, the values of K_M and V estimated from the data may differ significantly from their true values.

In this work, we seek to address the issue of estimating parameters from progress curves of single-substrate, single-enzyme-catalyzed reactions quantitatively. In Section 2, we review the validity of the steady-state and reactant-stationary assumptions, and quantify errors incurred by making these assumptions. In Section 3, we discuss the inverse problem associated with estimating parameters based on the MM equation. Numerical experiments are then conducted in Section 4 to verify and quantify the range of experimental conditions that allow for veracious estimations of K_M and V. We conclude with a discussion of the results in Section 5.

2. The forward problem: the Michaelis–Menten equation and the conditions for its validity

In the simplest, single-enzyme and single-substrate reaction, the enzyme E reacts with the substrate S to form and intermediate complex C, which then, under the action of the enzyme, forms a product P and releases the enzyme,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_{\text{cat}}}{\rightarrow} E + P \tag{1}$$

where k_1 and k_{-1} are microscopic rate constants, and k_{cat} is the catalytic constant [4]. Applying the law of mass action to reaction mechanism (1)

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

yields four rate equations

$$\dot{e} = -k_1 es + k_{-1}c + k_{\text{cat}}c \tag{2a}$$
$$\dot{s} = -k_1 es + k_{-1}c \tag{2b}$$
$$\dot{c} = k_1 es - k_{-1}c - k_{\text{cat}}c \tag{2c}$$
$$\dot{p} = k_{\text{cat}}c, \tag{2d}$$

where lowercase letters represent concentrations of the corresponding uppercase species. Typically, in test tube enzyme binding assays the initial conditions are taken to be

$$(e, s, c, p)\,|_{t=0} = (e_0, s_0, 0, 0)\,. \tag{3}$$

Additionally, the system obeys two conservation laws, the enzyme and substrate conservation laws,

$$e\,(t) + c\,(t) = e_0 \tag{4a}$$
$$s\,(t) + c\,(t) + p\,(t) = s_0. \tag{4b}$$

Using (4a) to decouple the enzyme concentrations, the redundancies in the system (2) are eliminated to yield

$$\dot{s} = -k_1\,(e_0 - c)\,s + k_{-1}c \tag{5a}$$
$$\dot{c} = k_1\,(e_0 - c)\,s - (k_{-1} + k_{\text{cat}})\,c \tag{5b}$$

where $e(t)$ and $p(t)$ are readily calculated once $s(t)$ and $c(t)$ are known. If, after an initial, rapid buildup of c, the rate of depletion of c approximately equals its rate of formation, c is assumed to be in a quasi-steady state [3], i.e.

$$\dot{c} \approx 0 \quad \text{for} \quad t > t_c, \tag{6}$$

where t_c is the timescale associated with the initial transient buildup of c [10]. The steady-state assumption (6), in combination with (5), leads to

$$c = \frac{e_0 s}{K_M + s} \tag{7a}$$
$$\dot{s} = -\frac{V s}{K_M + s}, \tag{7b}$$

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where $V = k_{\text{cat}} e_0$ and $K_M = (k_{-1} + k_{\text{cat}})/k_1$. Hence, the system (2) is reduced to an algebraic-differential equation systems with one single differential equation for s. However, since (7) is only valid after the initial transient time period, t_c, a boundary condition for s at $t = t_c$ must be supplied. To do this, it is assumed that very little substrate is consumed during the initial transient period (the reactant-stationary assumption) such that

$$s(t < t_c) \approx s_0, \tag{8}$$

which provides an initial condition for (5a) under the variable transformation $t \to t - t_c$. Substituting (7a) into (2d), one obtains, the rate of the reaction (1)

$$\dot{p} = v = \frac{Vs}{K_M + s}, \tag{9}$$

relating the rate of product formation to the substrate concentration. Equations (9) is the MM equations, and the system of equations (7a), (7b), and (9) govern the dynamics the complex, substrate, and product, respectively, under the steady-state assumption.

The conditions under which the steady-state assumption (6) and reactant-stationary assumption (8) are valid have been extensively studied. Segel [10] showed that the steady-state assumption is valid so long as

$$\frac{e_0}{K_M + s_0} \ll \left(1 + \frac{K_S}{K}\right)\left(1 + \frac{s_0}{K_M}\right), \tag{10}$$

where $K_S = k_{-1}/k_1$, and $K = k_{\text{cat}}/k_1$. For the reactant-stationary assumption to be valid, they derived the condition

$$\frac{e_0}{K_M} \ll \left(1 + \frac{s_0}{K_M}\right), \tag{11}$$

which is more stringent than condition (10), and hence dictates the conditions under which the MM equation can be applied. Interestingly, it has been shown that condition (11) is independent from (10) [11].

2.1. Quantitative analysis of the errors induced by the steady-state and reactant-stationary assumptions

To gain a quantitative understanding of the inequalities expressed in (10) and (11), an accurate assessment of the difference between the solution to

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

system (5) and the reduced equations (7) is required. For our analysis, we compare progress curves of the substrate calculated with numerical solutions to the exact law of mass action system (5a) and the reduced equation (7b) under the steady-state assumption. Note that the reduced rate equation (7b) is effectively the MM equation for the substrate depletion. The *concentration error* as a function of time is calculated as

$$\text{error}(t) = \left| \frac{s(t) - s_{MM}(t)}{s_0} \right|, \tag{12}$$

where s_{MM} is the substrate concentration calculated using the reduced equation (7b) and $||$ denotes the absolute value. To form a scalar measure of the error, we use the maximal value of the concentration error over the time course of the reaction. Contours of the maximum concentration error in the plane of initial enzyme and substrate concentrations (normalized by K_M) are shown in **Fig. 1**. Additionally, conditions (10) and (11) are plotted for the cases when the right-hand sides are ten times the left-hand sides to represent the much less condition numerically. For all values of $\kappa = k_{-1}/k_{\text{cat}} = K_S/K$, condition (11) is sufficient to guarantee small errors when using the MM equation. However, **Fig. 1A** shows that when κ is small – implying the reverse step in reaction (1) is negligible – small values of s_0/K_M yield small errors, regardless of the initial enzyme concentration.

