ABSTRACT

REHM, KERI LEIGH. Multiscale Modeling of Plant Growth Combining Enzyme Kinetics and Whole Plant Dynamics and Experimental Design Applications. (Under the direction of H. T. Banks.)

While models of photosynthesis and physiological models of crop growth abound, the questions of how the productivity of photosynthesis drives plant growth and how the environment (outside of PAR, CO_2, and O_2 availability [49]) affects these cellular level reactions remain largely unexplored. We construct two environment-dependent candidate model of plant metabolism and growth based on previous work modeling the dark reactions of photosynthesis [49], light reactions of photosynthesis [28], and whole-plant metabolism [26], along with information on carbon metabolism from [4] and [42] and whole-plant response to environmental factors from [39]. The carbon uptake rate as predicted by the model is used to relate photosynthetic productivity with plant leaf area as predicted by a logistic growth model.

We use these models with a least-squares optimization algorithm to estimate parameters that would replicate recorded growth of *Arabidopsis thaliana* [35] and soybean [39]. The leaf area predicted by the model agrees with the data and exhibits changes in behavior between day and night. Moreover, metabolites involved in photosynthetic processes exhibit different concentrations when exposed to daytime conditions or nighttime conditions. To investigate potential relationships between parameters in the models and overall model behavior, we use a C3 cycle model [49] and one of the two constructed models to conduct further parameter estimation problems and calculation of sensitivity matrices describing how much an equation changes when a parameter is varied some small amount from a predetermined value.

Additionally, we formulate a Fisher information matrix-based optimal design methodology for the selection of the best state variables to observe and optimal sampling times for parameter estimation problems involving complex nonlinear dynamical systems. An iterative algorithm for implementation of the resulting methodology is proposed. Its use and efficacy is illustrated on two applied problems of practical interest, a dynamic model of HIV and immune response [1] and a C3 cycle model [49].

Multiscale Modeling of Plant Growth Combining Enzyme Kinetics and
Whole Plant Dynamics and Experimental Design Applications

by
Keri Leigh Rehm

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2013

APPROVED BY:

Lorena Bociu	Ralph Smith

Hien Tran	H. T. Banks
	Chair of Advisory Committee

DEDICATION

To my husband, for his love and encouragement through every step of my graduate career.
And to my family, for their loving support in all my endeavors.

BIOGRAPHY

The author spent her early childhood in the towns of Endicott, NY, and Poughkeepsie, NY, before moving to Cary, NC, at the age of 9 with her parents and two brothers. She graduated from Green Hope High School in Morrisville, NC, in the top 1% of her class in 2004 and from Meredith College in Raleigh, NC, with a B.S. in Mathematics, B.A. in Religion and Philosophy, and minors in Statistics and Computer Science, summa cum laude, in 2008. During her time at Meredith College and at North Carolina State University, she has been very active in a variety of organizations, including service in the Meredith College Academic Technology, Lillian Parker Wallace Lecture Series, and Sonia Kovalevsky Day committees and being an executive officer in the University Graduate Student Association at North Carolina State University for 2.5 years in addition to being a leader in the worship ministry at her local church. When at home, she enjoys playing video games and dabbling in crafts ranging from crocheting to woodworking.

ACKNOWLEDGEMENTS

I'd like to acknowledge the many people who have contributed to my successes over the past 26 years. My family and my husband, first and foremost, for their constant support, advice, and encouragement in my academic and extracurricular pursuits. My friends, including my colleagues at North Carolina State University and those whom I consider my second family through church, for reminding me to enjoy life – including my work. My many teachers over the years who encouraged and inspired me to not only remain passionate about my education in but also take leadership roles in the classroom and in the community. My collaborators at Syngenta Biotechnology, Inc., for their advice and input. And most certainly my Ph. D. research advisor, for his guidance across several research projects and his mentoring as a mathematical researcher.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Multiscale Modeling of Plant Growth Combining Enzyme Kinetics and Whole Plant Dynamics

1.1 Introduction

The development and refinement of dynamical systems models of photosynthesis reflect the increased understanding of plant metabolism built through numerous in vivo, in vitro, and even in silico experiments. These models typically address light-dependent reactions that occur within the thylakoid or light-independent reactions of the Calvin Cycle in the stroma and cytosol; however, very few models address both sets of reactions - the exception being the work of A. Laisk in [27] and [26]. The questions of how these reactions drive plant growth and how the environment (outside of sunlight, CO_2, and O_2 availability) affects these reactions remain largely unexplored.

Utilizing the models of Zhu, et. al. [49], Lazar [28], and Laisk, et. al. [26], we construct two ODE candidate models of photosynthesis in C3 plants such as *Arabidopsis thaliana*, soybean, and spinach. One candidate model, labeled the comprehensive model, uses a set of 74 first order reaction equations developed in [28] to describe the light reactions. The other, labeled the compact model, utilizes four Michaelis-Menten enzyme kinetic equations to summarize these reactions. Both models use the C3 cycle model of [49] as well as other mechanism descriptions from [26]. We relate this cellular-level productivity to environmental conditions such as sunlight, CO_2, and O_2 availability and temperature by drawing upon mathematical models developed in the area of crop modeling [39] and knowledge of how environmental factors affect photosynthesis [4]. Changes in leaf area are determined using sigmoid-like functions scaled by cellular activity as calculated by these cellular-level reactions.

