Modeling and parameter subset selection for fibrin polymerization kinetics with applications to wound healing

Katherine J Pearce · Kimberly Nellenbach · Ralph C Smith · Ashley C Brown · Mansoor A Haider

Abstract During the hemostatic phase of wound healing, vascular injury leads to endothelial cell damage, initiation of a coagulation cascade involving platelets, and formation of a fibrin-rich clot. As this cascade culminates, activation of the protease thrombin occurs and soluble fibrinogen is converted into an insoluble polymerized fibrin network. Fibrin polymerization is critical for bleeding cessation and subsequent stages of wound healing. We develop a cooperative enzyme kinetics model for in vitro fibrin matrix polymerization capturing dynamic interactions among fibrinogen, thrombin, fibrin and intermediate complexes. A tailored parameter subset selection technique is also developed to evaluate parameter identifiability for a representative data curve for fibrin accumulation in a short duration in vitro polymerization experiment. Our approach is based on systematic analysis of eigenvalues and eigenvectors of the classical information matrix for simulations of accumulating fibrin matrix

Funding. This study was funded in part by grants DMS-1638521 and DMR-1847488 (CAREER) from the National Science Foundation and NHLBI R01HL146701 from the National Institutes of Health.

Katherine J. Pearce
Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA

Kimberly Nellenbach
Joint Department of Biomedical Engineering, North Carolina State University and The University of North Carolina at Chapel Hill, Raleigh, NC 27695, USA

Ralph C. Smith
Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA

Ashley C. Brown
Joint Department of Biomedical Engineering, North Carolina State University and The University of North Carolina at Chapel Hill, Raleigh, NC 27695, USA

Mansoor A. Haider (corresponding author)
Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA E-mail: mahaider@ncsu.edu, https://orcid.org/0000-0002-3096-1203

via optimization based on a least squares objective function. Results demonstrate robustness of our approach in that a significant reduction in objective function cost is achieved relative to a more ad hoc curve-fitting procedure. Capabilities of this approach to integrate non-overlapping subsets of the data to enhance the evaluation of parameter identifiability is also demonstrated. Unidentifiable reaction rate parameters are screened to determine whether individual reactions can be eliminated from the overall system while preserving the low objective cost. These findings demonstrate the high degree of information within a single fibrin accumulation curve, and a tailored model and parameter subset selection approach for improving optimization and reducing model complexity in the context of polymerization experiments.

Keywords Wound healing · Fibrin polymerization · Kinetics model · Parameter identifiability · Subset selection

Mathematics Subject Classification (2010) 62P10 · 65L09 · 92C45 · 92C05

1 Introduction

Wound healing is a complex process that occurs in four temporally overlapping phases (hemostasis, inflammation, proliferation, and remodeling), spanning several hours to weeks (Chester et al. 2019). During hemostasis, initial vascular injury leads to endothelial cell damage, platelet activation, and subsequent fibrin clot formation. Over the next 24-48 hours, inflammatory cells migrate into the wound bed to fight pathogens and remove foreign material. This is followed by 1-2 weeks of proliferation in which fibroblasts migrate into the fibrin clot to rebuild damaged tissue via biosynthesis and remodeling of extracellular matrix (ECM). Lastly, in the remodeling phase which spans months to years, provisional ECM is replaced with a more permanent collagen-rich matrix concurrent with additional fibroblast-mediated ECM remodeling. The successful design of novel therapeutic approaches can benefit from coordination with development of mathematical models for these underlying mechanisms, particularly in cases of degenerate wound healing (Hoffman et al. 2006). A variety of mathematical models have been developed to describe these mechanisms in one or more phases of wound healing (see Jorgensen and Sanders 2016; Valero et al. 2015; Weihs et al. 2016 for reviews). Most models focus on one (or just a few) phases of wound healing (Jorgensen and Sanders 2016). Models have been delineated (Valero et al. 2015) by those that are biochemical (kinetics or reaction-diffusion systems), mechanobiological (Tranquillo and Murray 1992), or discrete (e.g. Cellular Potts) or particle-based models that track individual cells (Vermolen and Gefen 2013). Recently, hybrid (continuum-discrete) models have also been developed to study coupling between cell dynamics, diffusive signaling and extracelullar matrix deposition (Wang et al. 2019).

This investigation considers hemostasis, during which the coagulation cascade is activated by the initial vascular injury and formation of a fibrin-rich

clot ensues. Specifically, this cascade culminates in activation of the protease (enzyme) thrombin, which converts soluble fibrinogen into an insoluble fibrin network (Janmey et al. 2009; Weisel 2004; Weisel 2005; Weisel and Litvinov 2013; Wolberg 2007). Formation of this network is critical, both initially as it stops bleeding, and for subsequent stages of wound healing. In particular, the initial polymerized fibrin network serves as a provisional ECM scaffold to support infiltration of cells involved in later healing stages (Chester and Brown 2017), thus influencing outcomes in both normal and degenerate wound healing. The viability of kinetic modeling of fibrin polymerization was demonstrated several decades ago (Weisel and Nagaswami 1992), but model interpretation was qualitative. In particular, model parameters were not estimated though a process of inverse analysis of data from experiments.

The focus of this study is two-fold. The first major focus is on the development of a kinetics-based mathematical model for fibrin matrix polymerization during initial clot formation. A cooperative enzyme kinetics model, comprising a system of ordinary differential equations (ODEs), is formulated to capture dynamic interactions among fibrinogen, thrombin, fibrin matrix and intermediate complexes. The second major focus is the development of a tailored parameter subset selection technique to assess those parameters which are identifiable in the sense that they are uniquely determined by data. The notion of identifiability specific to our problem is practical identifiability with respect to the underlying reaction rates (Brun et al. 2002; Quaiser and Mönnigmann 2009). Our approach is rooted in algorithms presented in studies based on analysis of eigenvalues and eigenvectors of the (ill-conditioned) information matrix (Burth et al. 1999; Quaiser and Mönnigmann 2009; Vajda et al. 1989). Prior works utilizing parameter subset selection techniques based on such approaches employed simulated data (Burth et al. 1999; Cintrón-Arias et al. 2009; Quaiser and Mönnigmann 2009), whereas our approach is tailored to our specific model and wound healing application using real data for fibrin matrix accumulation during polymerization. Note that our approach does not integrate the parameter estimation and subset selection procedure with uncertainty quantification of estimated parameter values, as in Cintrón-Arias et al. (2009). The technique developed in this study is applied in the context of parameter estimation via optimization of a least squares objective function comparing model predictions of fibrin to representative data for fibrin matrix accumulation in a short duration polymerization experiment.

2 Models and Methods

2.1 Fibrin Matrix Polymerization Experiment

We have the capability to analyze the polymerization of fibrin matrices via in vitro clot turbidity experiments (Sproul et al. 2018). In this study, we focus on the representative fibrin matrix data corresponding to measurements of polymerizaton rate (turbidity profiles) for initial concentrations of 2.5 mg/mL

for fibrinogen, 0.75 (Units/mL) for thrombin and 5mM CaCl2. Fibrinogen was mixed with 25 mM Hepes buffer (25 mM HEPES, 150 mM NaCl, 5 mM CaCl2, pH 7.4). Real time measurement of clot turbidity was taken every minute for roughly 4 hours (225 minutes) through absorbance readings (Abs350nm) using a plate reader (Brown et al. 2015; Sproul et al. 2018). Such measurements are known to have a direct relationship with proportional increases in fibrin matrix density (Wolberg et al. 2002). Specifically, clots are formed within a 96 well plate and the plate reader measures absorbance at 9 locations per well. As fibrinogen converts into fibrin, an increase in absorbance is detected, serving as a quantitative measurement of fibrin matrix density (Fig. 1). This method enables rapid investigation of early time system dynamics which is not possible using traditional approaches to measuring fibrin content.

2.2 Reaction System for Fibrin Matrix Polymerization

We idealize the aforementioned in vitro system for fibrin matrix polymerization as a set of chemical reactions involving interactions among fibrinogen, the enzyme *thrombin* (Thb) and *fibrin matrix* (FM). A unique feature of this system is that thrombin and fibrinogen are unbound (dissolved) species, whereas fibrin matrix is a bound species that forms a solid-like matrix. We will utilize an extended enzyme kinetics framework to account for the intermediate complexes that arise in the transition from unbound to bound system constituents which is known to involve a cooperative mechanism (Chernysh et al. 2012). Hence, our system will include a small number of intermediate complexes along with associated reactions.

The first stage of fibrin polymerization in the in vitro system involves the activation of *inactive fibrinogen* (Fbn_i) into *active fibrinogen* (Fbn_a), with the latter species combining with other substrates in subsequent reactions to form fibrin matrix. This activation of fibrinogen is assumed to follow Michaelis-Menten-type reaction kinetics (Michaelis and Menten 1913), mediated by interactions with the enzyme thrombin (Thb) and a single intermediate complex (C_0). The resulting reaction system is:

$$\mathrm{Fbn}_i + \mathrm{Thb} \underset{k^-}{\overset{k^+}{\rightleftharpoons}} \mathrm{C}_0 \xrightarrow{k} \mathrm{Fbn}_a + \mathrm{Thb}. \tag{1}$$

The right-most (forward) reaction in (1) occurs rapidly and is assumed to be irreversible once fibrinogen is activated. Active fibrinogen is a dissolved intermediate product that in turn reacts to form fibrin matrix; i.e. active fibrinogen does not return to an inactive form.

