0.1 Introduction:

0.1.1 Identifiability analysis: Definitions and Formulations

The identifiability of parameters in nonlinear models of physical processes can be classified into two categories: structural and practical identifiability. The effect of model structure and parameterization on the ability to infer true parameter values from experimental data is determined by the structural identifiability of the parameter. The effect of the available experimental data on the ability to estimate unique parameter values is determined by the practical identifiability of the parameter. Practical identifiability of a parameter is contingent upon the nature, quality and quantity of data available to estimate the parameter as opposed to the structure and parameterization of the model. The aforementioned distinction between practical and structural identifiability of parameters is illustrated in Figure. Thus, the identifiability of parameters in nonlinear models is dependent on the model structure, parameterization, and the quality and quantity of experimental data that is available for the purpose of estimation.

On the one hand, establishing the structural identifiability of parameters enables one to propose models that are not only appropriate representations of physical processes, but also are parameterized in such a way that the value of these parameters can be estimated. On the other hand, establishing practical identifiability of parameters in any model helps design experiments that are minimal, informative and useful for parameter estimation.

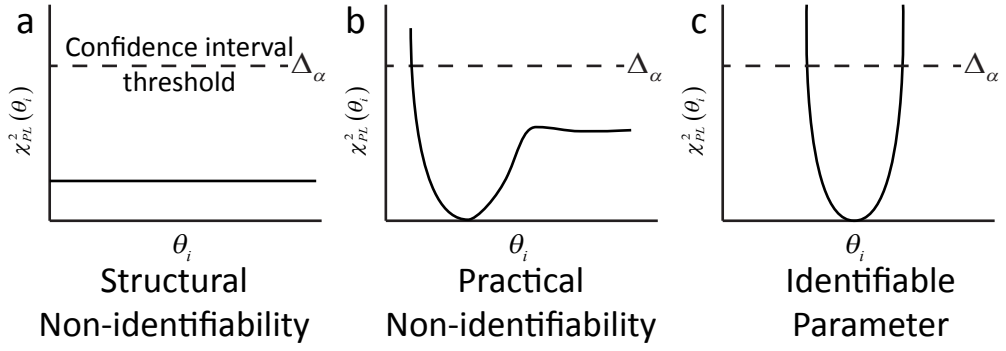

Figure 1: The profile likelihood estimates of a) a structurally non-identifiable, b) a practically non-identifiable and c) an identifiable parameter. The practically identifiability of a parameter is dependent on the quality and quantity of experimental data used to estimate the parameter.

We can use a profile likelihood-based approach to determine the structural and practical identifiability of parameters in any dynamic model. The identifiability of parameters is established through the confidence intervals of their estimates, $\left[\sigma_i^-, \sigma_i^+\right]$. The confidence intervals of a structurally non-identifiable parameter are unbounded, i.e., $[-\infty, +\infty]$, while the confidence intervals of a practically non-identifiable parameter are unbounded in at least one direction, i.e., $\left[\sigma_i^-, \sigma_i^+\right]$ where either $\sigma_i^- = -\infty$ or $\sigma_i^+ = +\infty$. If a parameter's estimates have a finite confidence interval then the parameter is said to be identifiable. The visualization of structurally and practically non-identifiable parameters using the profile likelihood approach is illustrated in Figure 1.

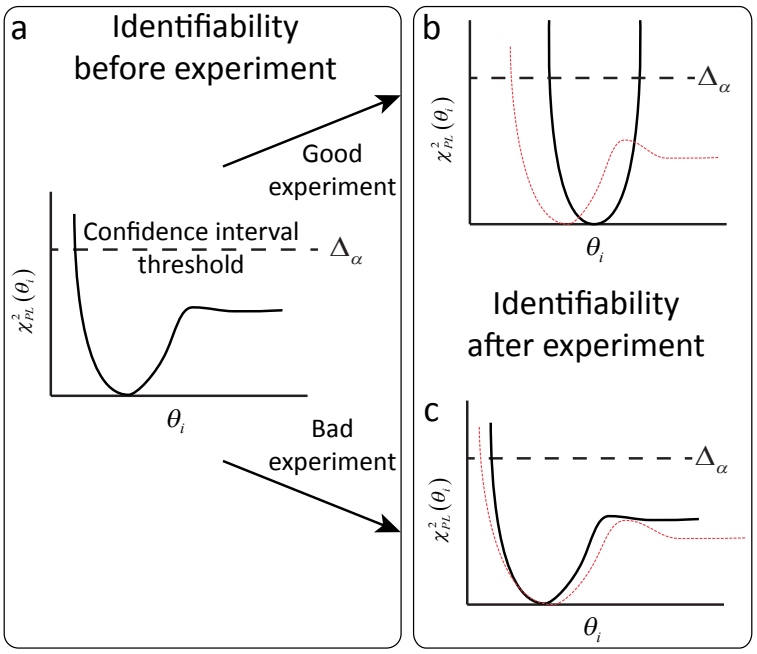

Figure 2: Cartoon illustrating the utility of identifiability analysis for experimental design. a) The profile likelihood of a practically non-identifiable parameter that needs to be estimated based on both good and bad experimental data. The changes in the profile likelihood of the parameter when estimated with b) good experimental data and c) bad experimental. The identifiability of the parameter does not change due to the poor quality/quantity of experiments.