1.1.1 Physiology of photosynthesis

Photosynthesis is a process by which plants convert sunlight, CO_2, and H_2O into energy that is usable by the plant and other organisms that may consume the plant. Photosynthesis primarily occurs in the leaves of a plant. A leaf is composed of several layers of cells, as pictured in Figure 1.1. The cuticle, the outermost layer of a leaf is not cellular - it is a waxy protective layer. The exterior layer of cells, the epidermis, controls the transaction of water, air, and other substances into the leaf but does not contribute much to photosynthetic productivity. The body of the leaf, known as the mesophyll, contains the cells that perform photosynthesis.
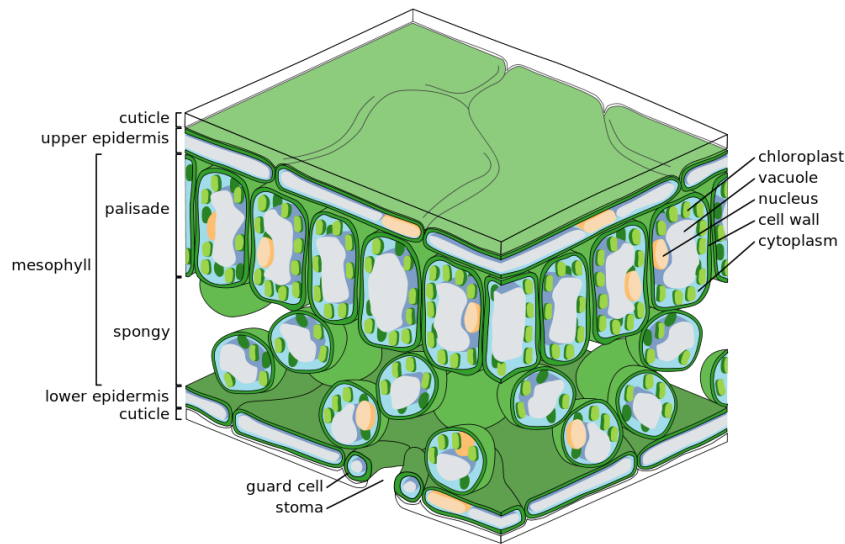
Figure 1.1: Depiction of a cross-section of a leaf [48].

A leaf cell is composed of many organelles that perform specific tasks within a cell. The names and general locations of these organelles are pictured in Figure 1.2. Photosynthesis occurs in the chloroplasts and cellular cytosol (also known as cytoplasm), and additional energy-producing reactions occur in the mitochondria. The chloroplast, represented in Figure 1.3, further contains substructures that support photosynthesis and other cellular functions. The light reactions of photosynthesis occur primarily in the thylakoid, which is a reaction center, and the cholorplast stroma, a liquid-filled space between the thylakoid (and other structures) and the chloroplast's membrane. The light-independent reactions, which are also known as the C3 cycle or Calvin cycle, occur in the stroma and in the cytoplasm outside the chloroplast.

When light strikes a leaf, it excites electrons contained in the thylakoid's interior known as the lumen. Once excited, the electron moves along an electron transport chain in which it slowly releases energy to power processes that occur along the thylakoid membrane. In Photosystem I, the movement of these electrons causes $NADP^+$ to gain a negative charge and attract a H^+ atom to become NADPH. Names of $NADP^+$, NADPH, and other metabolites involved in photosynthesis may be found in Section A.1 of the appendix. NADPH is used as a hydrogen supplier to help reduce CO_2 in the light-independent reactions. In Photosystem II, electron movement instead creates a charge that moves H^+ and supplies energy to the enzyme ATP synthase. When activated, ATP synthase catalyzes the combination of ADP and available phosphate, Pi, into ATP. The bond between ADP and the newly added phosphate stores a large amount of energy, which is used to fuel many other reactions in the cell.
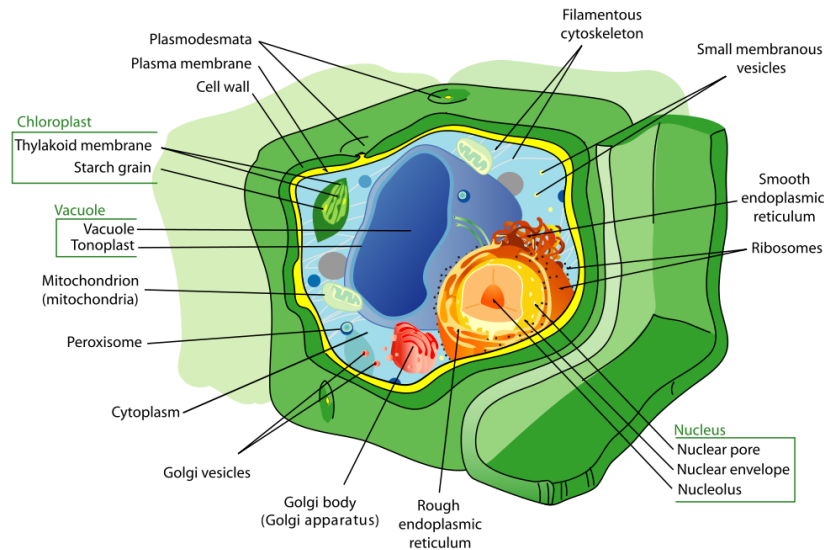
Figure 1.2: Depiction of a plant cell with organelles labeled [46].