The observed errors can be understood by considering the influence of small κ and s_0/K_M on the system (5). When $\kappa \ll 1$, reaction (1) strongly favors the production for P from C as opposed to the disassociation of C back to E and S. This reduces the reaction mechanism (1) to the van Slyke–Cullen mechanism [17] as $K_M \approx K$. The requirement $s_0/K_M \gg 1$ implies that the formation of C is slow compared to the formation P and the disassociation of C. Taken together, these two requirements provide an ordering of timescales such that the formation of C is slow compared to the action of the enzyme to form P, but fast compared to the disassociation of the intermediate complex, effectively reducing the rate equation for the substrate depletion (5a) to $\dot{s} \approx -k_1 e_0 s$. Similarly, under the same condition, the MM equation for substrate (7b) reduces to the same expression. Hence, under these conditions, the MM equation accurately represent the system dynamics, even though condition (11) is violated.

The condition for the validity of the reactant-stationary assumption (11) is a sufficient condition for the MM equation to be valid. In essence, this says that for a known set of parameter values, if the reactant-stationary assump-

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

tion is valid, the dynamics of the reduced system (7) will closely approximate the dynamics of the full system (5). However, the MM equation is often used to estimate K_M and V from experimental data, which requires solving an inverse problem. Solutions to the forward problem do not guarantee the existence or uniqueness of the inverse problem, hence it is not clear that the conditions for the validity of the reduced forward problem correspond to the conditions required to accurately estimate rate constants. This issue is investigated in the following section.

3. The inverse problem: Estimation of K_M and V

The experimental estimation of the parameters K_M and V is used to characterize enzyme-catalyzed reactions. In general, K_M and V can be estimated through either initial rate experiments [see 18, for a recent review] or direct analysis of time course data [19]. In initial rate experiments, a series of enzyme assays with differing substrate concentrations are performed and initial reaction rates are calculated from the linear portion of the progress curve (after the initial fast transient, t_c, and before substrate depletion becomes influential). The MM equation for either substrate or product is then fit to the initial rates as a function of initial substrate concentration, yielding K_M and V. When time course data is used, the integrated implicit [20] or closed-form [21] of the MM equations are fit directly to time series through nonlinear regression, providing estimates for K_M and V. Although initial rate experiments are more commonly used, they require numerous assays with different substrate concentrations to determine K_M and V. Alternatively, time course analyses have the advantage that K_M and V can be estimated from a single experiment, making them potentially much cheaper when expensive reactants are required, and less time consuming [22, 23, 24, 7]. Hence, in this work, we consider the problem of parameter estimation directly from progress curves, specifically, those for the concentration of substrate.

Inverse problems are typically formulated in terms of an operator, F, mapping the space of parameters, Q, to the space of observations, Y, i.e.

$$F(q) = y, \tag{13}$$

where $q \in Q$ is a vector of parameters, and $y \in Y$ is a vector of observed quantities. In general, $F = G \circ H$ is a composite of the solution operator S, which maps a parameter vector q to a solution vector $\bar{y}$ of the underlying

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ordinary differential equation for the rate equations, and and the observation operator R, which takes $\bar{y}$ to the observable y [25]. For example, if fluorescent markers are used to tag substrate molecules, and fluorescent intensity is measured at times t_i, G is then the mapping between the fluorescent intensity at times t_i and substrate concentration, and H is the solution to the rate equations (7). G effectively samples the solution to the rate equation model at the observation times and converts those concentrations to the experimental observables.

For the present study, we assume the concentrations are observed directly, hence G is simply a sampling of the integrated rate equations (5). Specifically, we consider the case in which the concentration of the substrate is measured at discrete times t_i and H is the solution to (7). The inverse problem consists of finding a parameter vector q solving (13). However, (13) is generally ill-posed due to experimental noise. Even in the absence of experimental error, the inverse problem will be ill-posed, because the MM equation only approximates the mass action rate equations (5), even when the steady-state and reactant-stationary assumptions are valid. The exact inverse problem must then be reformulated as a least-squares optimization problem to minimize the function

$$||y - F(q)||_Y^2, \tag{14}$$

where $||\cdot||_Y$ is the L_2 norm on Y. The sensitivity of (14) to changes in parameter values is measured by the local condition number for the first order optimality condition. The condition number is given by the ratio of the maximum and minimum eigenvalues of the matrix

$$J^*\left(q_*\right) J\left(q_*\right). \tag{15}$$

In the above expression, J is the Jacobian of the mapping F, q_* is the "true" parameter vector and J^* denotes the conjugate transpose of J. Ill-conditioning implies small errors in the data (or model) can result in large errors in the estimated parameters. Although many features of a problem can affect the conditioning (such as proper choice of units) [26], of particular importance when fitting the MM equation is the correlation of the parameters. When the parameters are highly correlated the model is incapable of uniquely determining the parameters because, as the correlation coefficient tends toward 1, the parameters become linearly dependent. In this case, at least one column of J will be approximately a linear combination of the others, and hence not invertible.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Effectively, this dictates when the mass action model (5), which depends on three parameters (k_1,k_{-1},k_{cat}), reduces to the MM model (7) with parameters (K_M,V). Under experimental conditions for which the reactant stationary assumption is valid, it is not possible to estimate all three rate constants from the mass action model using time course data. Similarly, within the region of validity for the reactant stationary assumption, there are sub-regions in which rank(J)≈ 1, and hence prohibit estimation of both K_M and V from time course data. To see where this rank deficiency occurs we consider two regions in the s_0/K_M–e_0/K_M plane. In both, the conditions for the validity of the reactant stationary assumption are met. Additionally, in the first case $s_0 \ll K_M$. Since $s < s_0$ for all t, we can expand (5a) in powers of s/K_M. Truncating this expansion at order two leads to

$$\dot{s} = -\frac{Vs}{K_M}\left(1 - \frac{s}{K_M}\right). \tag{16}$$

To lowest order, $\dot{s}$ depends only on the ratio of V to K_M, and hence the inverse problem of finding both parameters from time course data will become extremely ill-conditioned at small substrate concentrations (see, **Fig. 2A**).