Thrombin-mediated kinetics for fibrin assembly are known to exhibit a sigmoidal reaction velocity for the product (Chatterjee et al. 2010; Weisel and Nagaswami 1992), indicative of cooperative enzyme effects with a few intermediate complexes. Here, we assume the simplest case of a Hill-type model (Goutelle et al. 2008; Hoffman et al. 2006) with positive cooperativity (Hill coefficient $n = 2$) to model fibrin matrix accumulation. The associated reaction

system thus involves two additional intermediate complexes (C_1, C_2) and is given by:

$$\mathrm{Fbn}_a + \mathrm{Thb} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{C}_1 \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{Thb} + \mathrm{FM}, \tag{2}$$

$$\mathrm{Fbn}_a + \mathrm{C}_1 \underset{k_3^-}{\overset{k_3^+}{\rightleftharpoons}} \mathrm{C}_2 \underset{k_4^-}{\overset{k_4^+}{\rightleftharpoons}} \mathrm{C}_1 + \mathrm{FM}. \tag{3}$$

The reaction system for our model then comprises (1)-(3). Note that the 11 reaction rates $\{k, k^+, k^-, k_i^+, k_i^- \mid i = 1, \ldots, 4\}$ are all assumed to be non-negative constants. Some additional features of the reaction system (1)-(3) are now outlined.

In many enzyme kinetics models with cooperative effects, it is assumed that the product is continuously removed from the system (Keener and Sneyd 1998). When this assumption is applicable, the right-most reverse reactions in (2) and (3) can be neglected. By contrast, in our system the product (fibrin matrix) is a bound species that cannot be removed from the system; thus, in our model we retain the reverse reactions in (2) and (3). Moreover, due to inherent model complexity from cooperative effects, many models invoke a quasi-steady state approximation (QSSA; Briggs and Haldane 1925), assuming equal rates for formation and breakdown of intermediate complexes (Keener and Sneyd 1998). In our system (1)-(3), using the QSSA violates physical laws governing the system, admitting negative values for substrate and product concentrations in the model.

2.3 Mathematical Model

Initial ODE System. Using the law of mass action, which assumes leading order reaction kinetics (Holmes 2000), we map the 11 reactions in (1)-(3) to a system of 7 ordinary differential equations in 7 dependent variables, delineated as the four primary system species ($\mathrm{Fbn}_a, \mathrm{FM}, \mathrm{Thb}, \mathrm{Fbn}_i$) and three intermediate complexes ($\mathrm{C}_0, \mathrm{C}_1, \mathrm{C}_2$). In the mathematical model italicized symbols for all species denote their concentration, except for *FM* which uses absorbance

units (as a surrogate for concentration). The resulting system is

$$\frac{dFbn_a}{dt} = -k_1^+ Fbn_a \cdot Thb + k_1^- C_1 - k_3^+ Fbn_a \cdot C_1 + k_3^- C_2 + kC_0 \tag{4}$$

$$\frac{dFM}{dt} = k_2^+ C_1 - k_2^- Thb \cdot FM + k_4^+ C_2 - k_4^- C_1 \cdot FM, \tag{5}$$

$$\begin{aligned}\frac{dThb}{dt} &= -k_1^+ Fbn_a \cdot Thb + k_1^- C_1 + k_2^+ C_1 - k_2^- Thb \cdot FM \\ &\quad - k^+ Fbn_i \cdot Thb + k^- C_0 + kC_0,\end{aligned} \tag{6}$$

$$\frac{dFbn_i}{dt} = -k^+ Fbn_i \cdot Thb + k^- C_0, \tag{7}$$

$$\frac{dC_0}{dt} = k^+ Fbn_i \cdot Thb - k^- C_0 - kC_0, \tag{8}$$

$$\begin{aligned}\frac{dC_1}{dt} &= k_1^+ Fbn_a \cdot Thb - k_1^- C_1 - k_2^+ C_1 + k_2^- Thb \cdot FM + k_3^- C_2 \\ &\quad - k_3^+ Fbn_a \cdot C_1 + k_4^+ C_2 - k_4^- C_1 \cdot FM,\end{aligned} \tag{9}$$

$$\frac{dC_2}{dt} = k_3^+ Fbn_a \cdot C_1 - k_3^- C_2 + k_4^- C_1 \cdot FM - k_4^+ C_2. \tag{10}$$

Initial conditions for this system are as follows:

$$\begin{aligned}&Fbn_a(0) = 0,\ FM(0) = FM_0,\ Thb(0) = Thb_0,\ Fbn_i(0) = Fbn_{i0}, \\ &C_j(0) = 0,\ j = 0, 1, 2.\end{aligned} \tag{11}$$

Collectively, equations (4)-(11) contain 14 model parameters, of which 11 are (constant) reaction rates and the remaining 3 are prescribed initial species quantities. In (11), Thb_0 is the initial concentration of thrombin in the system based on properties of the batch obtained for use in the experiment (§2.1). We also assume that initially all fibrinogen is inactive, with prescribed initial value Fbn_{i0}. Due to the structure of our coupled reaction system, a small initial amount of fibrin matrix must be assumed at $t = 0$. This value for FM_0 is determined in the experiment by measuring the amount of fibrin matrix present just after initial polymerization occurs and is used to prescribe the corresponding initial condition in (11).

Conservation Laws. The stoichiometric matrix S used to determine conservation laws for the system of ODEs (4)-(10) is defined in terms of the net stoichiometric coefficient for each reactant and product involved in each reaction, assuming r species and n reactions ($S \in \mathbb{R}^{n \times r}$). Specifically, the net stoichiometric coefficient of species i in reaction j is $n_{ij} = n_{ij}^+ - n_{ij}^-$, where n_{ij}^+ is the product stoichiometric coefficient defined as the production rate for the species, and n_{ij}^- the reactant stoichiometric coefficient defined as the consumption rate of the species (Brendel et al. 2006). With these definitions, the (i, j)-th entry $S_{ij} = n_{ij}$. The rank of the null space of S^T is equal to the number of linearly independent conservation laws in the system of ODEs, and a specific set of conservation laws is obtained from basis vectors for the null

space (Holmes 2000). In our system, $n = 7$ and $r = 11$, and the matrix S yields two independent conservation laws (noting that these depend on the chosen basis vectors):

$$-Fbn_a - FM + 2Thb - Fbn_i + C_0 + C_1 = -Fbn_{i0} + 2Thb_0 - FM_0, \quad (12)$$

$$Fbn_a + FM - Thb + Fbn_i + C_2 = Fbn_{i0} - Thb_0 + FM_0. \quad (13)$$

Reduced ODE System. Given two independent conservation laws, it is possible to reduce our system from 7 to 5 ordinary differential equations. We use (12) and (13) to eliminate the intermediate complexes C_1 and C_2 and thus reduce the original system to 5 ordinary differential equations (4)-(8) with the respective initial conditions.

2.4 Data Observations and Statistical Model

The parameter subset selection techniques employed in this investigation were developed by optimizing model predictions of $FM(t)$ using a single representative experimental data curve for accumulation of fibrin matrix (FM) (Fig. 1, see also §2.1).

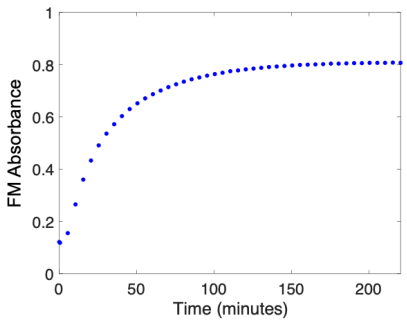

Fig. 1: Representative fibrin matrix absorbance data for model development and analysis. As outlined in §2.1, data was collected every minute for a total of 225 minutes and averaged at each time point across all 9 locations within one well. The data was downsampled to retain every 10^{th} point.

The representative data curve exhibits sigmoidal features consistent with a Hill coefficient ($n > 1$) associated with positive cooperative (enzyme) effects. In particular, this curve exhibits an inflection point, corresponding to a maximum reaction rate, occurring within the first hour of the reaction. To increase algorithmic efficiency, we have downsampled the data by considering every k^{th} point. We denote the data collected in this manner by $\{y_i^D\}_{i=1}^N$, where N is the total number of data points after downsampling.