While the reactions of the C3 cycle don't require light as a direct input, the rate of these reactions is dependent upon the concentration of the products from the light reactions. The core reactions of the C3 cycle occur in the chloroplast stroma and utilize the ATP and NADPH generated in the light reactions to create more stable energy storage metabolites in the form of sugars and starches. Glucose-based starch is generated inside the chloroplast and transported to

3
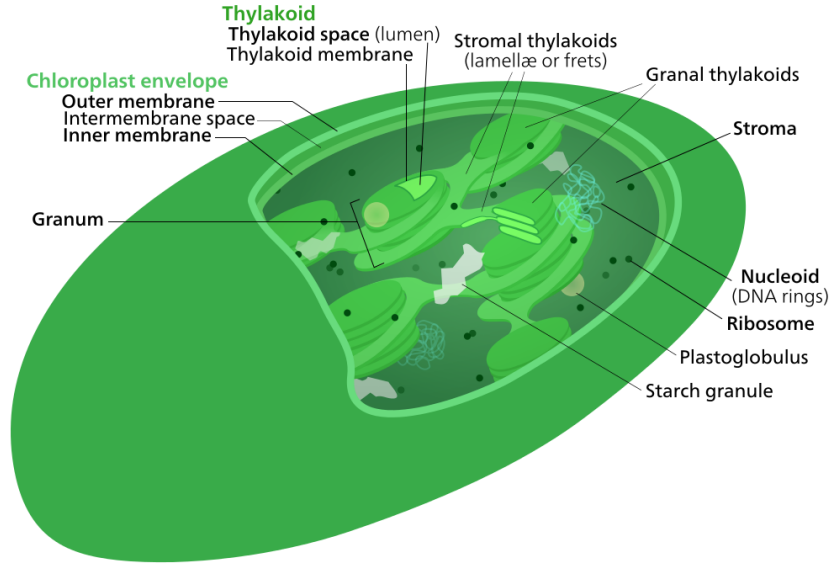
Figure 1.3: Depiction of a chloroplast with subsystems labeled [41].

other areas of the plant to fuel reactions and growth. Sucrose is generated in the cytoplasm and is primarily used to fuel processes in other areas of the same cell. Photorespiration, a process that competes for the same resources that are used in the C3 cycle, dissipates the energy in ATP and NADPH, and due to its high activity and impact on plant productivity, must also be considered when modeling the productivity of the C3 cycle. The reactions involved in the C3 cycle, sucrose production, and photorespiration are listed in [4] and [49].

1.1.2 Models of photosynthesis

The earliest steady-state models of C3 photosynthesis focus on key reactions involved in ATP, starch, and sugar generation and assume that most reactions exhibit Michaelis-Menten [29] type dynamics (the primary exception being reactions catalyzed by RuBisCO). Farquhar, et. al. [20] relate a leaf's CO_2 assimilation rate to radiance and temperature. Based on previous knowledge gathered through experiments on chloroplasts harvested from spinach, eucalyptus, and barley, the model calculates the rates of RuBP, ADP, ATP, $NADP^+$, NADPH, and PGA production and consumption based on electron transport that is fueled by radiance and controlled by temperature and atmospheric CO_2 and O_2. The model is tested at different atmospheric concentrations of CO_2 and O_2 while varying the temperature and level of light absorbed by the chloroplasts. While this model relates energy output to energy input with few equations, it does not utilize the full system of reactions that are known to occur in C3 photosynthesis [4].

4

The C3 cycle model for the synthesis of sucrose and starch proposed by Laisk, et. al. [25] is foundational to many current models. While much more comprehensive than [20], it makes assumptions that the NADPH:NADP$^+$ ratio, UTP:UDP ratio, and total Pi remain constant and it removes the light-dependent reactions found in Laisk's earlier work [27]. This model uses 20 ODE's and 6 concentration balance equations to represent the dynamics of 23 state variables, including metabolite concentrations in the stroma and cytosol. Parameters are adjusted so that the metabolite concentrations achieve a steady state indicative of that seen in experimentally obtained data, and then the model's performance is tested by varying atmospheric CO_2 and O_2 levels as well as maximum activity of ATP synthase. This model, while approaching the level of complexity needed to model the Calvin cycle at the subcellular level, removes the environmental factors of temperature (by assuming a constant temperature of $25°$ C) and radiance.

Pettersson and Ryde-Pettersson [31] propose a competing model for the C3 cycle at equilibrium. Using Michaelis-Menten type dynamics for all reaction velocities, the model uses 16 ODES, 2 concentration balance equations, and 11 reaction equilibrium equations to describe the concentrations of 19 metabolites. This model is used to examine the effect of Pi concentration on other metabolites' concentrations. Like that of [25], this model shows great improvement in modeling photosynthetic production, but the lack of environmental factors, assumptions made to enforce a steady state, and C3 reactions ignored by the model render it insufficient for in silico experiments with varying environmental conditions.

Poolman, et. al [32] extend the model of [31] to include responses to the environmental factors of Pi and light. Additionally, many of the steady-state assumptions are removed. The assumptions used in this model are very similar to those of [25]: the NADPH and NADP$^+$ concentrations are fixed, and level of light is reflected by varying the maximum activity of ATP synthase instead of modeling the photosystems. The model is used to examine CO_2 assimilation and starch production rates when the concentrations of RuBisCO, SBPase, Pi, and triose phosphate translocator are varied.