Next, consider the case in which the substrate is in great excess, i.e. $s_0 \gg K_M$. Initially, $s \approx s_0$, allowing for an expansion of $\dot{s}$ in powers of K_M/s_0, which, to second order, gives

$$\dot{s} \approx -V\left(1 - \frac{K_M}{s}\right). \tag{17}$$

Hence, so long as $s \gg K_M$, the substrate concentration will decrease linearly with rate $-V$. Eventually, the progress curve must deviate from the initial linearity, and presumably, this curvature should contain information about K_M, allowing for both parameters to be estimated. However, if the time over which the progress curve is nonlinear is small, or equivalently, the initial linear regime very nearly approaches substrate depletion, parameter estimation will fail. Large s_0/K_M can be shown to imply this by comparing the timescale for significant substrate depletion, t_S, with the timescale of high curvature t_Q. The substrate depletion timescale is given by [10]

$$t_S = \frac{\Delta s}{|\dot{s}_{\max}|} = \frac{K_M + s_0}{V}. \tag{18}$$

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The high-curvature timescale can be estimated with the aid of the second derivative of the substrate concentration,

$$\ddot{s} = \frac{V^2 K_M s}{(K_M + s)^3}. \tag{19}$$

t_Q is defined as the ratio of the total change in velocity of the reaction to the maximum acceleration. The maximum acceleration, found by equating (19) with zero, occurs when $s = K_M/2$ for $s_0 \geq K_M/2$, and s_0 otherwise. Since the present analysis concerns high s_0/K_M, the high-curvature timescale is given by

$$t_Q = \frac{\Delta V}{\ddot{s}|_{s=K_M/2}} = \frac{27 K_M s_0}{4V\left(K_M + s_0\right)}, \tag{20}$$

where ΔV is the change in reaction velocity through the region of curvature and is equal to V. Estimation of parameters from time course data will not be possible when $t_Q \ll t_S$, or, upon substitution of (18) and (20), when

$$\frac{4\dfrac{s_0}{K_M}}{27\left(1 + \dfrac{s_0}{K_M}\right)} \ll 1. \tag{21}$$

Therefore, as the initial substrate concentration is increased, the proportion of the time course that can yield information about K_M decreases, and measurements will require greater resolution in both time and concentration. **Fig. 2B** shows the condition number and the ratio of the substrate depletion timescale to the high-curvature timescale for a large range of s_0/K_M. At small values of s_0/K_M, ill-conditioning makes parameter extraction intractable, while at large s_0/K_M, measurements must be increasingly precise. Thus, substrate concentrations close to K_M are desirable when determining parameters.

4. Numerical experiments

To demonstrate and quantify the regions in which the conditioning of the inverse problem is poor, and the necessary measurements become intractable, we present a systematic numerical analysis of progress curve experiments in this section.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4.1. Methodology for numerical progress curve experiments

Numerical experiments consist of first generating progress curve data from the mass action rate equations with a known set of rate constants. Then, the values of K_M and V corresponding to those rate constants are estimated by fitting the MM equations to the progress curve. To generate experimental progress curves it is necessary to choose a set of rate constants $(k_1, k_{-1}, k_{\text{cat}})$, and experimental protocol. The experimental protocol consists of defining initial conditions, (s_0, e_0, c_0, p_0), a time span for the experimental observation, t_{obs}, and a sampling frequency ω. The system of equations (5) are integrated numerically from $t \in [0, t_{\text{obs}}]$ and substrate concentrations are recorded every ω^{-1} time units, leading to $t_{\text{obs}}\,\omega$ data points $\{s_i\,(t_i)\}$.

The data is then fit using the numerically integrated form of (5a). The nonlinear regression used to calculate the parameters (K_M, V) is performed using the Levenberg–Marquardt algorithm as implemented in SciPy (version 0.17.1, http://www.scipy.org). In many cases, supplying good initial conditions for the optimizer used for the regression is crucial to finding accurate parameter estimates. Since, in actual experiments the values of K_M and V are not known a priori, we attempt to roughly estimate their values from the time course data to provide initial conditions for the optimization. To do this, $\{s_i\,(t_i)\}$ is differentiated numerically by central differences to give approximate rates $\{\dot{s}_i\,(t_i)\}$. Then, using (5a), data at any two time points, t_i and t_j can be used to estimate the parameters through

$$K_M = \frac{(\dot{s}_j - \dot{s}_i)\, s_i s_j}{\dot{s}_i s_j - \dot{s}_j s_i} \tag{22}$$

$$V = \dot{s}_i \left(\frac{K_M}{s_i} + 1 \right). \tag{23}$$

In theory, any two points can be used to estimate K_M and V, however, it is best to use data for which the velocity is changing at that greatest rate. Hence, we additionally numerically calculate $\{\ddot{s}_i\,(t_i)\}$ and choose the times directly on either side of the maximum to substitute into (22) and (23). To avoid using data points in the transient regime before the system reaches a quasi-steady state, we consider only the regime for which $s(t_i) < s_0/2$. For actual experiments, noise can make calculations of derivatives subject to large errors, hence smoothing techniques must be used. Additionally, numerous pairs of data points can be used to generate a distribution of estimates, which can then be averaged to give initial conditions for the optimization, similar to

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[27]. Once the initial conditions for the optimization routine are established, the best-fit values of K_M and V can be systematically estimated.

We note that when experimental conditions do not lie in a region for which the reactant stationary assumption is valid, the above technique will naturally provide poor estimates for K_M and V. In these regions, we have also used the true values K_M and V, calculated from the known rate constants, as initial conditions. Both methods provide qualitatively similar results throughout the regions of parameter space investigated here, and quantitatively agree in the region for which the reactant stationary assumption is valid.