Let $y(t; \mathbf{q_0}, \mathbf{q_k})$ be the solution curve for the fibrin matrix equation in (4)-(8), where

$$\mathbf{q_0} = [FM_0, Thb_0, Fbn_{i0}] \tag{14}$$

are the three (experimentally prescribed) initial condition parameters, and

$$\mathbf{q_k} = [k_1^+, k_1^-, k_2^+, k_2^-, k_3^+, k_3^-, k_4^+, k_4^-, k, k^+, k^-] \tag{15}$$

are the 11 reaction rate parameters. The statistical model for the observation process can be formulated as

$$y_i^D = y(t_i; \tilde{\mathbf{q}}_\mathbf{0}, \tilde{\mathbf{q}}_\mathbf{k}) + \epsilon_i,\ i = 1, \ldots, N, \tag{16}$$

where $\tilde{\mathbf{q}}_\mathbf{0}, \tilde{\mathbf{q}}_\mathbf{k}$ are the true but unknown parameter values. The errors ϵ_i are random variables assumed to be independently and identically distributed.

2.5 Parameter Subset Selection Approach

In practice, we prescribe values for $\mathbf{q_0} \approx \tilde{\mathbf{q}}_\mathbf{0}$ from (16) based on initial states in the fibrin polymerization experiment. Hence, our focus is on estimating parameters $\mathbf{q_k} \in \mathbb{R}^M$ from (15) for $M = 11$. To find an optimal $\mathbf{q}_\mathbf{k}^* \approx \tilde{\mathbf{q}}_\mathbf{k}$ in (16) we use an approach that, within each iteration, systematically decreases the number of parameters j ($\leq M$) to be estimated as we determine which $M - j$ parameters are unidentifiable in the sense that they are not uniquely determined by the data. In §2.5.4, we outline the main parameter subset selection procedure used to find $\mathbf{q}_\mathbf{k}^*$, but first we introduce some helpful concepts and notation.

Let $\mathbf{q} \in \mathbb{R}^j$ be the vector consisting of the current j parameters to be estimated, and let $\bar{\mathbf{q}} := \mathbf{q_k} \backslash \mathbf{q}$ be the vector comprising the remaining $M - j$ fixed parameters in $\mathbf{q_k}$. Thus, we may write $\mathbf{q_k} = \langle \mathbf{q}, \bar{\mathbf{q}} \rangle$[1], where $\langle\ ,\ \rangle$ denotes the merging of vectors that preserves the original parameter index ordering in (15). For convenience, the amount of fibrin matrix at time t is henceforth denoted as $y(t; \mathbf{q_0}, \langle \mathbf{q}, \bar{\mathbf{q}} \rangle)$.

2.5.1 Parameter Estimation

The parameters $\mathbf{q} \in \mathbb{R}^j$ are estimated via the following minimization of the ordinary least squares objective function $J(\mathbf{q}; \bar{\mathbf{q}})$:

$$\mathbf{q}^* = \arg \min_{\mathbf{q} \in \mathbb{R}^j_{\geq 0}} J(\mathbf{q}; \bar{\mathbf{q}}),\ \text{where } J(\mathbf{q}; \bar{\mathbf{q}}) = \frac{1}{N} \sum_{i=1}^{N} \left(y(t_i; \mathbf{q_0}, \langle \mathbf{q}, \bar{\mathbf{q}} \rangle) - y_i^D \right)^2. \tag{17}$$

[1] For example, if $\mathbf{q} = [k_4^+, k_2^-, k_1^+, k, k_2^+]$ and $\bar{\mathbf{q}} = [k_3^-, k_3^+, k_1^-, k^-, k_4^-, k^+]$, then $\mathbf{q_k} = \langle \mathbf{q}, \bar{\mathbf{q}} \rangle = [\mathbf{q}[3], \bar{\mathbf{q}}[3], \mathbf{q}[5], \mathbf{q}[2], \bar{\mathbf{q}}[2], \bar{\mathbf{q}}[1], \mathbf{q}[1], \bar{\mathbf{q}}[5], \mathbf{q}[4], \bar{\mathbf{q}}[6], \bar{\mathbf{q}}[4]]$. Note that this notation throughout the paper does not denote an inner product.

In evaluating $J(\mathbf{q};\bar{\mathbf{q}})$ in (17), the solution y is determined approximately using a numerical solver for stiff systems of ODEs (see §3). The optimization problem that determines $\mathbf{q}^*$ is solved via a Nelder-Mead direct search simplex algorithm (Nelder and Mead 1965) applied to the objective function (17).

2.5.2 Sensitivities and Information Matrix.

Once we obtain an optimal parameter vector $\mathbf{q}^*$, we compute the $j \times j$ information matrix $\chi^T\chi$ (Rothenberg 1971), where $\chi := \nabla_{\mathbf{q}} y(\{t_i\}_{i=1}^N; \mathbf{q_0}, \langle \mathbf{q}^*, \bar{\mathbf{q}}\rangle)$ is the $N \times j$ sensitivity matrix. However, the variety of units and time scales inherent to our model lead to a wide range of parameter (reaction rate) magnitudes to be estimated. To circumvent this, we re-scale our model by writing. $y_s(t;\mathbf{q_0},\langle\theta,\bar{\mathbf{q}}\rangle) = y(t;\mathbf{q_0},\langle\mathbf{s}(\theta),\bar{\mathbf{q}}\rangle)$, where θ is the scaled parameter vector and $\mathbf{s}(\theta)$ is the linear mapping from the intervals $[0,1]^j$ to $\prod_{k=1}^j[(1-\alpha)q_k^*,(1+\alpha)q_k^*]$. The value of α is prescribed ($0<\alpha<1$; see §3.1). The scaled sensitivity matrix χ_s is then defined as

$$\chi_s := \nabla_\theta y_s(\{t_i\}_{i=1}^N;\mathbf{q_0},\langle\theta,\bar{\mathbf{q}}\rangle) \tag{18}$$

where, via the Chain Rule, $\frac{\partial y_s}{\partial \theta_k} = \frac{\partial y}{\partial q_k^*}\frac{\partial q_k^*}{\partial \theta_k}$. It can be readily shown that $\frac{\partial q_k^*}{\partial \theta_k} = 2\alpha q_k^*$. The derivative $\frac{\partial y}{\partial q_k^*}$ is approximated using a complex-step approximation (Lyness 1967; Lyness and Moler 1967; Martins et al 2003; Squire and Trapp 1995)

$$\frac{\partial y}{\partial q_k^*}(t_l;\mathbf{q_0},\langle\mathbf{q}^*,\bar{\mathbf{q}}\rangle) \approx \frac{\mathrm{Im}[y(t_l;\mathbf{q_0},\langle\mathbf{q}^*+ih\mathbf{e}_k,\bar{\mathbf{q}}\rangle)]}{h},\ \ k=1,\ldots,j;\ l=1,\ldots,N, \tag{19}$$

where $\mathrm{Im}[\cdot]$ denotes the imaginary part of the function with standard basis vectors $\mathbf{e}_k \in \mathbb{R}^j$ and $0 < h(=10^{-10}) \ll 1$. Complex-step derivative approximation is second-order accurate without the drawback of (numerator) subtractive cancellation that can occur with standard finite differences (Banks et al. 2015).

We now briefly summarize the relevance of the matrix $\chi^T\chi$ (or analogously, $\chi_s^T\chi_s$) to parameter identifiablilty. When $y(t;\mathbf{q_0},\langle\mathbf{q},\bar{\mathbf{q}}\rangle)$ is expanded about $\mathbf{q}=\mathbf{q}^*$, we obtain via a Taylor expansion the result

$$J(\mathbf{q}^*+\Delta\mathbf{q};\bar{\mathbf{q}}) \approx \frac{1}{N}\Delta\mathbf{q}^T\chi^T\chi\Delta\mathbf{q} \approx \frac{\lambda}{N}||\Delta\mathbf{q}||_2^2. \tag{20}$$

Here $\Delta\mathbf{q} = \mathbf{q}-\mathbf{q}^*$ and χ is evaluated approximately using (19). We note that $\chi^T\chi$ is symmetric and non-negative definite. The second approximate equality in (20) follows when $\Delta\mathbf{q}$ is an eigenvector of $\chi^T\chi$; i.e. $\chi^T\chi\Delta\mathbf{q} = \lambda\Delta\mathbf{q}$ for some real eigenvalue $\lambda \geq 0$.

From (20), we have that $J(\mathbf{q}^*+\Delta\mathbf{q};\bar{\mathbf{q}}) \to 0$ if $\lambda \to 0$, and the perturbation directions $\Delta\mathbf{q}$ associated with negligible eigenvalues of $\chi^T\chi$ are indicative of parameters that are unidentifiable.