The work by Zhu, de Sturler, and Long in [49] extends [25], [31], and [32] to reflect recent advances in understanding of plant metabolism. In [49], the reactions for fructose and sucrose synthesis in the Calvin Cycle as well as those of photorespiration (not to be confused with cellular respiration) are quantified in terms of Michaelis-Menten enzyme kinetic equations (or similar mechanisms for reactions known to not fulfill Michaelis-Menten assumptions) and is one of the most comprehensive C3 models available, containing 31 state variables and 158 parameters. The greatest weaknesses of this model are the conservation equations used to enforce constant levels of phosphates, NADPH+NADP$^+$, and ATP+ADP in the system. Parameters are found either in existing literature or estimated using a genetic algorithm, and then the model is solved at varying levels of atmospheric CO_2 and O_2 to observe predicted carbon uptake rates.

In [47], von Caemmerer, Farquhar, and Berry explain a slightly modified version of the

model from [20] and extend it to the leaf level, where light intensity has a more pronounced and nonlinear effect. In this later work, this model is expanded to better explain the effects of temperature change on the reaction rate equations and apply the model to canopy-level systems, but does not include any more metabolites. While this model may be an acceptable approximation of carbon fixation for plants in which the Calvin Cycle is not understood, models similar in scale to [49] may be preferable for species whose Michaelis-Menten constants are available through previous research or may be estimated.

Similar to the efforts in C3 cycle modeling, a number of models have also been developed to describe the light reactions of photosynthesis. These reactions also influence the CO_2 uptake rate and O_2 production [22] and produce ATP and NADPH necessary for fueling the C3 cycle. Many of these models focus on chlorophyll fluorescence for the first second of exposure to light. Like CO_2 uptake in the C3 cycle, chlorophyll fluorescence is a strong indicator of productivity of the photosystems. During this small time frame, not all mechanisms of photosynthesis have been activated; this simpler system composed of only the activated components is easier to model [22]. These models are then used to qualitatively compare long-term photosynthetic performance (with the expectation that higher initial output implies higher long-term output).

The current generation of the foundational model by Laisk ([27], [25]) is described and used by Laisk in [26]. This model, while simpler than [49] in its description of the C3 cycle, includes simple mechanisms to describe the photosystems, ATP generated by photophosphorylation, and the synthesis of some organic and amino acids. Parameters are experimentally determined using intact leaves or taken from previous versions of the model, and then the model is used to explore the effects of varying light and CO_2 on the CO_2 uptake rate. Like [49], this model also relies on conservation equations to maintain constant concentrations of phosphates, ADP+ATP, and PSI and PSII reaction centers. While this model's quantitative description of the photosystems is appealing and approximates long-term fluorescence yields, its simpler implementation of the C3 cycle may disagree with [49] in predicted metabolite concentrations.

In [22], Govindjee discusses the nearly 60 years of research that has been informative in his efforts to model Chlorophyll (Chl) *a* fluorescence. While he discusses many topics in his review, he expounds most upon the observed phases of fluorescence and their relation to experimentally observed photochemical processes. He lists the several phases of fluorescence activity as determined by local maxima and minima in fluorescence data and explains his proposed names and definitions of those phases (in an attempt to standardize the multiple naming schemes in existence at the time of the paper's creation). Then, using experimental findings from a number of papers, he consolidates the knowledge of energy transfer between different reaction centers and the time scales on which these reactions occur after light is introduced. The timing of these transfers is then compared to the fluorescence phases to link which electron transfers increase or decrease chlorophyll fluorescence. Many contemporary researchers now follow the fluorescence

phase naming scheme of [22].

Zhu and Govindjee [50] construct an ODE model of PSII, cyt b6/f, and the OEC containing 20 state variables. The reactions included in the model are assumed to be bidirectional and are mathematically described using first-order kinetics. While the model shows correlations between the saturation of state variables and different fluorescence phases, the predicted fluorescence did not correspond well to observed fluorescence in both the start time of the fluorscense phase or the level of fluorescense. This lack of fit may be attributed to parameters taken from literature and not estimated using numerical optimization.

In [28], Lazar presents a model of PSI, PSII, cyt b6/f, FNR, and the OEC. While the electron transfers that occur in the photosystems are catalyzed by enzymes like the C3 cycle, they are written using a compartmental model and modeled using bi-directional first-order kinetic expressions. The model contains approximately 40 states and 20 parameters. Some parameters are found in the literature, while others are estimated using least squares minimization. Lazar then uses the model to investigate the rate of photosynthesis at different light levels. The timing of fluorescences phases as predicted by the model agrees well with observed fluorescence; however, the levels of fluorescence were often overestimated by the model.

Rubin and Riznichenko [36] formulate a model of PSI, PSII, and cyt b6/f using a master equation approach. The parameters used in the model are primarily from literature, however, some are manually estimated. While the model produces a qualitatively acceptable curve, the level of fluorescence predicted differed. The authors then propose a scheme based on Brownian motion to model the movement of chemicals within reaction centers. Simulations are run to show a model trajectory of a plastoquinone molecule; however, these results are not then related back to fluorescence levels.

Energy production in plant cells is not limited to photosynthesis - it also occurs in the mitochondria through cellular respiration. In a recent review of current photosynthesis research, Amthor [4] calculates the number of photons needed to generate ATP. He considers percent of useful radiation absorbed by plants, ATP generated in the C3 (and if applicable, C4) cycle, and cellular respiration. He also estimates the amount of ATP used in starch building, mineral uptake, and protein synthesis. Amthor documents and balances the chemical reaction equations for the C3 cycle, C4 modification, glycolysis, and the TCA cycle. While the equations are not coded nor written as enzyme kinetics equations, they are easily converted should any additional mechanisms be needed to accurately model plant metabolism, and Michaelis-Menten constants may be found in other sources or estimated.