4.2. Errors in parameter estimates can be large even when the reactant stationary assumption is valid

Despite the validity of the reactant stationary assumption being sufficient for the MM equation model to closely align with the solution to the mass action governing equations, the inverse problem does not provide accurate estimates for parameters within the same range. Fig. 3 shows errors in estimates of K_M and V for a wide range of e_0/K_M and s_0/K_M. Below and above the range plotted for s_0/K_M, the solutions become numerically unstable due to the conditioning problems discussed in Section 3. It is clear however, that even within the range defined by large and small values of s_0/K_M, significant errors are present. At high s_0/K_M and e_0/K_M, V can be accurately determined, but K_M begins to show significant deviation. This is anticipated from the high-s_0/K_M approximation of the substrate rate equation, which depends only on V. Additionally, when $s_0/K_M < 1$, the error contours follow a line for which $e_0 \approx s_0$. The condition that enzyme concentration is small relative to that of the substrate was one of the earliest conditions for the validity of the MM equations derived from singular perturbation theory [28]. For the forward problem, Segel [10] showed this condition to be overly restrictive, yet it appears to be appropriate for the inverse problem.

An explanation for the condition $e_0 \ll s_0$ can be found by comparing the integrated form of the MM substrate equation with an exponential progress curve that is the limiting solution to the MM equations as s_0/K_M approaches zero. The integrated closed-form of (7b), known as the Schnell–Mendoza equation [21], can be written explicitly in terms of the Lambert-W function

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[29]

$$\frac{s\left(t\right)}{s_0} = \left(\frac{s_0}{K_M}\right)^{-1} W\left[\frac{s_0}{K_M} e^{\left(\frac{s_0}{K_M} - \frac{V}{K_M} t\right)}\right]. \tag{24}$$

Expanding the above expression about zero and truncating at first order leads to

$$\frac{s(t)}{s_0} \approx e^{-\frac{k_{\text{cat}} e_0 t}{K_M}\left(1 - \frac{s_0}{k_{\text{cat}} e_0 t}\right)}, \tag{25}$$

where we have used the definition of V to explicitly show the dependence on e_0. The exponential solution takes the form

$$\frac{s_{\text{exp}}\left(t\right)}{s_0} = e^{-\frac{k_{\text{cat}} e_0 t}{K_M}}. \tag{26}$$

Comparing (25) and (26) shows that the correction provided by the MM solution over the exponential progress curve becomes decreasingly significant as e_0/s_0 becomes large. **Fig 4A** compares the mean concentration errors between the best fit solutions and the "true" solutions for both the MM equation and an exponential fit. At small values of e_0/s_0, the MM equation provides a distinctly better fit than the exponential solution, allowing both K_M and V to be estimated from a single progress curve. As e_0/s_0 increases, the two fitting functions eventually provide the identical fits. This corresponds to an exponential increase in the variance of the estimated parameters (**Fig. 4B**), and indicates that only the ratio V/K_M can be determined in this range.

4.3. Fitting the initial substrate concentration does not significantly alter estimates of K_M and V

Even when the reactant stationary assumption is valid, a small amount of substrate will be consumed in the initial transient period. Hence, knowing the substrate concentration at the start of the reaction may not exactly correspond to that at the start of the quasi-stead-state phase. Although this difference is small, it is not clear whether this can noticeably alter the estimation of K_M and V. Additionally, time course measurements often employ

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

optical techniques to collect concentration data. Without time consuming calibration curve experiments to relate the fluorescent intensity to concentration directly, only relative concentrations are known. For these reasons, s_0 can be treated as an additional unknown parameter for the regression analysis [30].

Fig. 5 shows error contours for estimates of K_M, V and s_0 for different experimental conditions. Similar to when s_0 is assumed known, the errors in K_M and V follow lines of constant e_0/s_0 at low substrate concentration. Additionally, **Fig. 5C** shows that the best-fit value of s_0 corresponds to the true value of the initial substrate concentration for conditions where the reactant stationary assumption is valid. These results indicate that including s_0 as a free parameter can yield similar information about the constants K_M and V, even in those cases when no definite concentrations are known.

4.4. Data noise further reduces the range of conditions providing accurate estimates of K_M and V

In any physical experiment, some finite amount of measurement error will be present. To understand how signal noise affects the estimation of K_M and V, we add noise to the numerically-calculated solution of (5a) such that the data becomes

$$\{s_i(t_i)\}_\delta = \{s_i(t_i)\} = \eta\,(\delta)\,, \tag{27}$$

where η is a pseudo-random number drawn from a Gaussian distribution of mean zero and standard deviation δ. The data is then fit as described in Section 4.1. However, the noise in the data precludes the use of the method described for estimating good initial conditions for the solver. Without a smoothing procedure, the difference formulas (22) and (23) can lead to large errors. In order to eliminate possible uncertainty arising from the determination of good initial guesses from experimental data, we chose the "true" values of K_M and V as the starting point for the optimization algorithm.

Contour plots of the errors in the estimated values of K_M and V for the case of $\delta = 0.01$ are shown in **Fig. 6**. Qualitatively, they exhibit the same behavior as the noise-free error contours (**Fig. 3**), yet the magnitude of the error increases, shrinking the range of initial conditions for which the error will be below a given threshold. A meaningful characterization of the quality of a fit is the variance in the estimated model parameters. To calculate the variance of K_M and V, the covariance matrix is first calculated as

$$\mathrm{Cov} = \left(\bar{J}^T \bar{J}\right)^{-1}, \tag{28}$$

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where $\bar{J}$ is the Jacobian evaluated numerically at the terminal point of the optimization. The variance for K_M and V are then the diagonal elements of Cov. As shown in **Fig. 7**, the range of experimental conditions leading to precise estimates of K_M becomes significantly constrained when even a small amount of measurement error is present. Only in the region where $s_0/K_M \approx 1$ and $e_0/K_M \ll 1$ are the estimated K_M values robust. At larger initial substrate concentrations, the noise in the data sufficiently smears the sharply curved region of the substrate progress curve, making extraction of K_M prone to uncertainty. At small initial substrate concentrations, the added noise reduces the distinction between the exponential and MM solution branches shown in **Fig. 4A**, making independent determination of K_M and V more difficult. Hence, even with only slight measurement error the reliability of estimated parameters falls significantly as the ratio s_0/K_M departs from unity.