2.5.3 Parameter Candidate Selection (PCS)

Our main procedure to find, at each iteration, parameters $\mathbf{q}^*$ (identifiable) and $\bar{\mathbf{q}}$ (unidentifiable) satisfying $\mathbf{q}^*_{\mathbf{k}} = \langle \mathbf{q}^*, \bar{\mathbf{q}} \rangle$ relies on at least one call in every iteration to an underlying algorithm, termed the Parameter Candidate Selection Algorithm (PCS, Algorithm 1). This algorithm is based on techniques from prior works utilizing parameter subset selection (Burth et al. 1999; Cintrón-Arias et al. 2009; Quaiser and Monnigmann 2009), that is tailored to our specific application and unique features of our data. The PCS Algorithm is

Algorithm 1 Parameter Candidate Selection (PCS)

Input: Vector $\mathbf{q}^* \in \mathbb{R}^j$ for given j; observations $\{y_i^D\}_{i \in I}$ where $N = |I|$

Output: L=PCS($\mathbf{q}^*, \{y_i^D\}_{i \in I}$) where subset $L \subseteq \{1, \ldots, j\}$ are indices of candidate least identifiable parameters

1: Specify tolerances $0 < \eta \ll 1$ and $0 \ll \delta < 1$.
2: Construct $N \times j$ sensitivity matrix $\chi_s(\mathbf{q}^*)$.
3: Compute ordered eigenvalues $\lambda_1 \leq \ldots \leq \lambda_j$ of $\chi_s^T \chi_s(\mathbf{q}^*)$.
4: $L \leftarrow \emptyset$
5: **if** $\lambda_1 \geq \eta$ **then**
6: **return** L
7: **else**
8: Find index m for which $\lambda_m < \eta$ and $\lambda_{m+1} \geq \eta$.
9: **for** $i = 1 : m$ **do**
10: Compute normalized eigenvector $\mathbf{v}_i$ corresponding to λ_i.
11: **if** $\exists\ v_{l_i}$ component of $\mathbf{v}_i$ with $|v_{l_i}| \geq \delta$ **then**
12: $L \leftarrow L \cup \{l_i\}$
13: **end if**
14: **end for**
15: **return** L
16: **end if**

invoked within every iteration of our main procedure to determine indices of unidentifiable parameters for a given $\mathbf{q}^*$. Equation (20) provides the foundation for the algorithm. If the (scaled) information matrix has full rank, all corresponding parameters used in its evaluation are identifiable. For our purposes, we set a threshold η for eigenvalues of the information matrix (line 1); if all eigenvalues have magnitude above η, then $\chi_s^T \chi_s$ has full rank, numerically. In this case, all parameters in $\mathbf{q}^*$ are identifiable, and the PCS algorithm returns the empty list (lines 4-6). Otherwise, there is some index for which all of the preceding eigenvalues, ordered from smallest to largest, are less than the designated threshold η (lines 7-16). We then compute the normalized eigenvectors that correspond to these eigenvalues (line 10). The PCS Algorithm flags the components of these eigenvectors that are sufficiently large and returns their indices (lines 11-13); the index of any component having magnitude greater than $\delta < 1$ is a candidate for being deemed as unidentifiable within each iteration of our main procedure.

2.5.4 Outline of Main Parameter Selection Procedure

We now outline our main procedure to find parameter vectors $\mathbf{q}^*$ (identifiable) and $\bar{\mathbf{q}}$ (unidentifiable) such that $\mathbf{q}^*_{\mathbf{k}} = \langle \mathbf{q}^*, \bar{\mathbf{q}} \rangle$.

First, we initialize $\bar{\mathbf{q}} = [\,]$ and set the number of parameters to be estimated $j = M$ (here $M = 11$). The first iteration solves (17) for $\mathbf{q}^* \in \mathbb{R}^j$ (§2.5.1), with a prescribed initial vector $\mathbf{q}_{\text{init}} \in \mathbb{R}^j$. Algorithm 1 is then called with this vector $\mathbf{q}^*$.

Each time it is invoked, Algorithm 1 returns either: (a) a list L that is the empty set; or (b) a subset of the full index set $\{1, \ldots, j\}$ with indices of parameters that are flagged as candidates for being unidentifiable.

In case (a), the matrix $\chi_s^T \chi_s(\mathbf{q}^*)$ has full rank, and we deem all parameters in $\mathbf{q}^*$ identifiable. Hence the PCS procedure would yield $L = \emptyset$ and so $\mathbf{q}^*_{\mathbf{k}} = \langle \mathbf{q}^*, \bar{\mathbf{q}} \rangle$.

In case (b), consider the set of candidates for non-identifiability with corresponding indices stored in L. We wish to determine which candidate influences minimization of the objective function in (17) the least. For each index $l \in L$, we remove the l^{th} component q_l^* from $\mathbf{q}^*$, i.e. moving it temporarily to $\bar{\mathbf{q}}$, and repeat the estimation (17) over $\mathbf{q} \in \mathbb{R}^{j-1}$ using the remaining components of $\mathbf{q}^*$ as the initial parameter vector. With this approach, the solver in each loop iterate is guaranteed to output a cost that is no greater than the previous iterate's best cost. Let $m \in L$ be the index corresponding to the lowest cost. The current iteration then completes by updating $j \leftarrow j - 1$ and by moving q_m^* into $\bar{\mathbf{q}}$.

At the start of the next iteration, we update $\mathbf{q}_{\text{init}}$ to be $\mathbf{q}^*$ and then set $\mathbf{q}^*$ to be the solution of (17) using $\mathbf{q}_{\text{init}}$. Algorithm 1 is invoked again with the new $\mathbf{q}^*$, and we repeat this procedure until its terminal iteration when the information matrix has full rank (i.e. $L = \emptyset$).

As the dimension j of the parameter search space is successively reduced (i.e. as the dimensions of $\mathbf{q}$ and $\bar{\mathbf{q}}$ decrease and increase, respectively), a more robust optimization can be achieved. Metrics for evaluating such success include reduction in both the magnitude of the objective function and computational cost of the optimization as the subset selection procedure proceeds.

2.6 Model Reduction

The procedure described in §2.5.4 returns candidates for parameters that are unidentifiable in the curve-fitting routine for our model. Traditionally, unidentifiable parameters are fixed at nominal values, which coincide with the starting values for $\mathbf{q}_{\text{init}}$ passed to the optimization routine. Because our initial parameter values for this problem are unknown except by a relative ordering of magnitudes (see §3.1), we treat the outputs of the PCS Algorithm as the more realistic set of nominal values for our parameters. The parameters deemed unidentifiable are fixed at these values (when being moved into $\bar{\mathbf{q}}$) to ensure feasibility in the (local) parameter landscape. Among these parameter

candidates in $\bar{\mathbf{q}}$, we want to determine those, if any, that can be eliminated from the model by systematically fixing them at zero and attempting to optimize over the remaining parameters. For example, if q_l^* in $\mathbf{q}^*$ is flagged by the above procedure as the most unidentifiable parameter, we attempt to fix the corresponding l^{th} parameter at 0 to analyze that parameter's influence on the model in a more global sense. However, it is noted that a similar quality fit to the fibrin matrix data in terms of comparable objective costs is a necessary but insufficient condition for model reduction via parameter elimination. In addition, we must consider the effects of setting a parameter to zero not only on the values and sensitivities of the remaining model parameters but also on the reaction system as a whole. We note that these additional considerations are a result of physical constraints in our application; the only experimental observations for (17) correspond to the product (FM). The advantage of this approach in the context of our problem, if unidentifiable parameters can be set to zero, is that each parameter corresponds to a reaction rate in the model. By setting parameters equal to 0, we eliminate reactions from (1)-(3) and reduce the complexity of the reaction system.

3 Results

3.1 Implementation and Overview

The mathematical model of our reaction kinetics system (4)-(10) was incorporated into an implementation of the parameter estimation, parameter identifiability, and model reduction techniques outlined in §2.5 and §2.6 using Matlab® 2020b. For all parameter estimations carried out, the routine `fminsearch` was used to minimize the objective function in (17). Function and step size tolerances were set at 10^{-12} and 10^{-8}, respectively. In the context of the available data, the model exhibits a diverse range of time scales and the system of ODEs is stiff. Hence, the variable-step, variable-order (VSVO) ODE solver `ode15s` was used to numerically solve the system in all cases considered.

Our initial attempt at minimizing the objective function in (17), i.e. over $\mathbb{R}^{11}$, revealed a violation of underlying physical properties of the system (1)-(3). Specifically, the property that an enzyme concentration near steady state should be very close to its initial value, i.e. $Thb(t_N) \approx Thb_0$, was not preserved. Consequently, in all subsequent simulations we augmented the cost function (17) with a penalty term enforcing this constraint. Incorporation of this term precluded any further violations of the aforementioned condition in all results presented.

The details and progression of our iterative procedure for parameter estimation and identifiability described in §2.5.4 with $\eta = 10^{-10}$ and $\delta = 0.80$ in the PCS algorithm (and $h = 10^{-10}, \alpha = 0.2$ for the parameter scaling map) is shown in Tables 1 and 2. Initial values of all 11 parameters in $\mathbf{q}_{\text{init}}$ were prescribed as multiples of 10 based on presumed reaction rate orders of magnitude. Our procedure terminated in 6 iterations, i.e. with 6 parameters estimated in

the optimization algorithm at the final iteration.

Note that, within each iteration (Table 1), three calls to the function PCS (Algorithm 1) were made to identify a set of candidates for unidentifiable parameters (see §3.2). When, within each iteration, multiple eigenvalues were below the threshold η, eigenvector components (exceeding the threshold δ) for the case with the lowest cost were used to generate the set of unidentifiable candidates. Via this approach, we ensure that the cost in (17) decreases (Table 2) as the iterative procedure (Table 1) proceeds. In Fig. 3, a representative fit to our fibrin matrix data is shown for the case with the lowest cost (Table 1, Iter. 6).