1.1.3   Physiological crop modeling

The field of physiological crop modeling approaches the question of predicting plant productivity at a more tangible level: standard approaches in crop modeling relate crop yields or biomass production to a plethora of environmental factors such as water, sunlight, soil nutrients, and crop management practices. The relatively simple Aquacrop model [43] utilizes basic mathematics and patterns observed in plant growth to mainly relate plant growth to water consumption using the formula $B = WP \times \sum Tr$, where B is biomass, WP is water productivity of the crop, and Tr is crop transpiration. As this model does not include factors that relate to carbon fixation efficiency, it would be difficult to integrate this model with cellular-level kinetics models, but it includes simple formulations about important ideas such as growth based on humidity and sunlight and the existence of canopy cover from neighboring plants, which inhibits growth.

The decision support system for agrotechnology transfer (DSSAT) [23] utilizes a modular algorithm and includes many more factors in plant growth and has been tested on about 16 crops in its history from 1988 to 2003. It utilizes the CROPGRO algorithm, which was originally developed for soybean and peanut growth but has been successfully extended to wheat. The DSSAT program and its modular components are sold, so the source code or mathematics of its current algorithm are not publicly available.

J. Jones, a developer of the DSSAT, however, has a history of investigating and mathematically describing relationships between environmental factors and plant growth since the 1970's. In [33], the leaf area and biomass production of young individuals of several species of plants are examined in three different temperature regimes. The first four weeks of growth of these species in each temperature regime was modeled using exponential curves with an acceptable fit. The authors concluded that the plants of the same species in different temperature conditions grew at different rates. Additionally, they observe that leaf area expansion and biomass production are directly related.

Plant growth at later periods of the seedling and vegetative stages, however, does not follow an exponential pattern. At around the same time as [33], T. Sinclair was also developing a mathematical model that relates the basic environmental factors of temperature, available water, nitrogen, and light levels to leaf area, biomass, and emergence of seeds. This model, appearing in [40] in an application to modeling soybean growth, makes use of a system of simple mathematical equations to describe daily changes in plant size and weight for multiple individuals of one species.

The mechanisms affecting leaf area change in soybean in [40] are expanded in [39], where the authors model how temperature and water availability affect leaf growth and senescence at an individual plant level using logistic functions to model net gained and net lost leaf area. The model results may be interpreted as plant area change over time. The model uses a beta

function to relate to daily temperature to plant growth efficiency, allowing it to be used to predict how plants will grow in different temperature regions. Like [40], this model also focuses on daily changes in plants; however, its formulation is very close to that of a system of ordinary differential equations. The model is fit to three sets of soybean leaf area data, each covering the 120 days of plant development after emergence from the soil, with relatively successful results.

1.1.4 Dynamical systems modeling and experimental design

The application of a model to approximate trends seen in data raises many methodological questions. Seber and Wild [44] provide an overview of the mathematical and statistical tools commonly used and refined by current research. While basic topics of different types of models (such as linear ODEs, sigmoid and nonlinear functions, and even stochastic models) and parameter estimation techniques are covered, multiple aspects of analyzing and improving and model's performance are also addressed. Additionally, [44] shows theoretical background to confirm the reliability of these techniques, offers some explanations of the standard numerical algorithms used to solve the parameter estimation problem, and discusses some of the common difficulties encountered in parameter estimation.

Banks and Tran [17] provide a primer on the basics of modeling and parameter estimation. While it does not address the same breadth of mathematical and statistical modeling topics as [44], it is a very useful introduction to the concept and implementation of the iterative modeling process to often-encountered situations in the physical and biological sciences. Banks and Tran [17] contains several detailed examples of deriving a system of equations based on physical laws, performing an experiment that exemplifies the physical behavior, and analyzing how well the model describes the experimental data using different mathematical and statistical tools. The examples include heat conduction through a metal rod, sound wave propagation in an enclosed space, and size-structured population dynamics.

Inverse problem methodologies are discussed in [11] and earlier in [13] in the context of dynamical system or mathematical model parameter estimation when a sufficient number of observations of one or more states (variables) are available. The choice of method depends on assumptions the modeler makes on the form of the error between the model and the observations (the statistical model). The most prevalent source of error is observation error, which is made when collecting data. (One can also consider model error, which originates from the differences between the model and the underlying process that the model describes. But this is often quite difficult to quantify.) Measurement error is most readily discussed in the context of statistical models. The three techniques commonly addressed are maximum likelihood estimators (MLE), used when the properties of the error distribution are known; ordinary least squares (OLS), for error with constant variance across observations; and generalized least squares (GLS), used when

the variance of the data can be expressed as a nonconstant function. Uncertainty quantification is also described for optimization problems of this type, namely in the form of observation error covariances, standard errors, residual plots, and sensitivity matrices. Techniques to approximate the variance of the error are also included in these discussions.

Experimental design using the Fisher Information Matrix (FIM), which is based on sensitivity matrices, is described in [12] for the case of scalar data. Sensitivity matrices are composed of functions that relate the change in a variable to the change in the parameter. The first order quantifications of these relations are called traditional sensitivity functions and are useful in suggesting when a variable should be sampled to get the most information for estimating a particular parameter, especially when the first order sensitivity functions are used in conjunction with the so-called second order sensitivity functions. This work also examines the usefulness of generalized sensitivity functions [45], which are calculated using the FIM, that are known to describe how information about the parameters is distributed across time for each variable. Both types of sensitivity functions are then used in [12] in numerical simulations to determine the optimal final time for an experiment of a process described by a logistic curve.