Due to the dependence of practical parameter identifiability on the experimental data, the profile likelihood approach can be used to design experiments in such a way that the observables that are derived from these experiments can improve the practical identifiability of the parameters. We show how experimental

design can have a meaningful impact on parameter identification and estimation in Figure 2.

0.2 Methods:

We use a profile likelihood-based approach to establish structural and practical identifiability of parameters in nonlinear kinetic models of metabolism. Briefly, the approach seeks to establish the existence/non-existence of bounds in confidence intervals for the estimates of parameters in nonlinear models. The profile likelihood is calculated based on Equation 1 for each parameter θ_i where $\chi^2(\theta_i)$ is given by Equation 2.

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j\neq i}} \left[\chi^2(\theta)\right] \tag{1}$$

$$\chi^2(\theta) = \sum_{k=1}^{m}\sum_{l=1}^{d}\left(\frac{y^*_{kl} - y_{kl}}{\sigma^*_{kl}}\right)^2 \tag{2}$$

Equation 2 is typically used as the objective to be minimized in a typical parameter estimation problem where y^*_{kl} is the available experimental time course data for each observable state k at each l time point. The difference between this data and the model estimates at these time points, y_{kl} is weighted by the experimental error σ^*_{kl}. An algorithm to calculate $\chi^2_{PL}(\theta_i)$ based on Equation 1 is given below.

Algorithm:

1. Start by solving the numerical optimization problem in Equation 1 for a fixed value of θ_i to determine the initial point on the profile likelihood curve.
2. Take an increasing/decreasing step θ_{step} in the direction of θ_i.
3. Re-optimize $\theta_{j\neq i}$ using Equation 1.
4. Repeat the last two steps until a desired threshold Δ_α is exceeded or a maximal amount of steps is reached.

Adaptive step change:

The increasing/decreasing steps in θ_i can be adapted to the shape of the profile likelihood curve that is encountered while running the above algorithm. Accordingly, the step θ_{step} should satisfy Equation 3 below,

where θ^{k-1} refers to parameter estimates obtained from iteration $k-1$ of the above algorithm, and $q \in [0,1]$.

$$\chi^2\left(\theta^{k-1} + \theta_{step}\right) - \chi^2(\theta^{k-1}) \approx q.\Delta_\alpha \tag{3}$$

Identifiability analysis of parameters in a kinetic model of gluconeogenesis:

The proposed model for acetate consumption through gluconeogenesis and its corresponding kinetic model is used as a case study to illustrate the utility of identifiability analysis for the design of experiments for estimating parameters in kinetic models of metabolism. The kinetic model is described below.

$$\frac{d}{dt}pep = v_1 - v_2 - v_4 \tag{4}$$

$$\frac{d}{dt}fdp = v_2 - v_3 \tag{5}$$

$$\frac{d}{dt}E = v_{e,max}\left(\frac{1}{1+\left(\frac{fdp}{K_e^{fdp}}\right)^{n_e}}\right) - dE \tag{6}$$

The kinetic expressions for fluxes v_1 through v_4 are given below. The consumption of acetate through v_1 and conversion of *pep* through v_2 are expressed in Equations (7) and (8) respectively using Michaelis-Menten kinetics. The acetate flux through v_1 is also governed by the quantity of available enzyme E.

$$v_1 = k_1^{cat}E\frac{acetate}{acetate + K_1^{acetate}} \tag{7}$$

$$v_2 = V_2^{max}\frac{pep}{pep + K_2^{pep}} \tag{8}$$

$$v_3 = V_3^{max}\frac{f\tilde{d}p\left(1+f\tilde{d}p\right)^3}{\left(1+f\tilde{d}p\right)^4 + L_3\left(1+\frac{pep}{K_3^{pep}}\right)^{-4}} \tag{9}$$

The allosterically regulated flux v_3 for the consumption of *fdp* is expressed in Equation (9) using the Monod-Wyman-Changeux (MWC) model for allosterically regulated enzymes, where $f\tilde{d}p$ refers to the ratio of *fdp* with respect to its allosteric binding constant K_3^{fdp}. The added flux v_4 for the export of *pep* is expressed as a linear equation dependent on *pep* in Equation (10).

$$v_4 = k_4^{cat}.pep \tag{10}$$

Simulated data

0.2.1 Parameter estimation in kinetic models of metabolism:

The formulation for parameter estimation assumes that the concentrations and the fluxes corresponding to each perturbation are variables in the optimization problem formulated to estimate kinetic parameters.