We first motivate incorporation of a data partition into our main parameter selection procedure (§3.2) and then summarize key details of our parameter subset selection procedure by illustrating outcomes corresponding to several rows of Table 1 (§3.3-3.5). Lastly, we emphasize how unidentifiable parameters can be used to aid in model reduction, as outlined in §2.6, in the context of the given data (§3.6).

3.2 Data Partitioning for Additional Sensitivity Information

As detailed in §2.1, the duration of the experiment informing our model was roughly 4 hours (225 minutes), and rapid polymerization was observed within the first hour. Furthermore, parameter sensitivities for our initial curve fitting results over $\mathbb{R}^{11}$ (Fig. 2) exhibited two types of sensitivities: those that achieve their largest magnitude within the first hour of the experiment and those do so at later times. It is noted that some information about parameter identifiability was lost when computing sensitivities across the full data set. This is reflective of the multiple interacting time scales inherent to our application. To mitigate loss of information when using the full data set, we exploited the data-driven structure of the sensitivity matrix as well as the flexibility of the PCS algorithm to receive subsets of the full data as input (Algorithm 1). Specifically, the dataset for the full duration of the experiment, the data subset for the first hour of the experiment, and data subset for the remainder of the experiment (~ 3 hours) were all used in the main parameter selection procedure to generate a comprehensive set of candidates for non-identifiability.

3.3 The First Iteration

Our first parameter estimate $\mathbf{q}^* \in \mathbb{R}^{11}$ (Table 1, Iter. 1) results from minimization of the cost (17) using the prescribed initial guess $\mathbf{q}_{\text{init}}$ (Table 1, Iter. 0), noting that $\bar{\mathbf{q}} = \emptyset$. Via Algorithm 1, PCS($\mathbf{q}^*, \{y_i^D\}_{i=1}^N$) is then called on the three time intervals to determine a set of unidentifiable parameter candidates (Fig. 4).

Five eigenvalues of the full-time information matrix ($\chi_s^T \chi_s$) were found to lie below the specified threshhold η and are shown in Fig. 4a. The five corresponding eigenvectors were then examined (Fig. 4b-e), and components that

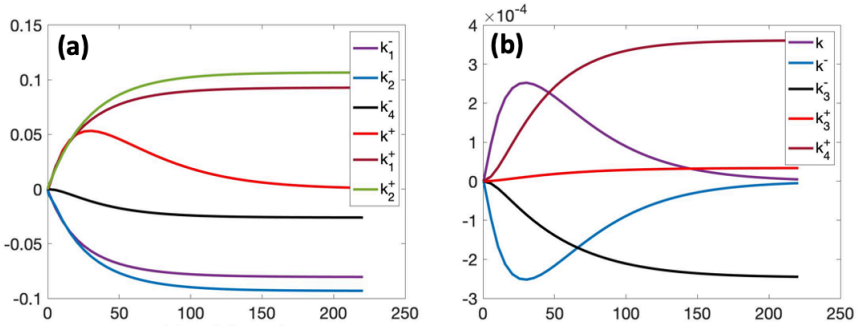

Fig. 2: Scaled sensitivities for parameters after initial curve fitting in $\mathbb{R}^{11}$ (evaluated at the values in Iter. 1 of Table 1). Parameters deemed identifiable by our main parameter subset selection procedure are shown in (a) while the unidentifiable parameters are shown in (b).

were above the threshold $\delta = 0.80$ were flagged (red). By repeating this process over all three time intervals, the candidate parameters $\{k_3^-, k_2^-, k_4^+, k, k^-, k_3^+\}$ were deemed to be potentially unidentifiable. Next, we determined which one of these parameters should be moved into $\bar{\mathbf{q}}$, i.e. fixed at a nominal value, at the next iteration. To this end, each of the six candidate parameters was (in turn) fixed at its estimated value within the current iteration (Table 1, Iter. 1), and a new parameter estimation in $\mathbb{R}^{10}$ was carried out. The parameter k_3^+ yielded the lowest cost in (17) and, hence, was chosen to move to $\bar{\mathbf{q}}$ at the next iteration (Table 2, Iter. 1).

3.4 Second Iteration

We next describe results for the second iteration of our overall algorithm to find $\mathbf{q}_\mathbf{k}^* = \langle \mathbf{q}^*, \bar{\mathbf{q}} \rangle \in \mathbb{R}^{11}$. Via Algorithm 1, PCS($\mathbf{q}^*, \{y_i^D\}_{i=1}^N$) was called with $\mathbf{q}^* \in \mathbb{R}^{10}$ and $\bar{\mathbf{q}} = [k_3^+]$ for all three time intervals (Table 1, Iter. 2). Initial

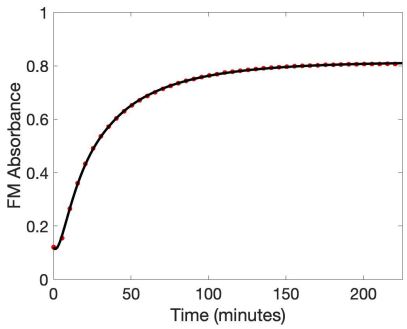

Fig. 3: Best fit for model predictions of fibrin matrix accumulation to the representative data (Fig. 1) based on the procedure for parameter estimation and identifiability (§2.5.4). The vector of parameters $\mathbf{q}_\mathbf{k}^*$ and final cost J correspond to Iter. 6 in Tables 1 and 2.

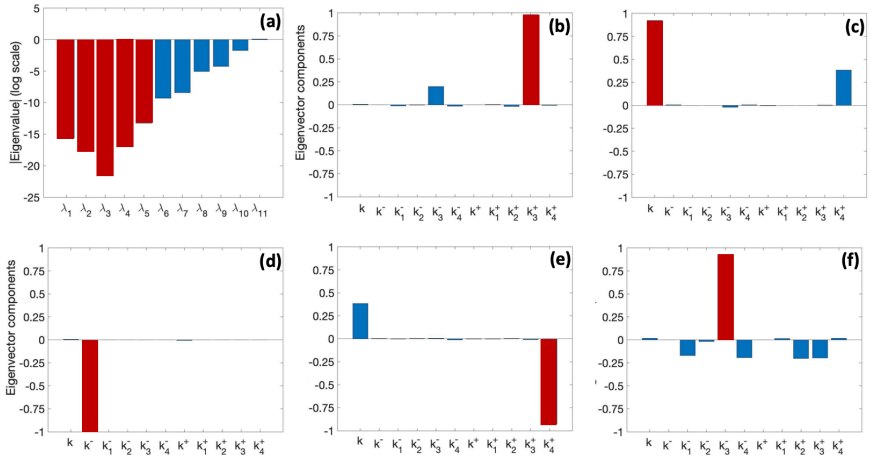

Fig. 4: Eigenvalues and eigenvectors of $\chi_s^T \chi_s$ for the full-time data set evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^{11}$ for Iter. 1 (Table 1). (a) five eigenvalues that are smaller than the threshold $\eta = 10^{-10}$ were flagged (red); (b-f) components of the associated eigenvectors having magnitude greater than $\delta = 0.80$ were flagged (red) to generate candidates for unidentifiable parameters (Table 2, Iter. 1)

values for the 10 components of $\mathbf{q}^*$ were set based on the previous iteration (Table 1, Iter. 1). The PCS results for the early- and full-time data sets yielded identical sets of unidentifiable parameter candidates; this is illustrated for the full-time data set in Fig. 5. For the late-time data set, an additional unidentifiable parameter candidate (k_3^-) was generated (Fig. 6). Hence, at this iteration, the candidate parameters $\left\{k^-, k_4^+, k_3^-, k\right\}$ were deemed to be potentially unidentifiable.

The cost comparison procedure (end of §3.3) was then run to determine the least identifiable parameter, i.e. to be fixed at a nominal value at the next iteration; the resulting parameter was k^-. Consequently, k^- was chosen as the parameter to move to $\bar{\mathbf{q}}$ at the next iteration (Table 2, Iter. 2).