5. Discussion

It is now well established that the MM equation accurately captures the kinetics of simple enzyme-catalyzed reactions when the reactant-stationary assumption holds true. However, as we have shown here, this fact does not imply that K_M and V can be obtained from experimental progress curves conducted within the range of validity for the reactant stationary assumption. This highlights an important problem encountered in parameter estimation. Even when the MM equation very accurately fits the experimental data, the fitted parameters may not accurately represent their true values. Without a thorough analysis of the inverse problem, it is not possible to distinguish between good fits that provide poor parameter estimates, and good fits that accurately estimate parameter.

Most of the research done on the analysis of enzyme progress curves has focused on the nonlinear regression analysis and algorithms to fit progress curves data [31, 23, 19, 32]. Additional research has investigated the design of progress curve experiments from a computational and theoretical standpoint [33]. In these works, either experimental data is collected, or artificial data is generated by adding noise to numerical solutions the integrated MM equation for prescribed values of K_M and V. Then, the artificial data is fitted in order to estimate K_M and V. Although this procedure can identify values of K_M and V for which progress curves can be well-fit by the integrated MM equation, it makes no connection to the underlying microscopic rate

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

constants. Hence, these studies do not directly assess whether the predicted values of K_M and V are connected to their microscopic definitions. In the present study we have addressed this issue by extracting data from numerical solutions to the underlying mass-action system for prescribed microscopic rate constants, then comparing the predicted values of K_M and V with those derived from the prescribed values of k_1, k_{-1}, and k_{cat}.

The detailed error analysis presented in this work provides guidelines for the ranges of experimental conditions allowing for true parameter estimation. Specifically, we see that, in order for both K_M and V to be derived from substrate progress curve measurements:

- The initial substrate concentration must be within approximately an order of magnitude of the Michaelis constant, that is $s_0 = O(K_M)$, especially when significant noise is present in the data.

- When the initial substrate concentration is in great excess of the Michaelis constant, that is $s_0 \gg K_M$, a linear fit to the initial velocity will yield V, but provide no information about K_M.

- When the initial substrate concentration is small compared to the Michaelis constant, that is $s_0 < K_M$, an exponential fit to the progress curve will provide an estimate for the ratio of V to K_M, but neither parameter independently.

The recommendations presented here coincide with those of Duggleby and Clarke [33], who recommend an initial substrate concentration of approximately $2.5K_M$. However, we additionally provide error contours for parameters estimated from experiments conducted under conditions far from this optimal value. This analysis shows that reasonable estimates can be expected so long as the initial substrate concentration is within an order of magnitude of $2.5K_M$. Furthermore, noise in the data restricts this range to be significantly smaller that the theoretical range for the validity of the MM equation.

In general, since these requirements depend on K_M, they cannot be assessed before conducting an experiment. However, they do provide useful checks that can reduce the number of experiments required, especially when compared to parameter estimation based on initial rate experiments. If a progress curve for a given initial substrate concentration cannot be fit by an exponential, and has a curvature that can be resolved, nonlinear regression

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of the progress curve will provide accurate estimates of both K_M and V. If, say, the progress curve can reasonably be fit by an exponential, a second experiment with substantially larger initial substrate concentration should be performed. Then, the second progress curve, so long as the increase in initial substrate concentration is great enough to surpass the substrate range for with the kinetics are exponential, should yield either a curve from which both parameters can be estimated, or a curve from which V can be estimated. Hence, the two experiments are sufficient to determine both K_M and V. This is in contrast to initial rate experiments, which require a large number of experiments such that a curve of the initial reaction velocity as a function of s_0 can be produced. For accurate measurement of K_M and V from initial rate experiments, both large and small values of the substrate (relative to K_M) must be used [14, 18]. Hence, progress curve analysis will always require fewer experiments than initial rate experiments. Additionally, if initial rate experiments are used, progress curve analysis can be used as a check the accuracy of the estimates. Values of K_M and V obtained from fitting (5a) to the initial rate data should correspond to those values obtained from progress curve analysis of the experiments for which the initial rates are intermediate between 0 and V.

In conclusion, this work both advocates and cautions the use of progress curve analysis in determining kinetic parameters for enzymatic reactions. Progress curve assays can greatly reduce the number of experiments (and hence the cost and quantity of reagents) needed, while still providing accurate measurements. However, it is essential to not conflate an accurate fit with an accurate estimate of K_M and V. If this is kept in mind, progress curve analysis has significant advantages over the use of initial rate experiments.

Acknowledgments

This work is supported by the University of Michigan Protein Folding Diseases Initiative.

References

[1] A. Cornish-Bowden, The origins of enzyme kinetics, FEBS Lett. 587 (2013) 2725–2730.

[2] A. Cornish-Bowden, One hundred years of Michaelis-Menten kinetics, Persp. Sci. 4 (2015) 3–9.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[3] G. E. Briggs, J. B. S. Haldane, A note on the kinetics of enzyme action, Biochem. J. 19 (1925) 338–339.

[4] A. Cornish-Bowden, Current IUBMB recommendations on enzyme nomenclature and kinetics, Persp. Sci. 1 (2014) 74–87.

[5] D. L. Purich, Enzyme kinetics: Catalysis & Control: A reference of theory and best-practice methods, Academic Press, London, 2010.

[6] A. Cornish-Bowden, Analysis and interpretation of enzyme kinetic data, Persp. Sci. 1 (2014) 121–125.

[7] S. Schnell, P. K. Maini, A century of enzyme kinetics. Reliability of the K_M and $v_{\max}$ estimates, Comments Theor. Biol. 8 (2003) 169–187.

[8] S. Schnell, Validity of the Michaelis-Menten equation – Steady-state, or reactant stationary assumption: that is the question, FEBS J. 281 (2014) 464–472.

[9] K. van Eunen, B. M. Bakker, The importance and challenges of in vivo-like enzyme kinetics, Persp. Sci. 1 (2014) 126–130.

[10] L. A. Segel, On the validity of the steady-state assumption of enzyme kinetics, Bull. Math. Biol. 50 (1988) 579–593.

[11] S. M. Hanson, S. Schnell, Reactant stationary approximation in enzyme kinetics, J Phys Chem A 112 (2008) 8654–8658.

[12] L. A. Segel, M. Slemrod, The quasi-steady-state assumption: a case study in perturbation, SIAM Rev. 31 (1989) 446–477.