Let $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{p} \in \mathbb{R}^l$ and $\mathbf{v} \in \mathbb{R}$ be the vector of concentrations, parameters and fluxes respectively. If the number of perturbations used for parameter estimation is in $\mathbb{R}^p$, then the vector of concentrations, fluxes and parameters that need to be estimated in the optimization problem changes to $\mathbf{x} \in \mathbb{R}^{mp}, \mathbf{v} \in \mathbb{R}^p$ and $\mathbf{p} \in \mathbb{R}^l$ respectively.

Let $i \in \mathbb{R}^{mp}$, $j \in \mathbb{R}^p$ and $k \in \mathbb{R}^l$ represent the indices for concentrations, fluxes and parameters respectively in Equations (11)-(13).

Formulation a):

$$\min_{\mathbf{x},\mathbf{p},\mathbf{v}} \ \|\mathbf{v} - \mathbf{v}^*\| \tag{11a}$$

$$\text{st } N(\mathbf{x},\mathbf{p}) - v_j D(\mathbf{x},\mathbf{p}) = 0 \qquad \forall\ j \in \mathbb{R}^p \tag{11b}$$

$$x_i^*(1-\epsilon) \le x_i \le x_i^*(1+\epsilon) \qquad \forall\ i \in \mathbb{R}^{mp} \tag{11c}$$

$$v_j^*(1-\epsilon) \le v_j \le v_j^*(1+\epsilon) \qquad \forall\ j \in \mathbb{R}^p \tag{11d}$$

$$\mathbf{p}_{min} \le \mathbf{p} \le \mathbf{p}_{max} \tag{11e}$$

Formulation b):

In this formulation, in Equation (12), the flux and concentrations are constrained as earlier with a noise term.

$$\min_{\mathbf{x},\mathbf{p},\mathbf{v}} \ w_1 \|\mathbf{v} - \mathbf{v}^*\| + w_2 \|\mathbf{x} - \mathbf{x}^*\| \tag{12a}$$

$$\text{st } N(\mathbf{x},\mathbf{p}) - v_j D(\mathbf{x},\mathbf{p}) = 0 \qquad \forall\ j \in \mathbb{R}^p \tag{12b}$$

$$x_i^*(1-\epsilon) \le x_i \le x_i^*(1+\epsilon) \qquad \forall\ i \in \mathbb{R}^{mp} \tag{12c}$$

$$v_j^*(1-\epsilon) \le v_j \le v_j^*(1+\epsilon) \qquad \forall\ j \in \mathbb{R}^p \tag{12d}$$

$$\mathbf{p}_{min} \le \mathbf{p} \le \mathbf{p}_{max} \tag{12e}$$

Formulation c):

A third formulation where the noise terms (ϵ) used as bounds for the flux and the concentrations are included in the optimization objective as a variable to be penalized is given in Equation (13) below. Notice that the noise term has been separated into two components, ϵ_x and ϵ_v, corresponding to noise in the concentrations and the fluxes, respectively. w_3 and w_4 are the weights associated with the noise terms in the objective function.

$$\min_{\mathbf{x},\mathbf{p},\mathbf{v},\epsilon_x,\epsilon_v} \; w_1 \left\|\mathbf{v}-\mathbf{v}^*\right\| + w_2 \left\|\mathbf{x}-\mathbf{x}^*\right\| + w_3\epsilon_x + w_4\epsilon_v \tag{13a}$$

$$\text{st } N(\mathbf{x},\mathbf{p}) - v_j D(\mathbf{x},\mathbf{p}) = 0 \qquad \forall\, j \in \mathbb{R}^p \tag{13b}$$

$$x_i^*(1-\epsilon_x) \le x_i \le x_i^*(1+\epsilon_x) \qquad \forall\, i \in \mathbb{R}^{mp} \tag{13c}$$

$$v_j^*(1-\epsilon_v) \le v_j \le v_j^*(1+\epsilon_v) \qquad \forall\, j \in \mathbb{R}^p \tag{13d}$$

$$\mathbf{p}_{min} \le \mathbf{p} \le \mathbf{p}_{max} \tag{13e}$$

Formulation d):

Due to the inability of Equation (11) to get feasible results, we modified the problem to include the L2-norm of the concentrations as well in the objective function in Equation (14a). In Equation (14) the flux and the concentrations are free variables, i.e., $\mathbf{x} \in [\mathbf{0}, +\infty]$.