3.5 Final Iterations

We then continued the procedure outlined above with results summarized in Tables 1 and 2 (Iters. 3-4). Consider the iteration for PCS on $\mathbf{q}^* \in \mathbb{R}^7$ (Table 1, Iter. 5). Parameter candidates identified from the full- (Fig. 7a-b) and early-time (Fig. 7c-d) data sets illustrated the capability of our data partitioning technique to extract a greater amount of information. Specifically, no parameter candidates exceeded the δ threshold for the full time data set (Fig. 7b),

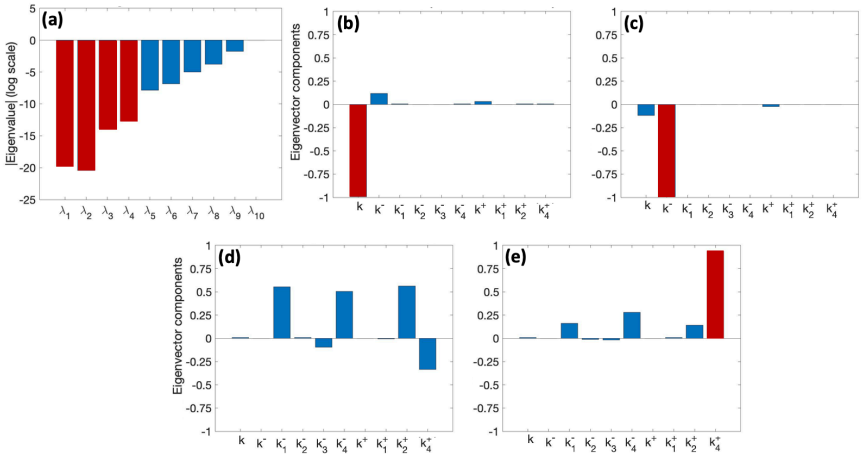

Fig. 5: Eigenvalues and eigenvectors of $\chi_s^T \chi_s$ for the full-time data set evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^{10}$ for Iter. 2 (Table 1). (a) four eigenvalues that are smaller than the threshold $\eta = 10^{-10}$ were flagged (red); (b-e) components of the associated eigenvectors having magnitude greater than $\delta = 0.80$ were flagged (red) to generate candidates for non-unidentifiableidentifiable parameters.

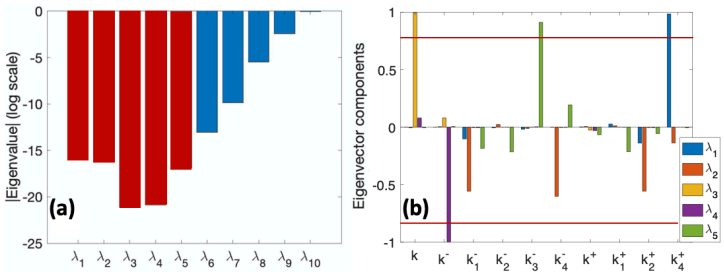

Fig. 6: Eigenvalues and eigenvectors of $\chi_s^T \chi_s$ for the late-time data set evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^{10}$ with $\bar{\mathbf{q}} = [k_3^+]$. (a) six eigenvalues were smaller than the threshold $\eta = 10^{-10}$, but for readability, the largest of these eigenvalues was excluded (it generated no additional parameter candidates); (b) components of the first five eigenvectors are shown, and those having magnitude greater than $\delta = 0.80$ were used to generate candidates for unidentifiable parameters (Table 2, Iter. 2).

whereas the early-time data set found one unidentifiable parameter candidate (k_3^-, Fig. 7d).

After moving k_3^- into $\bar{\mathbf{q}}$ (Table 2, Iter. 5), we then proceeded to the sixth (and final) iteration of our procedure (Tables 1-2, Iter. 6). Here, calling PCS using the full data set returned eigenvalues that were all above the thresh-

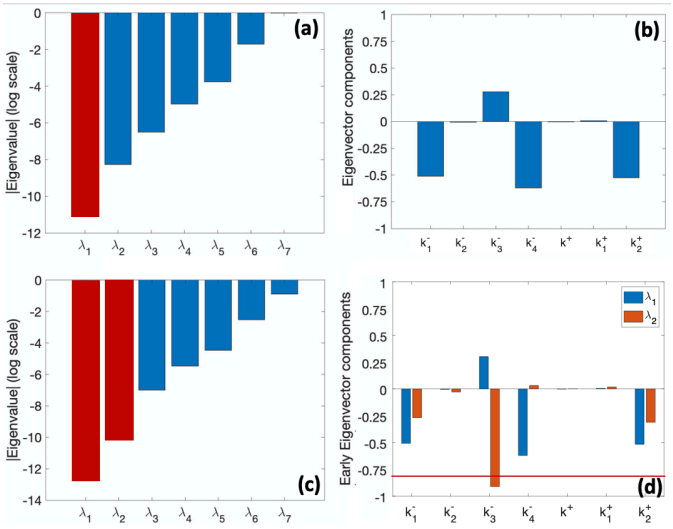

Fig. 7: Eigenvalues and eigenvectors of $\chi_s^T \chi_s$ for Iter. 5 (Table 1) of our procedure evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^7$ for the: (a-b) full-time data set; (c-d) early-time data set. In (b), no eigenvector components exceeded the threshold $\delta = 0.80$. In (d), a single parameter k_3^- was found (in the second eigenvector) to exceed the threshold $\delta = 0.80$.

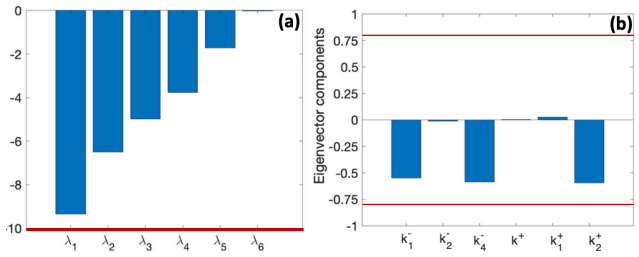

Fig. 8: Eigenvalues and eigenvectors of $\chi_s^T \chi_s$ for Iter. 6 (Table 1) of our procedure evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^6$ for the full-time data set. (a) There were no eigenvalues below the threshold $\eta = 10^{-10}$. (b) For the smallest eigenvalues, no eigenvector components exceeded the threshold $\delta = 0.80$.

old η (Fig. 8a). To ensure that the procedure was complete, we checked the eigenvector corresponding to the smallest eigenvalue to verify that none of its components were above the threshold δ (Fig. 8b). The PCS algorithm yielded no additional parameter candidates for early- and late-time data sets. Lastly, we varied the two threshold parameters η and δ to demonstrate that our set of identifiable parameters was preserved over reasonable ranges of η and δ (Table 3).

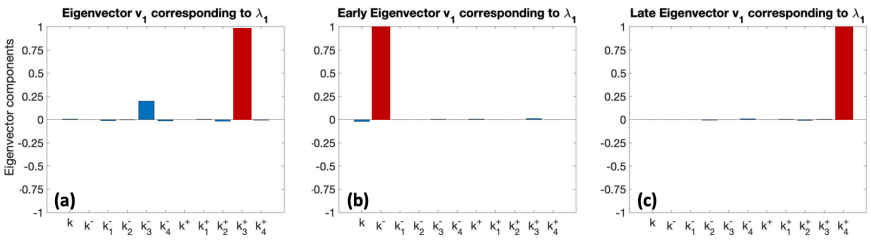

Fig. 9: Eigenvectors corresponding to smallest eigenvalue of each information matrix $\chi_s^T \chi_s$ evaluated at the optimal vector $\mathbf{q}^* \in \mathbb{R}^{11}$ for the 11-parameter run described in §3.6 on data sets for: (a) full time; (b) early time; (c) late time.

3.6 Model Reduction

Upon termination of the main parameter selection procedure (§2.5.4), which determined the set of unidentifiable parameters (Table 2, Iter. 6), we also considered using the results obtained for model reduction. Specifically, we aimed to identify which, if any, of the unidentifiable parameters could be set equal to 0, i.e. eliminating a reaction from our overall reaction system (1)-(3). However, in carrying out such a reduction, considerations must be made for model plausibility in the context of the application and the available data used for parameter estimation.

To this end, we used the output $\mathbf{q}^*_{\mathbf{k}} = \langle \mathbf{q}^*, \bar{\mathbf{q}} \rangle$ (Table 1, Iter. 6) as the starting point $\mathbf{q}_{\text{init}}$ for optimization over $\mathbb{R}^{11}$, as in the first iteration of our main procedure. In this additional optimization run, the objective cost $J = 4.324 \cdot 10^{-6}$ was preserved (Table 2, Iter. 6), and the identifiable parameter values were effectively the same (changed by less than $10^{-5}\%$ relative to Table 2, Iter. 6). We then ran one iteration of the PCS procedure on the resulting optimal parameter vector $\mathbf{q}^* \in \mathbb{R}^{11}$ (Fig. 9).

To determine the most unidentifiable parameters as potential candidates for model reduction, the eigenvector for the smallest eigenvalue (below η) was evaluated for each of the three time intervals (Fig. 9). The three parameters identified in this manner were k_3^+ (Fig. 9a), k^- (Fig. 9b) and k_4^+ (Fig. 9c). We next considered how the model is affected by setting, separately, each of these three parameters to zero. Specifically, for each case, we used optimal parameter values from the $\mathbb{R}^{11}$ fit described above as the starting point to optimize over the remaining 10 parameters, with the 11th parameter fixed at zero. All three cases yielded a final objective cost J in the range $(4.0, 4.6) \cdot 10^{-6}$. We also checked that no other parameters except for those in Fig. 9 could be set to 0 without resulting in a poor fit to the data.