[13] R. G. Duggleby, J. F. Morrison, Experimental designs for estimating the kinetic parameters for enzyme-catalysed reactions, J. Theor. Biol. 81 (1979) 671–684.

[14] R. J. Ritchie, T. Prvan, A simulation study on designing experiments to measure the K_M of the Michaelis–Menten kinetics curves, J. theor. Biol. 178 (1996) 239–254.

[15] J. A. Jacquez, The inverse problem for compartmental systems, Math. Comput. Simulation 24 (1982) 452–459.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[16] W. Stroberg, S. Schnell, On the validity and errors of the pseudo-first-order kinetics in ligand-receptor binding, bioRxiv (2016) 051136.

[17] D. D. Van Slyke, G. E. Cullen, The mode of action of urease and of enzymes in general, J. Biol. Chem. 19 (1914) 141–180.

[18] H. Bisswanger, Enzyme assays, Persp. Sci. 1 (2014) 41–55.

[19] C. T. Goudar, J. R. Sonnad, R. G. Duggleby, Parameter estimation using a direct solution of the integrated Michaelis–Menten equation, Biochim. Biophys. Acta 1429 (1999) 377–383.

[20] A. Cornish-Bowden, Fundamentals of enzyme kinetics, third edition ed., Portland Press, London, 2004.

[21] S. Schnell, C. Mendoza, Closed form solution for time-dependent enzyme kinetics, J. theor. Biol. 187 (1997) 207–212.

[22] R. G. Duggleby, J. F. Morrison, The analysis of progress curves for enzyme-catalysed reactions by non-linear regression, Biochim. Biophys. Acta 481 (1977) 297–312.

[23] R. G. Duggleby, Analysis of enzyme progress curves by nonlinear regression, Methods Enzymol. 249 (1995) 61–90.

[24] S. Schnell, C. Mendoza, A fast method to estimate kinetic constant for enzyme inhibitors, Acta Biotheor. 49 (2001) 109–113.

[25] H. W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, P. Schuster, Inverse problems in systems biology, Inverse Problems 25 (2009) 123014, 51.

[26] P. L. Bonate, Pharmacokinetic-Pharmacodynamic Modeling and Simulation, Springer, New York, 2011.

[27] T. L. Toulias, C. P. Kitsos, Fitting the Michaelis–Menten model, J. Comput. Appl. Math. 296 (2016) 303–319.

[28] F. G. Heineken, H. M. Tsuchiya, R. Aris, On the mathematical status of the pseudo-steady hypothesis of biochemical kinetics, Math. Biosci. 1 (1967) 95–113.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[29] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth, On the Lambert W function, Adv. Comput. Math. 5 (1996) 329–359.

[30] F. G. Heineken, H. M. Tsuchiya, R. Aris, On the accuracy of determining rate constants in enzymatic reactions, Math. Biosci. 1 (1967) 115–141.

[31] R. G. Duggleby, A nonlinear regression program for small computers, Anal. biochem. 110 (1981) 9–18.

[32] R. G. Duggleby, Quantitative analysis of the time courses of enzyme-catalyzed reactions, Methods 24 (2001) 168–174.

[33] R. G. Duggleby, R. B. Clarke, Experimental designs for estimating the parameters of the Michaelis-Menten equation from progress curves of enzyme-catalysed reactions, Biochim. Biophys. Acta 1080 (1991) 231–236.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

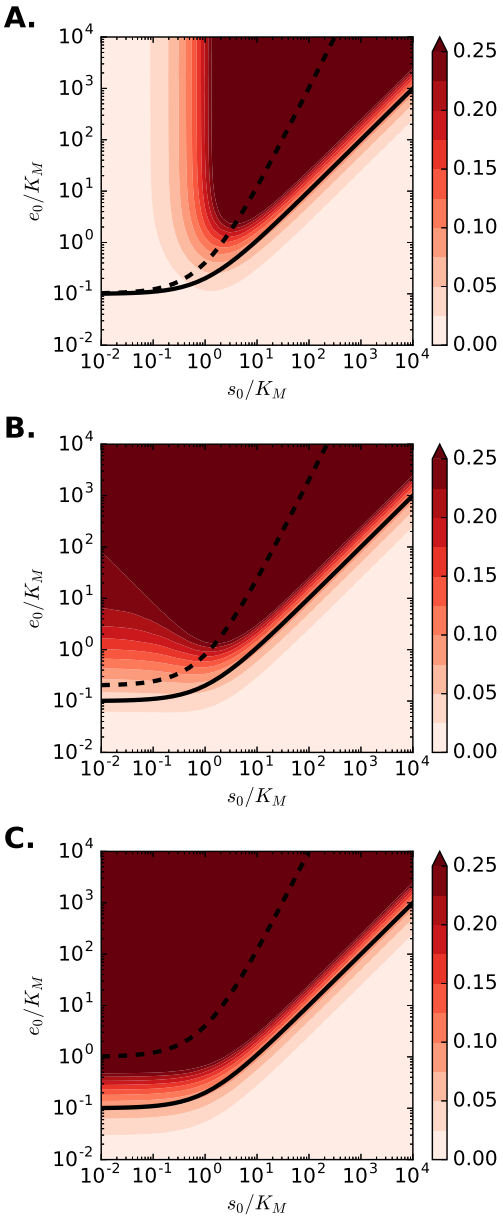

Figure 1: **Concentration error contours in the e_0/K_M–s_0/K_M plane.** The maximal concentration errors are plotted in the plane of initial enzyme and substrate concentrations, normalized by K_M. The dashed black line corresponds to the condition for steady-state assumption (10), while the solid black line corresponds to the reactant stationary condition (11). Each panel shows different values of K_S and K, while $K_M = 1$ for all cases. Panel **A**: $K_S = 0.1$, $K = 0.9$; Panel **B**: $K_S = 0.5$, $K = 0.5$; Panel **C**: $K_S = 0.9$, $K = 0.1$.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