$$\min_{\mathbf{x},\mathbf{p},\mathbf{v}} \; w_1 \left\|\mathbf{v}-\mathbf{v}^*\right\| + w_2 \left\|\mathbf{x}-\mathbf{x}^*\right\| \tag{14a}$$

$$\text{st } N(\mathbf{x},\mathbf{p}) - v_j D(\mathbf{x},\mathbf{p}) = 0 \qquad \forall\, j \in \mathbb{R}^p \tag{14b}$$

$$0 \le x_i \le +\infty \qquad \forall\, i \in \mathbb{R}^{mp} \tag{14c}$$

$$0 \le v_j \le +\infty \qquad \forall\, j \in \mathbb{R}^p \tag{14d}$$

$$\mathbf{p}_{min} \le \mathbf{p} \le \mathbf{p}_{max} \tag{14e}$$

After testing the problem in Equation (14) above for the acetate uptake flux v_1, we found that all solutions using IPOPT exceeded the maximum number of iterations (set at 5000) and never produced an optimal solution. Hence, we use a modified form of the bounds in Equation (14) above in Equation (12).

Table 1: Table showing the perturbed values of all fluxes used for parameter estimation.

Designation	Perturbed Fluxes	Perturbed Values
P1	v_1	2
P2	v_2	0.2
P3	v_3	0.5

0.3 Results:

Table 2: Table showing the ability of different formulations mentioned above to estimate enzyme kinetic parameters for different fluxes based on the perturbations given in the above table.

Flux	Data Type	Formulation (b)		Formulation (c)	
		$\epsilon_x(\%)$	$\epsilon_v(\%)$	$\epsilon_x(\%)$	$\epsilon_v(\%)$
flux 1	no noise	30	5	33.33	0
	noisy	x	x	0	33.33
flux 2	no noise	90	30	0	31
	noisy	85	40	0	30
flux 3	no noise	15	10	17	0
	noisy	x	x	17	0

0.3.1 Estimation of flux 1:

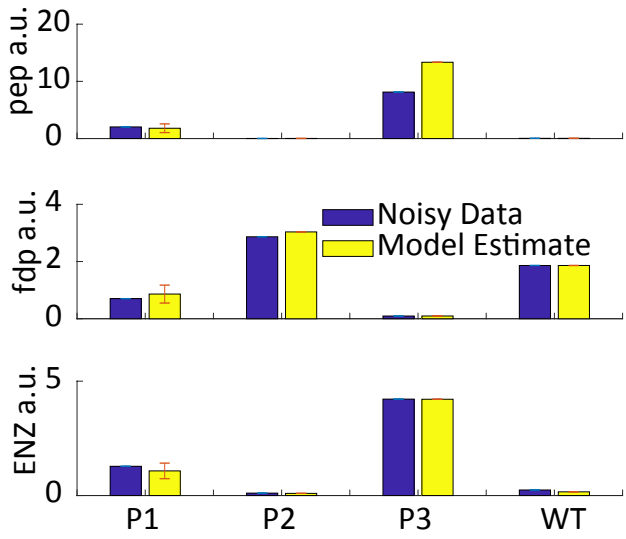

Figure 3: Plot comparing noiseless model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_1 using formulation (c).

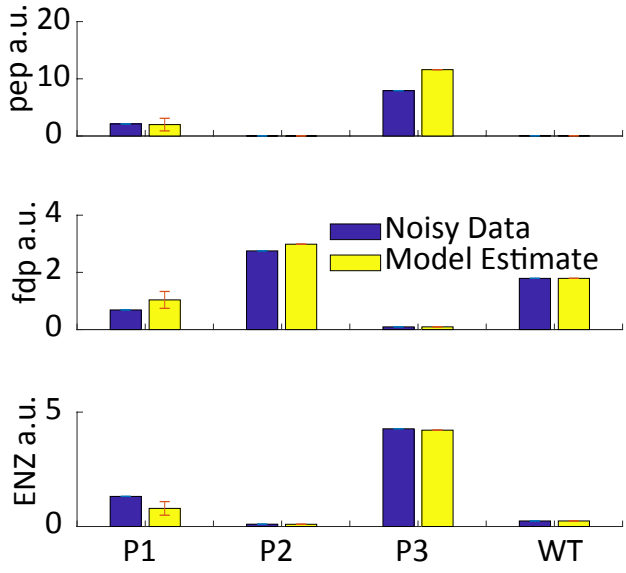

Figure 4: Plot comparing noisy model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_1 using formulation (c).