In assessing the physical plausibility of model reductions for the parameters in Fig. 9, we view our reaction system as comprising three subsystems: a Michaelis-Menten-type subsystem (1), a cooperative enzyme kinetics subsystem with a single intermediate complex (2), and a cooperative enzyme kinetics

subsystem with two intermediate complexes (2-3). Of our three parameter candidates for model reduction, k^- can be set to 0 without impeding progression of the reaction system towards synthesis of the product (FM). The parameters k_3^+ and k_4^+ are forward reactions in (3) that we observed could (simultaneously) both be fixed at zero while maintaining the same quality of fit to the fibrin matrix data. Doing so would suggest that an entire mechanism for synthesis of the product might be excludable from the overall model. However, given that this subsystem also contained an identifiable parameter k_4^-, we proceed conservatively with model reduction (see also §4) and conclude that k^- is the only parameter in the context of our data that can be set equal to 0 (Table 4).

4 Discussion and Conclusions

In the context of a single representative data set (Fig. 1), we have demonstrated the advantages of a tailored approach to parameter estimation, parameter identifiability and model reduction for a dynamic enzyme kinetics model (with cooperativity) (4)-(11) of fibrin matrix accumulation in polymerization experiments.

4.1 Methodological Considerations

We can contrast the approach used in this study to a more ad hoc curve-fitting procedure that would, effectively, commence with a reasonable initial guess (Table 1, Iter. 0) and obtain a single set of estimated parameter values with an associated cost (Table 2, Iter. 0). Ultimately, the approach developed in this study was more robust in that it yielded a significant reduction in objective function cost, i.e. $J = 4.514 \cdot 10^{-6}$ (Table 3) versus $J = 58.68 \cdot 10^{-6}$ (Table 2, Iter. 0). Concurrently, it categorized 5 of the 11 reaction rate parameters as being unidentifiable (Table 2, Iters. 5-6) and eliminated one of these parameters (k^-), and its associated reaction in (1), from the overall system. The potential perils of incorporating the information matrix into parameter estimation are well known (Vallisneri 2008). Thus, our findings demonstrate the high degree of information within a single data set, as well as the manner in which it can be systematically extracted and utilized to both improve the underlying optimization procedure, and to reduce complexity of the overall model (§3.6).

In particular, the mathematical property that the information matrix $\chi_s^T \chi_s$ is built from the sensitivity matrix χ_s, which has the number of data points as one of its dimensions, allows for tailoring of the technique to the problem under consideration. For models such as ours with a moderate number of parameters, multiple interacting time scales, and limited data for parameter estimation, local sensitivities on their own can be difficult to interpret.

The (symmetric) information matrix imposes additional algebraic structure by admitting bases in terms of model parameters for the unidentifiable and identifiable (complementary) subspaces of the admissible parameter space. In contrast to local sensitivities, this additional structure combined with our data partitioning technique enables aggregation of sensitivity information across time intervals (or sub-intervals) and can also yield covariance estimates for uncertainty quantification (Cintrón-Arias et al. 2009). Bayesian techniques for uncertainty quantification, such as Markov Chain Monte Carlo, can exhibit more accurate and robust performance when carried out subsequent to identifiability analysis via parameter subset selection (Smith 2013).

4.2 Biological Considerations

We can also view, in the context of the biological application, the collection of five parameters that were wholly eliminated (k^-) or were deemed to be unidentifiable $\left\{k_3^+, k_3^-, k_4^+, k\right\}$. Notably, k^- is the rate for the first reverse reaction in (1). The ability to eliminate it from the overall reaction system is supported by the observation that, once the process of activating (inactive) fibrinogen commences, it rarely deactivates due to rapid conversion of the first intermediate complex (C_0) to active fibrinogen (Fbn_a). A few observations regarding the remaining sets of four unidentifiable parameters $\left\{k_3^+, k_3^-, k_4^+, k\right\}$ and six identifiable parameters $\left\{k_1^+, k_1^-, k_2^+, k_2^-, k_4^-, k^+\right\}$ can also be made. Specifically, three of four parameters in the reaction subsystem (3) were deemed unidentifiable whereas, in the reaction subsystem (2) all four parameters were found to be identifiable. Of note here is that the reaction system (2) involves species with non-zero initial states (Thb, FM) and for which the parameter estimation procedure accounted for underlying constraints (Thb); by contrast all species but FM in (3), i.e. $(\mathrm{Fbn}_a, \mathrm{C}_1, \mathrm{C}_2)$, are intermediate species in the overall reaction system. We also note that our model contains multiple timescales, corresponding to each reaction rate constant. For a given data set for a quantity of interest, the subset of identifiable (and thus influential) parameters could delineate inherent fast and slow scales of potential use in rescaling the model for further analytical investigation.

Taken together, these findings demonstrate that robust parameter estimation for dynamic (ODE) models of enzyme reaction kinetics in wound healing applications can depend on the scope of available data, i.e. capabilities or constraints for measuring underlying species in the model. Our findings also demonstrate dependence on the nature of each species, e.g. some species are primary quantities of interest with available data (e.g. fibrin matrix), others are intermediate complexes, and some are enzymes with inherent constraints (e.g. thrombin). Future studies tied to our fibrin polymerization experiments (§2.1) will explore expanded data sets examining accumulation rates and steady-state values of fibrin matrix content as a function of initial thrombin concentration. Specifically, the extent to which the subset of influential parameters identified in this study can delineate cases within (versus across)

sample groups with the same initial design (within a group) will be considered.

This overall approach will also be considered for an extended biomimetic system in which the polymerization experiment species are augmented with platelet-like (microgel) particles (PLPs) (Nandi et al. 2020). PLPs are comprised of advanced biomaterials coated with antibodies that rapidly deform and bind to fibrin matrix during polymerization. In a separate regime after polymerization, PLPs continue to slowly contract the polymerized clot network by tugging on nearby fibrin matrix fibers as each PLP particle returns to an energetically favorable conformation. Ultimately, these approaches could be integrated into models for in vitro or in vivo wound healing systems to investigate coupled effects among PLPs and fibroblasts during cell migration and matrix (clot) retraction; here, recent advances in single cell analysis and imaging can be highly beneficial (Guerrero-Juarez et al. 2019; Haensel et al. 2020; Nandi and Brown 2017). Overall, the approaches and techniques developed in this study can serve as a modeling and parameter estimation foundation for such extended systems as well other applications involving enzyme-mediated kinetics models of biochemical systems.

Parameter Values by Iteration											
Iter.	k	k^-	k_1^-	k_2^-	k_3^-	k_4^-	k^+	k_1^+	k_2^+	k_3^+	k_4^+
0	10000	100	100	100	100	100	1	1	100	1	1
1	21034	100	196	1.46	131	38.2	0.041	0.71	237	2.66	1.59
2	12906	366	242	1.68	3.18	83.5	0.039	0.97	310	□	0.46
3	1931	□	252	0.27	27.3	383	0.031	0.51	299	□	16.6
4	□	□	248	0.26	26.8	373	0.031	0.49	297	□	15.5
5	□	□	250	0.26	26.9	375	0.031	0.49	298	□	□
6	□	□	250	0.26	□	375	0.031	0.49	298	□	□

Table 1: Results after each iteration of our procedure for parameter estimation and identifiability (§2.5.4). The last row shows the final output $\mathbf{q}^*$. Entries with a □ reference the last numerical entry within the same column.

Objective Function Costs by Iteration		
Iter.	Candidates to Fix	J ($\cdot 10^{-6}$)
0	[ ]	58.68
1	$[k_3^-]$	47.31
	$[k_2^-]$	36.27
	$[k_4^+]$	32.35
	$[k]$	31.10
	$[k^-]$	30.12
	$[k_3^+]$ ■	26.11
2	$[k_3^+, k]$	22.45
	$[k_3^+, k_4^+]$	22.30
	$[k_3^+, k_3^-]$	5.268
	$[k_3^+, k^-]$ ■	4.683
3	$[k_3^+, k^-, k_3^-]$	4.683
	$[k_3^+, k^-, k_4^+]$	4.683
	$[k_3^+, k^-, k]$ ■	4.442
4	$[k_3^+, k^-, k, k_3^-]$	4.348
	$[k_3^+, k^-, k, k_4^+]$ ■	4.326
5	$[k_3^+, k^-, k, k_4^+, k_3^-]$ ■	4.324
6	$L = \emptyset$	**4.324**

Table 2: Intermediate fixed parameters and costs with each iteration of our procedure for parameter estimation and identifiability (§2.5.4). Within each iteration, ■ signifies the vector the procedure chooses as $\bar{\mathbf{q}}$. The procedure terminates at Iter. 6 when Algorithm 1 returns an empty list, indicating that there are no additional candidates for unidentifiable parameters.