In [16], the authors develop an experimental design theory using the FIM to identify optimal sampling times for experiments on physical processes (modeled by an ODE system) in which scalar or vector data will be taken. The experimental design technique developed is applied in numerical simulations to the logistic curve, a simple ODE model describing glucose regulation and a harmonic oscillator example.

The use of mathematical models to better understand photosynthesis is featured in [6] and [14]. Both works focus on applying the C3 cycle model of [49] to help determine a sampling regimen when performing an experiment, a topic of interest in the mathematical field of experimental design. Multiple methods of selecting the optimal metabolites to measure in a cellular-level experimental setting are discussed and evaluated in [6]. One method, which is an ad-hoc statistical method, determines which variables directly influence an output of interest at any one particular time. A model using a subset of variables is determined via multivariate linear regression, and the efficacy of the model is then measured using the Akaike Information Criterion. The variables that appear in the best models at the most time points are identified as the most important to measure. Such a method does not utilize the information on the underlying time-varying processes given by the dynamical system model. The second method, which utilizes the dynamics of the ODE model to determine what metabolites are most related to parameters via a measure of parameter variability known as the Fisher Information Matrix, is expanded upon and used [14] to include both metabolite and sampling time selection. Extension of the second method first suggested in [6], based optimal design ideas, is the subject of our experimental design algorithm development.

We use the C3 cycle model of [49] and light reaction model of [28], along with information

from [4], [26], and [39], and environmental data, to construct an environment-dependent model of metabolism in plants including the light and dark reactions of photosynthesis. The carbon uptake rate A as predicted by the model is used to relate photosynthetic productivity with plant leaf area as predicted by a logistic growth model. The model is fit to leaf area data of *Arabidopsis thaliana* [35] and soybean [39] in separate numerical experiments to judge the model's ability to describe growth patterns in C3 species whose growth behavior and enzyme kinetics have been largely quantified. While our current efforts compare the modeled leaf area to leaf area data, our hope is to be able to use the experimental design methodologies of [6] and [14] with this proposed model in order to better design greenhouse and field studies of crop performance based on mathematical predictions of overall plant performance under different environmental conditions.

1.2   Proposed models

While many existing models either describe the macro-scale productivity of a plant via leaf area, biomass, or yield predictions or the micro-scale productivity of photosynthesis and CO_2 uptake rates, very few – if any – relate cellular productivity to gains in biomass using a mathematical framework. The goal of our proposed model, which utilizes existing work by [26], [28], [39], and [49], is to accurately represent metabolic processes, in particular C3 photosynthesis, in leaf tissue so that it may be used to predict photosynthetic productivity under varying environmental conditions and reaction rates and relate this productivity to change in plant size. In order to distinguish between changes in productivity due to the environment and those due to genetic regulation and variability, the model must also include mechanisms that reflect how external factors such as sunlight, temperature, and water availability affect photosynthetic rates and plant development in addition to the Michaelis-Menten [29] mechanics that describe cellular processes that fuel plant growth. Our initial conceptual model framework is pictured in Figure 1.4.

1.2.1   Comprehensive model

Our initial efforts focused on only modeling the individual-level dynamics of leaf area (Figure 1.4, right column). Different formulations of sigmoid functions, including the logistic equation and the Gompertz equation, were tested for their ability to fit leaf area data from [35] and [39] in a least-squares minimization parameter estimation problem. Based on the resulting fits, we propose that a dual-logistic function with temperature influence be used to describe leaf area growth and senescence:

$$
\frac{dL}{dt}(t;\vec{\theta},\vec{E}(t)) = \left(\frac{dG}{d\tau}\frac{A(t;\vec{\theta},\vec{E}(t))}{A_{opt}} - \frac{dS}{d\tau}\right)\left(\frac{d\tau}{dT}\frac{dT}{dt}\right), \quad L_0 = G(0) + S(0)
$$

$$
\frac{dG}{d\tau}(t;\vec{\theta},\vec{E}(t)) = r_G G(\tau)\left(1 - G(\tau)/k_G\right), \quad G_0 = G(0) \tag{1.1}
$$

$$
\frac{dS}{d\tau}(t;\vec{\theta},\vec{E}(t)) = r_S S(\tau)\left(1 - S(\tau)/k_S\right), \quad S_0 = S(0)
$$

$$
\frac{d\tau}{dT}(t;\vec{\theta},\vec{E}(t)) = \begin{cases} \frac{2(T(t)-T_{min})^\alpha (T_{opt}-T_{min})^\alpha - (T(t)-T_{min})^{2\alpha}}{(T_{opt}-T_{min})^{2\alpha}}, & T_{min} \le T(t) \le T_{max} \\ 0, & T(t) < T_{min} \quad \text{or} \quad T(t) > T_{max}, \end{cases} \tag{1.2}
$$

where $L((t;\vec{\theta},\vec{E}(t))$ is the total leaf area of the plant, $\vec{E}(t)$ is a vector of environmental forcing functions, $G(t;\vec{\theta},\vec{E}(t))$ is the expanding leaf area, $S(t;\vec{\theta},\vec{E}(t))$ is the senescing plant leaf area, and $A(t;\vec{\theta},\vec{E}(t))$ is the CO_2 uptake rate based on the enzyme kinetics reactions (held constant at $A(t;\vec{\theta},\vec{E}(t))/A_{opt} = 1$ in the exploratory phase of model development). Explanations of the parameters are available in Section A.2 in the appendix. The model includes a mechanism $\tau(T(t))$ as described by (1.2), taken from [39], that describes the influence of temperature on growth and senescence rates; however, when this model was fit to the LAI data found in [39] of two different plantings of the same type of spinach, the growth and senescence parameters of the models best describing these data sets changed by $10 - 30\%$. This indicates that temperature is not the only factor that introduces variability into plant growth rates – factors such as genetic variability and other environmental conditions may also affect a plant's development.