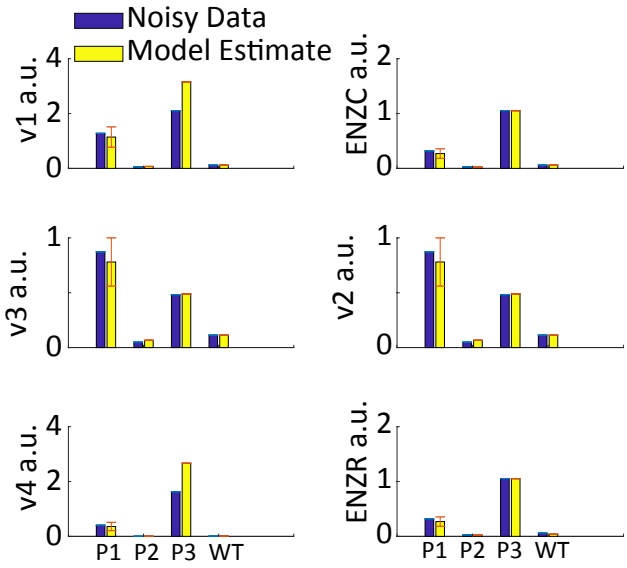

Figure 5: Plot comparing noiseless model generated steady state fluxes with model estimates obtained using an optimized model for v_1 using formulation (c).

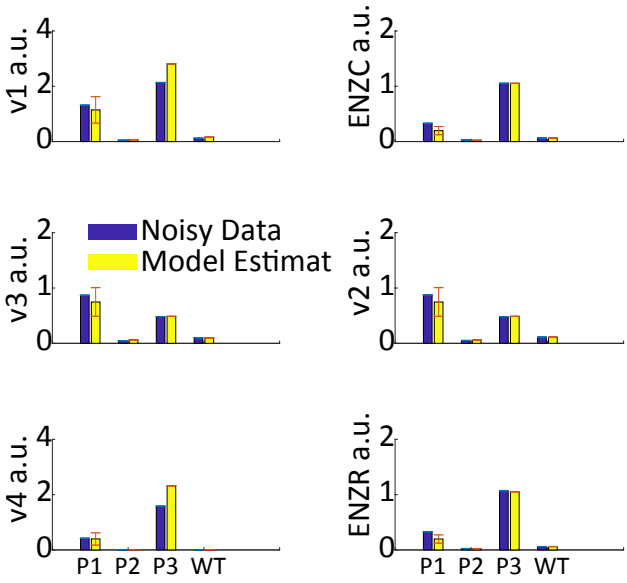

Figure 6: Plot comparing noisy model generated steady state fluxes with model estimates obtained using an optimized model for v_1 using formulation (c).

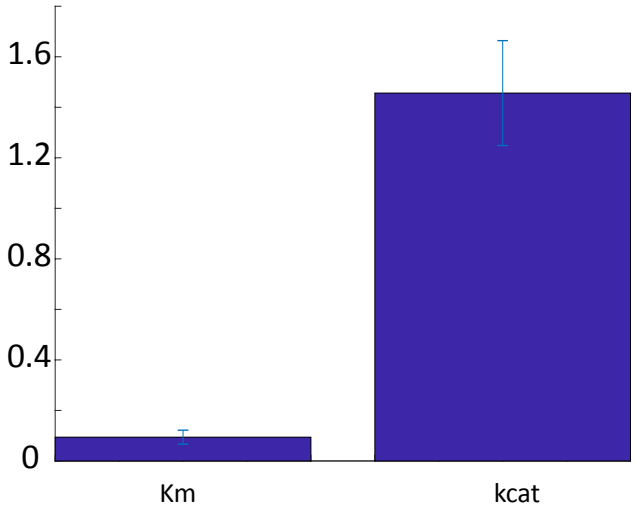

Figure 7: Plot comparing noiseless model generated parameters for v_1 through a multi-start search algorithm using formulation (c).

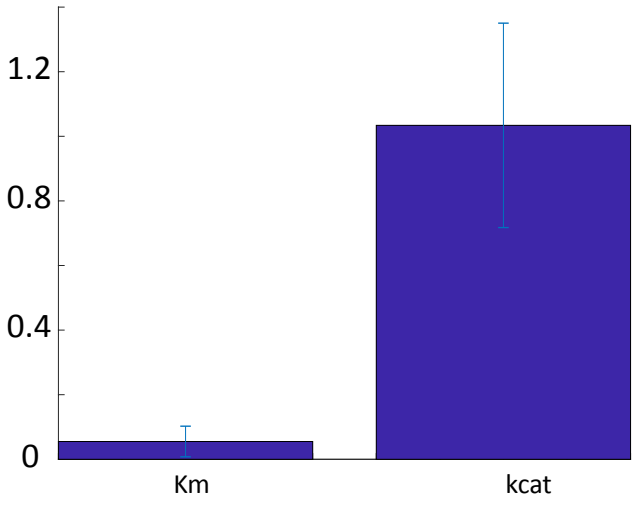

Figure 8: Plot comparing noisy model generated parameters for v_1 through a multi-start search algorithm using formulation (c).