η \ δ	0.75	**0.80**	0.85	0.90
10^{-8}	$[k_3^+, k^-, k, k_4^+, k_3^-]$ $4.324 \cdot 10^{-6}$	$[k_3^+, k^-, k, k_4^+, k_3^-]$ $4.324 \cdot 10^{-6}$	$[k_3^+, k^-, k, k_4^+]$ $4.348 \cdot 10^{-6}$	$[k_3^+, k^-, k, k_4^+]$ $4.348 \cdot 10^{-6}$
$\mathbf{10^{-10}}$	$[k_3^+, k^-, k, k_4^+, k_3^-]$ $4.324 \cdot 10^{-6}$	$[\mathbf{k_3^+, k^-, k, k_4^+, k_3^-}]$ $\mathbf{4.324 \cdot 10^{-6}}$	$[k_3^+, k^-, k, k_4^+]$ $4.348 \cdot 10^{-6}$	$[k_3^+, k^-, k]$ $4.442 \cdot 10^{-6}$
10^{-12}	$[k_3^+, k^-, k]$ $4.442 \cdot 10^{-6}$	$[k_3^+, k^-, k]$ $4.442 \cdot 10^{-6}$	$[k_3^+, k^-, k]$ $4.442 \cdot 10^{-6}$	$[k_3^+, k^-, k]$ $4.442 \cdot 10^{-6}$

Table 3: Unidentifiable parameters determined by our main parameter selection procedure with corresponding objective function values for a range of δ and η values.

Final Parameter Values										
k	k^-	k_1^-	k_2^-	k_3^-	k_4^-	k^+	k_1^+	k_2^+	k_3^+	k_4^+
2064	0	290	0.286	22.7	366	0.0284	0.484	300	2.64	15.5

Table 4: Final values of all estimated model parameters at the end of the parameter subset selection and model reduction procedures (for this case $J = 4.514 \cdot 10^{-6}$, Fig. 9).

References

1. Banks HT, Bekele-Maxwell K, Bociu L, Noorman M, Tillman K (2015) The complex-step method for sensitivity analysis of non-smooth problems arising in biology. Eurasian Journal of Mathematical and Computer Applications, 3:15–68.
2. Brendel M, Bonvin D, Marquardt W (2006) Incremental identification of kinetic models for homogeneous reaction systems. Chemical Engineering Science, 61:5404-5420.
3. Briggs GE, Haldane JBS (1925) A note on the kinematics of enzyme action. Biochemical Journal, 19:338-339.
4. Brown AC, Baker SR, Douglas AM, Keating M, Alvarez-Elizondo MB, Botvinick EL, Guthold M, Barker TH (2015) Molecular interference of fibrin's divalent polymerization mechanism enables modulation of multiscale material properties. Biomaterials, 49:27-36.
5. Brun R, Kühni M, Siegristm H, Gujer W, Reichert P (2002) Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter sets. Water Research, 36:4113-4127.
6. Burth M, Verghese GC, Vélez-Reyes M (1999) Subset selection for improved parameter estimates in on-line identification of a synchronous generator. IEEE Transactions on Power Systems, 14:218-225.
7. Chatterjee MS, Denney WS, Jing H, Diamond SL (2010) Systems biology of coagulation initiation: kinetics of thrombin generation in resting and activated human blood. PLoS Computational Biology, 6:e1000950.
8. Chester D, Brown AC (2017) The role of biophysical properties of provisional matrix proteins in wound repair. Matrix Biology, 60-61:124-140.

9. Chester D, Marrow EA, Daniele MA, Brown AC (2019) Wound healing and the host response in regenerative engineering. In: Narayan R (ed) Encyclopedia of biomedical engineering. Elsevier, Amsterdam, 1:1-12.
10. Chernysh IN, Nagaswami C, Purohit PK, Weisel JW (2012) Fibrin clots are equilibrium polymers that can be remodeled without proteolytic digestion. Scientific Reports 2:879
11. Cintrón-Arias A, Banks HT, Capaldi A, Lloyd AL (2009) A sensitivity matrix based methodology for inverse problem formulation. Journal of Inverse and Ill-posed Problems 17:545-564.
12. Goutelle A, Maurin M, Rougier R, Barbaut X, Bourguignon L, Ducher M, Maire P (2008) The Hill equation: a review of its capabilities in pharmacological modelling. Fundamental and Clinical Pharmacology 22:633-648
13. Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, Yamaga K, Shestova O, Gay DL, Yang Z, Kessenbrock K, Nie Q, Pear WS, Cotsarelis G, Plikus MV (2019) Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. Nature Communications, 10(1):650.
14. Haensel D, Jin S, Sun P, Cinco R, Dragan M, Nguyen Q, Cang Z, Gong Y, Vu R, MacLean AL, Kessenbrock K, Gratton E, Nie Q, Dai X (2020) Defining epidermal basal cell states during skin homeostasis and wound healing using single-sell transcriptomics. Cell Reports, 30:3932-3947.
15. Holmes MH (2000) Introduction to the foundations of applied mathematics. Springer, New York.
16. Hoffman M, Harger A, Lenkowski A, Hedner U, Roberts HR, Monroe DM (2006) Cutaneous wound healing is impaired in hemophilia B. Blood, 108:3053-3060.
17. Janmey PA, Winer JP, Weisel JW (2009) Fibrin gels and their clinical and bioengineering applications. J R Soc Interface, 6:1-10.
18. Jorgensen SN, Sanders JR (2016) Mathematical models of wound healing and closure: a comprehensive review. Medical & Biological Engineering & Computing, 54:1297-1316.
19. Keener J, Sneyd J (1998) Mathematical Physiology. Springer, New York.
20. Lyness J (1967) Numerical algorithms based on the theory of complex variables. In Proc. ACM 22nd Nat. Conf., 4:124-134.
21. Lyness J, Moler C (1967) Numerical differentiation of analytic functions. SIAM J. Numer. Anal., 4:202-210.
22. Martins JRRA, Sturdza P, Alonso JJ (2003) The complex-step derivative approximation. ACM Transactions on Mathematical Software, 29:245–262.
23. Michaelis L, Menten MI (1913) Die kinetik der invertinwirkung. Biochem Z, 39:333-369.
24. Nandi S, Brown AC (2017). Characterizing cell migration within three-dimensional in vitro wound environments. Journal of Visualized Experiments, 16: 126.
25. Nandi S, Sommerville L, Nellenbach K, Mihalko E, Erb M, Freytes DO, Hoffman M, Monroe D, Brown AC (2020) Platelet-like particles improve fibrin network properties in a hemophilic model of provisional matrix structural defects. Journal of Colloid and Interface Sciences, 577:406-418.
26. Nelder J, Mead R (1965) A simplex method for function minimization. Comput J, 7:308-313.
27. Quaiser T, Mönnigmann M (2009) System identifiability testing for unambiguous mechanistic modeling – application to JAK-STAT, MAP kinase, and NF-κ B signaling pathway models. BMC Systems Biology, 3:50.
28. Rothenberg TJ (1971) Identification in parametric models. Econometrica, 39:577-591.
29. Smith R (2013) Uncertainty Quantification: Theory, Implementation, and Applications. Society for Industrial and Applied Mathematics, Philadelphia.
30. Sproul EP, Hannan R, Brown AC (2018) Characterization of fibrin-based constructs for tissue engineering. In: Chawla K (ed) Biomaterials for tissue engineering: methods and protocols, Methods in molecular biology, Springer, New York, 1758:85-99.
31. Squire W, Trapp G (1998) Using complex variables to estimate derivatives of real functions. SIAM Review, 10:100-112.
32. Tranquillo RT, Murray JD (1992) Continuum model of fibroblast-driven wound contraction: inflammation-mediation. Journal of Theoretical Biology, 158:135-172.
33. Vajda S, Rabitz H, Walter E, Lecourtier Y (1989) Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chem. Eng. Commun., 83:191-219.

34. Vallisneri M (2008) Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. Physical Review D, 77: 042001.
35. Valero C, Javierre E, Garcia-Aznar JM, Menzel A, Gomez-Benito MJ (2015) Challenges in the modeling of wound healing mechanisms in soft biological tissues. Annals of Biomedical Engineering, 43:1654-1665.
36. Vermolen FJ, Gefen A (2013) A phenomenological model for chemico-mechanically induced cell shape changes during migration and cell-cell contacts. Biomechanics and Modeling in Mechanobiology, 12:301-323.
37. Wang Y, Guerrero-Juarez CF, Qiu Y, Du H, Chen W, Figueroa S, Plikus MV, Nie Q (2019) A multiscale hybrid mathematical model of epidermal-dermal interactions during skin wound healing. Experimental Dermatology, 28:493-502.
38. Weihs D, Gefen A, Vermolen FM (2016) Review on experiment-based two- and three-dimensional models for wound healing. Interface Focus, 6:20160038.
39. Weisel JW (2004) The mechanical properties of fibrin for basic scientists and clinicians. Biophysical Chemistry, 112:267-276.
40. Weisel JW (2005) Fibrinogen and fibrin. Advances in Protein Chemistry, 70:247-299.
41. Weisel JW, Litvinov RI (2013) Mechanisms of fibrin polymerization and clinical implications. Blood, 121:1712-1719.
42. Weisel JW, Nagaswami C (1992) Computer modeling of fibrin polymerization kinetics correlated with electron microscope and turbidity observations: clot structure and assembly are kinetically controlled. Biophysical Journal, 63:111-128.
43. Wolberg AS (2007) Thrombin generation and fibrin clot structure. Blood Reviews, 21:131-124.
44. Wolberg AS, Gabriel DA, Hoffman M (2002) Analyzing fibrin clot structure using a microplate reader, Blood Coagulation & Fibrinolysis, 13:533-539.