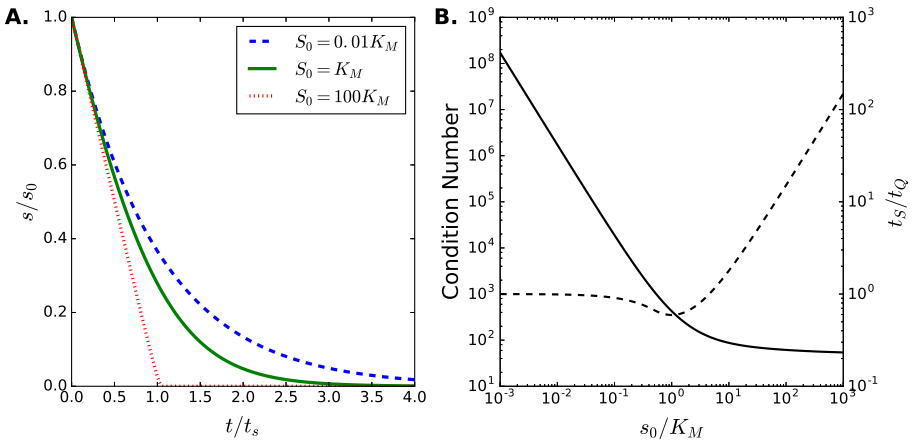

Figure 2: **(A) Substrate progress curves for high, intermediate and low initial substrate concentrations.** Substrate concentrations for differing values of S_0/K_M are plotted as a function of time. When the initial substrate concentration is large (red dotted line) the substrate depletion is linear until nearly all substrate has been depleted. With low initial substrate concentration (blue dashed line), the depletion follows a simple exponential. At intermediate values, the concentration follows the full hyperbolic rate law and both K_M and V can be uniquely identified through regression. For all curves, $V = 10^{-2}$ M sec^{-1}. **(B) Condition number (solid line) and timescale ratio t_S/t_Q as functions of the s_0/K_M.** At small values of s_0/K_M, the inverse problem becomes ill-conditioned. At large values of s_0/K_M, the region of the progress curve providing information about K_M becomes increasingly small.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

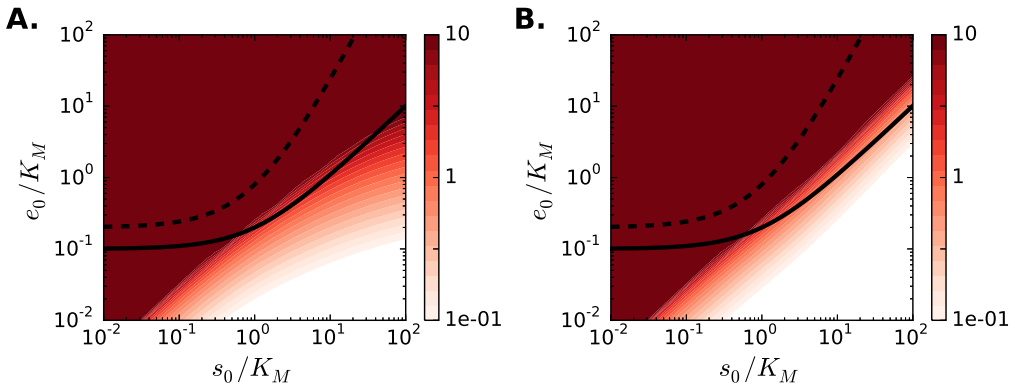

Figure 3: **Error contours of the estimated values K_M and V.** Errors in the predicted values of K_M and V for different initial substrate and enzyme concentrations are shown to deviate from the conditions for the validity of the reactant stationary assumption (shown as the solid line). The dashed line corresponds to condition (10). For s_0/K_M values lower than 10^{-2} and greater than 10^2, the fitting algorithm becomes unstable. Note that the color bar scale is logarithmic, showing errors can be significant. In this figure, $K_S = K = 10$.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

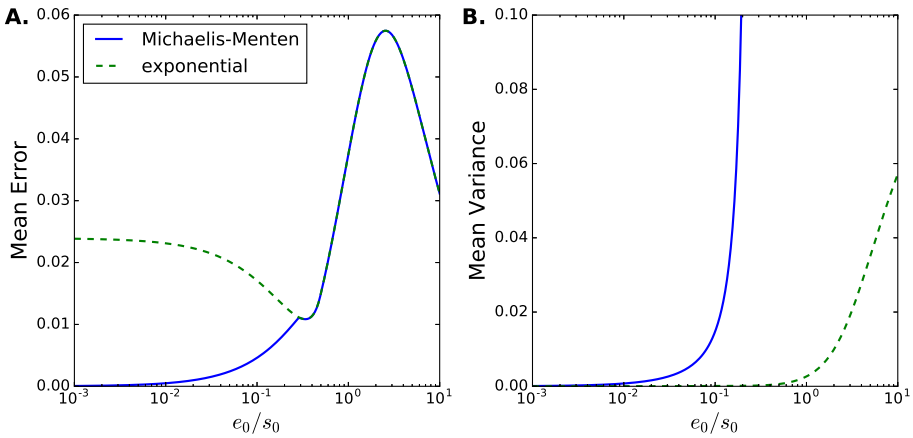

Figure 4: **(A) Mean concentration error**, and **(B) Mean variance in estimated parameters for the Michaelis–Menten equation and an exponential model.** For initial enzyme concentrations smaller than initial substrate concentrations, the Michaelis–Menten equation provides a noticeably better approximation of the true progress curve than does the exponential model, allowing for both V and K_M to be uniquely determined. Parameters for the case shown are: $(k_1, k_{-1}, k_{\text{cat}}) = (1.0, 0.5, 0.5)$, $t_{\text{obs}} = 3t_s$, $\omega = t_{\text{obs}}/1000$, $s_0 = 1$.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