0.3.2 Estimation of flux 2:

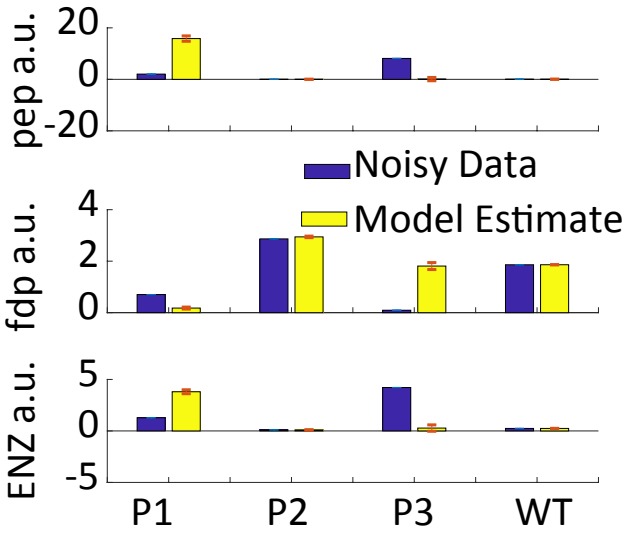

Figure 9: Plot comparing noiseless model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_2 using formulation (c).

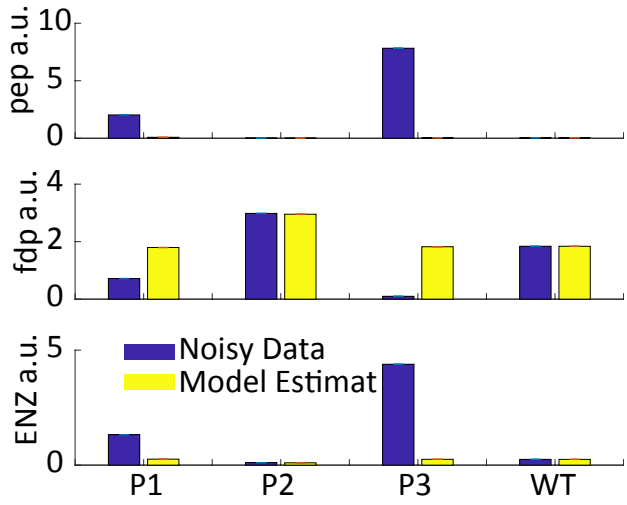

Figure 10: Plot comparing noisy model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_2 using formulation (c).

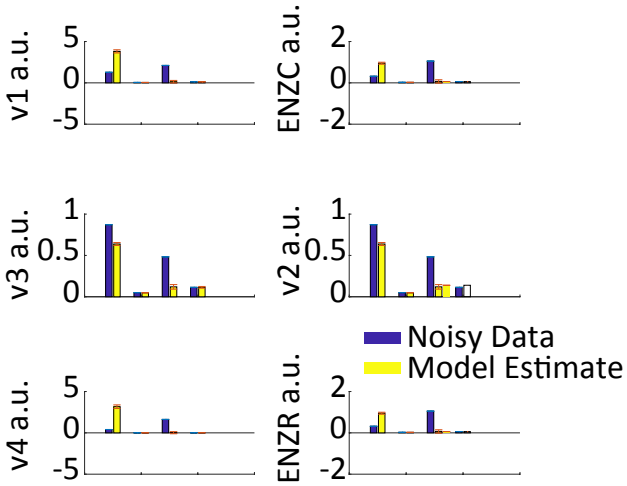

Figure 11: Plot comparing noiseless model generated steady state fluxes with model estimates obtained using an optimized model for v_2 using formulation (c).

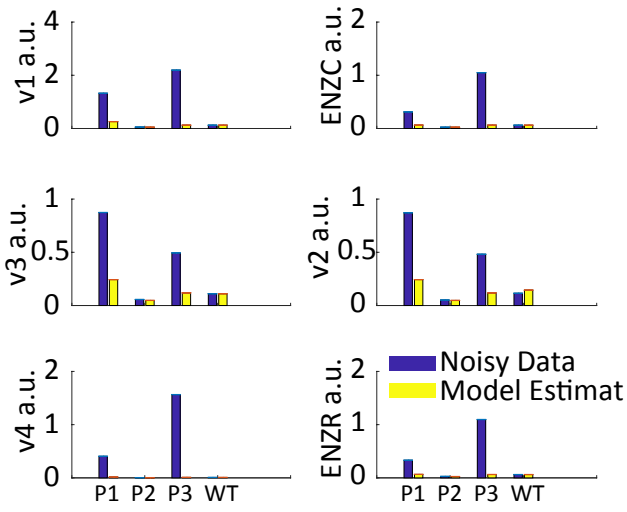

Figure 12: Plot comparing noisy model generated steady state fluxes with model estimates obtained using an optimized model for v_2 using formulation (c).