When environmental information is not considered (or held constant for the duration of the in silico experiment), the metabolite concentrations predicted by the model quickly stabilize. The CO_2 uptake rate quickly (within minutes) achieves a steady state that is not influenced by temperature or captured radiance (see [49] and [6] for solutions of the C3 model). Thus the CO_2 uptake mechanism effectively scales the growth curve of (1.1) by a constant when environmental conditions are not varied. Based on the models of [39], [40], and [49] - as well as basic knowledge about photosynthesis as described earlier in this chapter and information from [4] - we include in our model the external factors of light intensity, temperature (in degrees Celsius), atmospheric CO_2 and O_2 concentrations, and relative H_2O availability (Figure 1.4, left column). These factors are denoted as $\vec{E}(t) = [r(t) \quad T(t) \quad [CO_2] \quad [O_2] \quad w(t)]$ and shown in the left column of 1.4.

Similar to [26] and [49], gas solubility equations are used to calculate the concentration of CO_2 and O_2 in the plant based on atmospheric composition. Using Henry's law for gas solubility, the concentration $[G]$ in mol L^{-1} of a gas G (in this case, CO_2 or O_2) in an aqueous solution
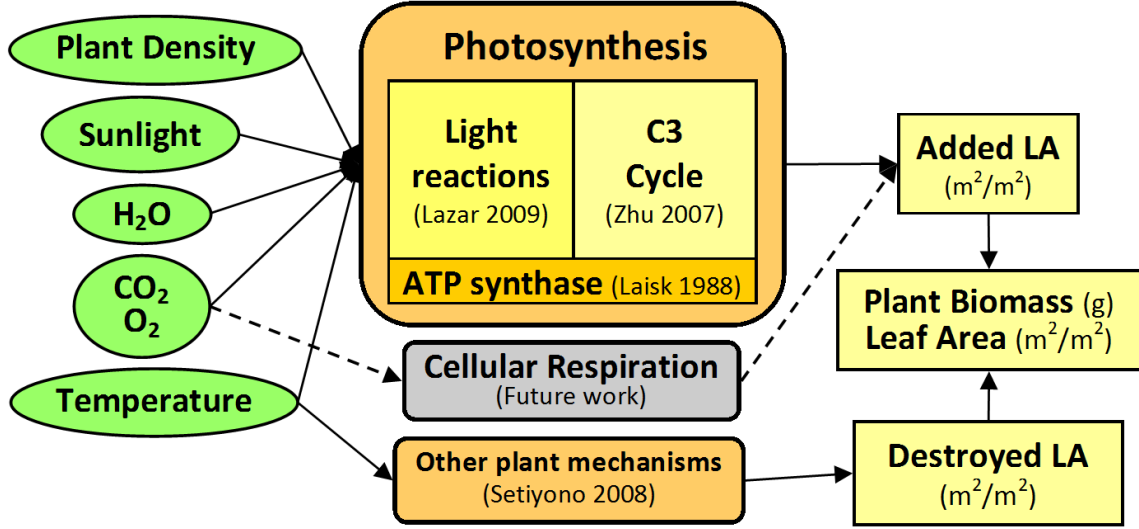
Figure 1.4: Schematic of how multiple existing models are combined for a comprehensive quantification of photosynthesis and then related to environmental factors and biomass change.

(the interior of leaf cell) may be described by

$$[G] = \frac{P(elev)M_G}{K_{HPC}(298.15, G)\exp\left[-C_G\left(\frac{1}{T(t)+273.15} - \frac{1}{298.15}\right)\right]}, \tag{1.3}$$

where $P(elev)$ is the pressure in atmospheric units (atm), M_G is the mole fraction of gas in the gaseous mixture, K_{HPC} and C_G are constants, and $T(t)$ is the temperature in degrees Celsius at time t. Assuming the atmosphere is an ideal gas, $M_{CO_2} = 0.00039$ (390 parts per million) and $M_{O_2} = 0.20946$. The values of the other constants are $K_{HPC}(298.15, CO_2) = 29.41$, $K_{HPC}(298.15, O_2) = 769.23$, $C_{CO_2} = 2400$, and $C_{O_2} = 1700$. We assume that the plant is grown in an environment with sufficient nutrient availability, soil drainage, and pest control so that these factors would not impact plant development.

Reconstructing the model of the light reactions [28] to include environmental factors is straightforward. The parameters k_{L1} and k_{L2}, which describe the rates of light-induced reactions, were changed to $k_{L1}r(t)$ and $k_{L2}r(t)$, where $r(t)$ is the radiance at time t scaled such that $r(t) = 1$ corresponds to 3000 μ mol m^{-2} s^{-1} of photosynthetically active radiation (PAR) and $r(t) = 0$ corresponds to 0 μ mol m^{-2} s^{-1} PAR, so that dynamic sunlight levels may be used to vary the rate of the light reactions. If knowledge about light intensity under the canopy is known but not recorded by above-canopy PAR measurements - or if the PAR sensors are affected by plant growth, we may also include knowledge about plant density by reducing or

13

increasing $r(t)$ when the data is interpolated to form a function.