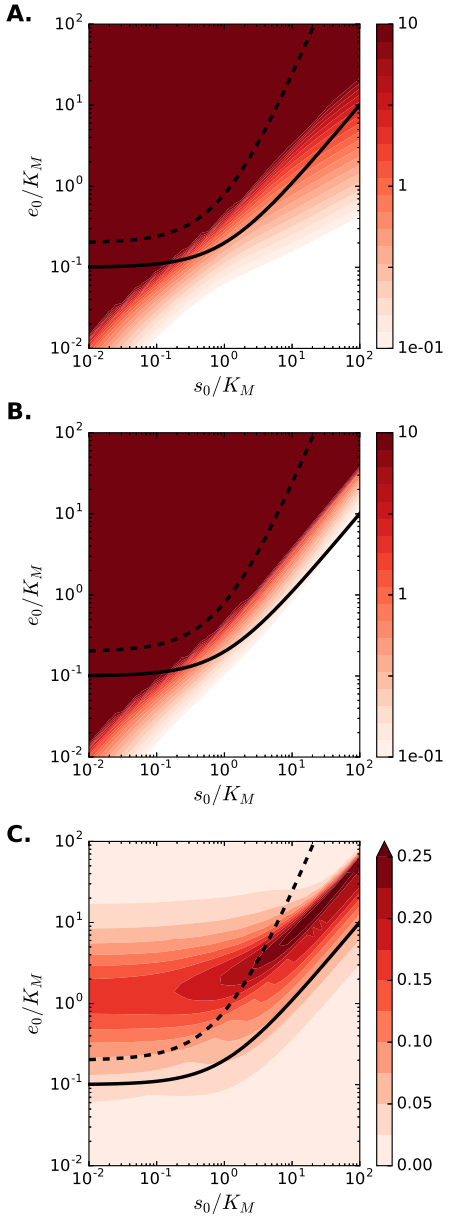

Figure 5: **Error contours when initial substrate concentration, s_0, is estimated from data.** K_M and V prediction errors (panels **A** and **B**, respectively) are qualitatively the same as those found when s_0 is known a priori. The error in estimating s_0 follows the reactant stationary condition, as well as providing accurate estimates when initial enzyme concentration is high and initial substrate concentration is low. Parameters for the case shown are: $(k_1, k_{-1}, k_{\text{cat}}) = (1.0, 0.5, 0.5)$, $t_{\text{obs}} = 3t_s$, $\omega = t_{\text{obs}}/100$.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

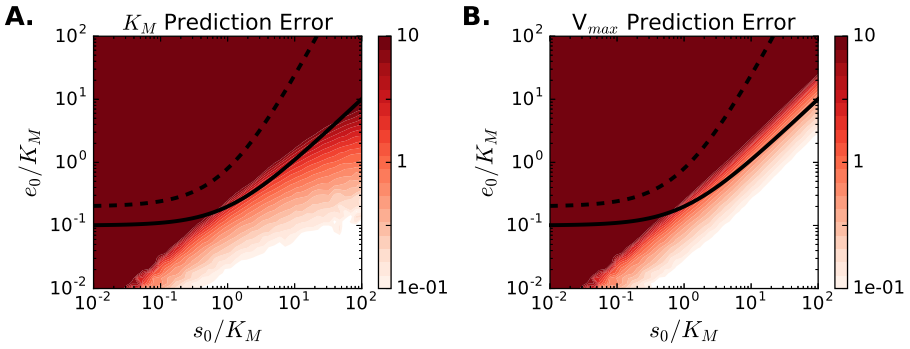

Figure 6: **Error contours for data with Gaussian noise.** When noise is added to the simulated data ($\delta = 0.01$), errors in the estimated parameters worsen compared to noise-free data. Parameters for the case shown are: $(k_1, k_{-1}, k_{\text{cat}}) = (1.0, 0.5, 0.5)$, $t_{\text{obs}} = 3t_s$, $\omega = t_{\text{obs}}/1000$.

bioRxiv preprint doi: https://doi.org/10.1101/068015; this version posted August 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

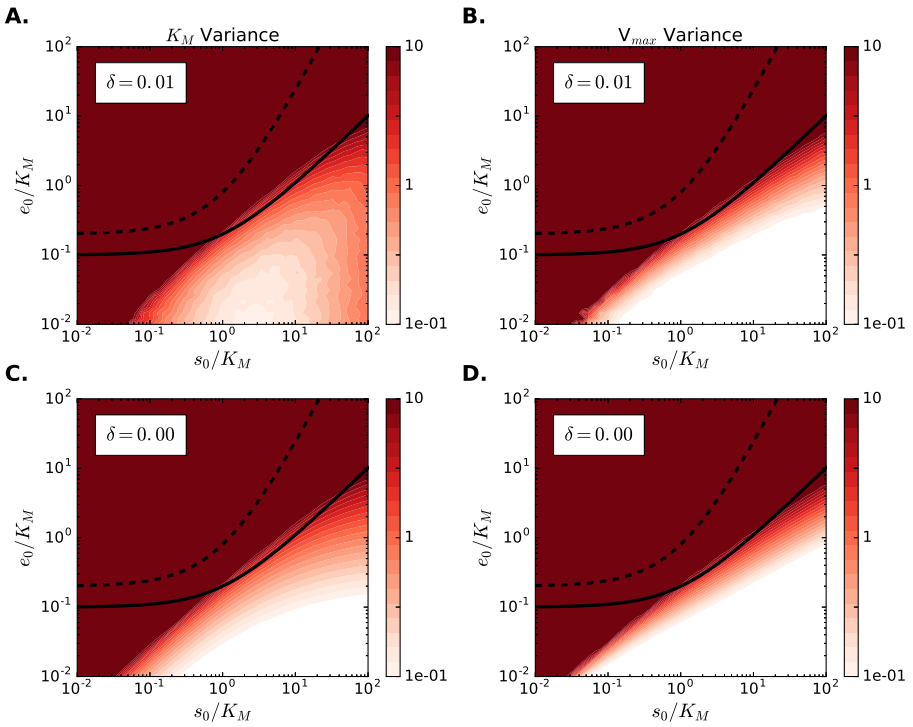

Figure 7: **Computed parameter variance for noisy and noise-free data.** Estimated variance in the parameters K_M (Panels **A** and **C**) and V (Panels **B** and **D**) for cases with $\delta = 0.01$ (**A** and **B**) and no noise (**C** and **D**). Even a small amount of noise restricts the range of conditions providing robust parameter estimates. Parameters for the case shown are: $(k_1, k_{-1}, k_{\text{cat}}) = (1.0, 0.5, 0.5)$, $t_{\text{obs}} = 3t_s$, $\omega = t_{\text{obs}}/1000$.