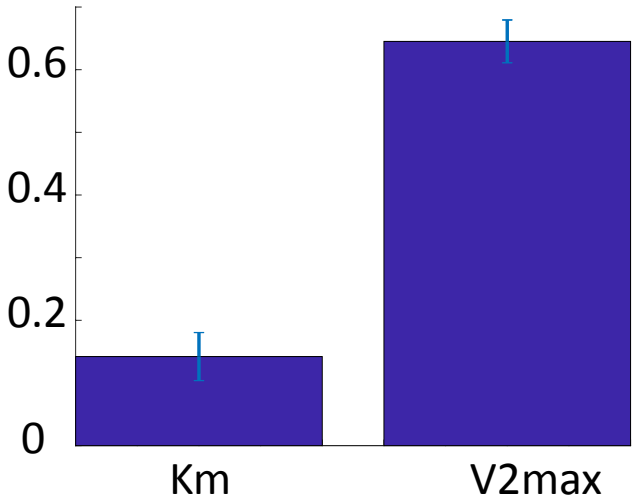

Figure 13: Plot comparing noiseless model generated parameters for v_2 through a multi-start search algorithm using formulation (c).

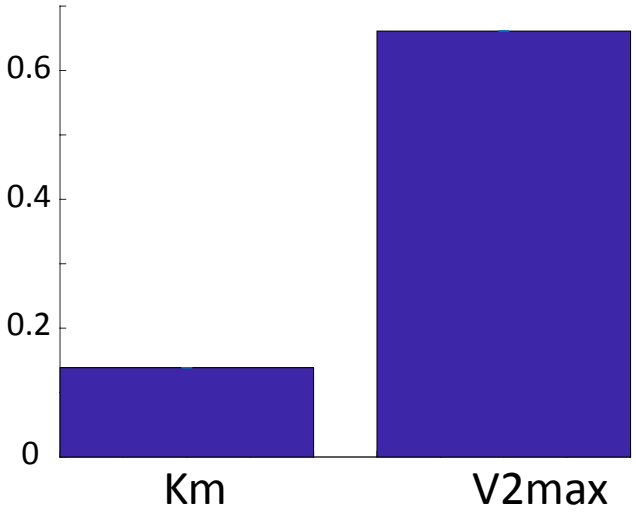

Figure 14: Plot comparing noisy model generated parameters for v_2 through a multi-start search algorithm using formulation (c).

0.3.3 Estimation of flux 3:

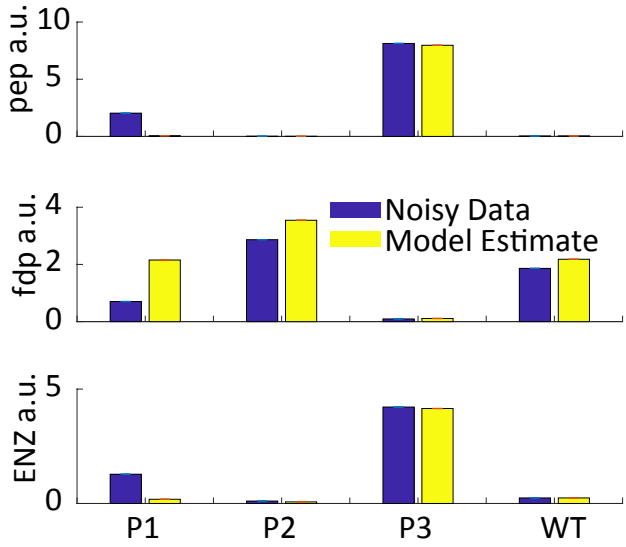

Figure 15: Plot comparing noiseless model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_3 using formulation (c).

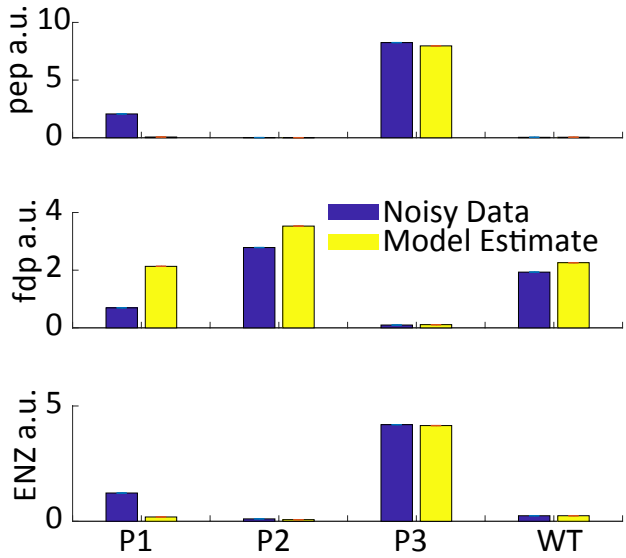

Figure 16: Plot comparing noisy model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_3 using formulation (c).

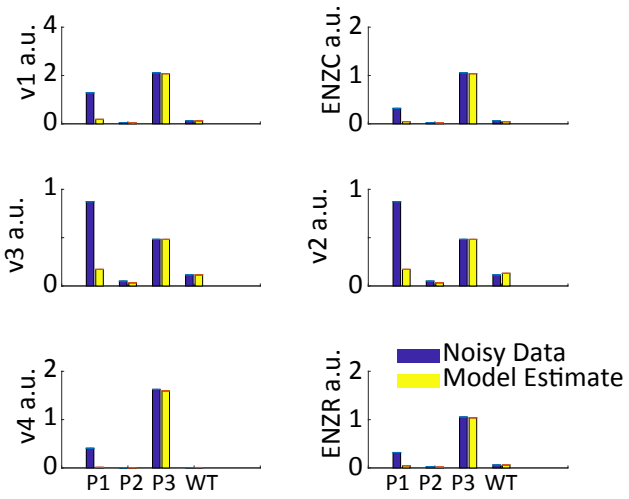

Figure 17: Plot comparing noiseless model generated steady state fluxes with model estimates obtained using an optimized model for v_3 using formulation (c).

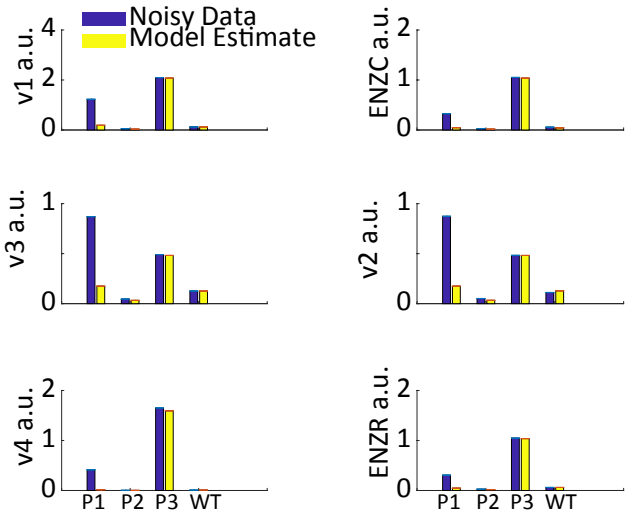

Figure 18: Plot comparing noisy model generated steady state fluxes with model estimates obtained using an optimized model for v_3 using formulation (c).

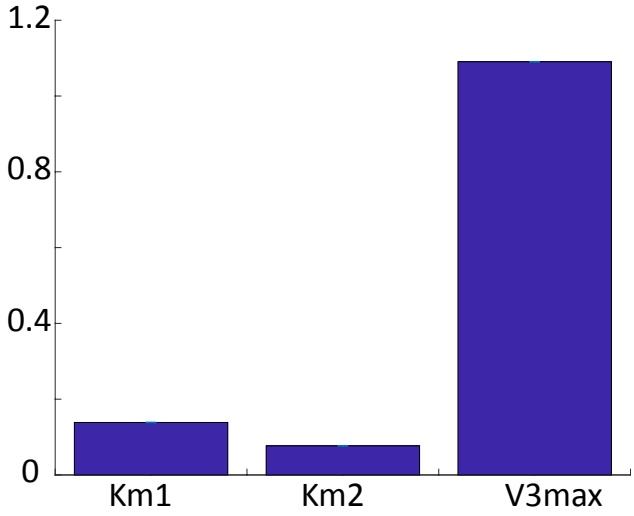

Figure 19: Plot comparing noiseless model generated parameters for v_3 through a multi-start search algorithm using formulation (c).

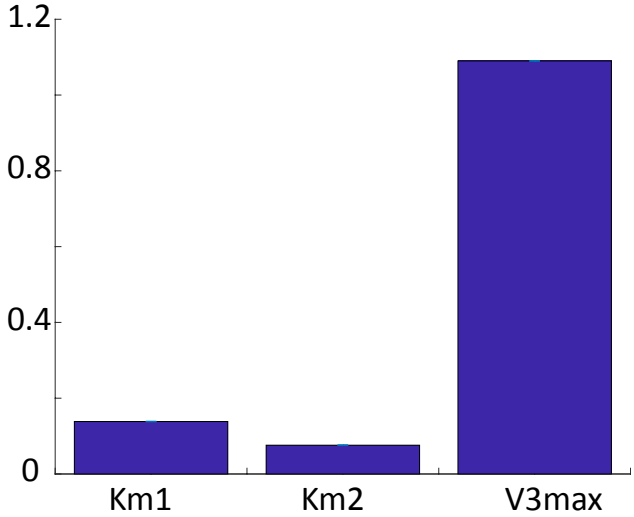

Figure 20: Plot comparing noisy model generated parameters for v_3 through a multi-start search algorithm using formulation (c).