The reaction implemented to reflect the transition from S_4 to S_0 in the OEC is scaled by $w(t)$, which is relative water availability ($w(t) < 1$ if there is a water shortage and $w(t) = 1$ if there is sufficient or abundant H_2O). As the influence of temperature on the light reactions is not very well understood, the $\tau(T)$ individual-level mechanism is assumed to be sufficient in scaling plant response rates; however, the study of Hill activity in [30] may serve as a useful reference for future changes to the model in this area.

As the intent of [49] was to study the effect of different atmospheric concentrations of CO_2 and O_2 on carbon uptake rate, modifying the C3 cycle and photorespiration models to include environmental conditions is also straightforward. Changes in CO_2 and O_2 concentration due to environmental conditions are calculated by (1.3). H_2O availability affects reactions involving fructose-bisphosphatase (EC 3.1.3.11) and phosphoglycolate phosphatase (EC 3.1.3.18), and so the velocities of these reactions are scaled by water availability $w(t)$. The influence of temperature on the Calvin cycle is also not very well understood, and so we assume the $\tau(T)$ individual-level mechanism and the changes in gas solubility at different temperatures are sufficient mechanisms to describe plant response to temperature during a typical growing season.

Equations from the models of [26], [28], and [49] are used to calculate the CO_2 uptake rate $A(t; \vec{\theta}, \vec{E}(t))$, which is a commonly used measure of photosynthetic productivity. While many of the proposed models in the literature suggest mechanisms that are appropriate for the systems they reflect and describe portions of plant behavior, we choose to build our cellular level enzyme kinetics model using the C3 model of [49] and the light reactions model of [28] because they include the most mechanisms. The models are linked using the photophosphorylation model of [26] to convert ADP to ATP and the production of NADPH from $NADP^+$ by ferredoxin-$NADP^+$-oxidoreductase (which was intentionally neglected in the model of [28] for simplicity). By combining these models we remove the conservation equations for ADP, ATP, $NADP^+$, and NADPH in [49].

The models of PSI, PSII, cyt b6/f, FNR, and the OEC from [28] remain mostly unchanged. The equations of [28] are listed in Sections B.2 and B.3 of the appendix, and parameters as taken from [28] and tuned to enable the model solution to remain physically feasible are listed in Section A.2 in the appendix. Instead of using a four-stage model (using stages S_0 - S_3) of the OEC, we add stage S_4 to expand it to a five-stage model that includes the splitting of H_2O to introduce more electrons and protons (H^+) into the chloroplast. The concentration of H^+ is quantified in this model because it is required for the components of the model from [26]. This requires altering some components of the light reaction model from [28], wherein a plentiful supply of H^+ is assumed. Finally, the generation of NADPH from $NADP^+$ during the oxidation of FNR is added, and the reaction equation is changed to scale with $NADP^+$ concentration so that the reaction will slow as $NAPD^+$ is consumed.

The photophosphorylation mechanism of [26], transport of H^+ across the thylakoid membrane due to the oxidation and reduction of plastoquinone, and changes in $NADP^+$, NADPH, and H^+ in the stroma due to FNR are used to connect the light reactions inside the thylakoid lumen to the dark reactions that occur in the chloroplast stroma. Photophosphorylation is modeled as

$$v_{11}(t; \vec{\theta}, \vec{E}(t)) = \frac{V_{11}\left(x_{11}y_{18} + \frac{x_8}{k_{E0}(x_{88}/x_{89})^{k_{HA}}}\right)}{k_{M11,1}k_{M11,2}\left(1 + \frac{x_{11}}{k_{M11,1}} + \frac{y_{18}}{k_{M11,2}} + \frac{x_8}{k_{M11,3}} + \frac{x_{11}y_{18}}{k_{M0,1}k_{M0,2}}\right)}, \tag{1.4}$$

where variables are defined in Section A.1 and parameters are defined in Section A.2 in the appendix.

All reaction equations formulated in [49] are used to model fructose and sucrose synthesis as well as photorespiration except for the reaction that approximates photophosphorylation (reaction 16 in [49]), which is replaced by (1.4). Additionally, we continue to use the balance equations of pentose phosphates, hexose phosphates, and inorganic phosphates to reduce the number of states for which we must calculate via differential equations. Like [49], the current model also holds the concentration of combined UDP and UTP constant because UTP is generated in the reactions of cellular respiration (which are not included in our model). Variables, parameters, and equations from [49] used in the current model may be found in Sections A.1, A.2, B.1, and B.3 in the appendix.

1.2.2 Compact light reaction model

As the dynamics of the light reactions are difficult to measure, the fine details included in the model of [28] would be difficult to confirm using data. Additionally, because of its complexity, the model of [28] adds a significant amount of time to the calculation of one forward solution to the model as compared to solving the C3 model of [49] alone. Therefore, replacing this model with a simpler mathematical description of the light reactions (as depicted in Figure 1.5, center column) to reduce computational time would not negatively impact our ability to use available metabolite data in parameter estimation problems.

Based on numerical experimentation with the comprehensive model, we alter the conceptual model (Figure 1.4) and temperature response functions used in this model. The temperature response function $\tau(T(t))$ used to describe temperature's response to both growth and senescence in the comprehensive model is changed to $\tau_G(T(t))$, that is, a temperature response function for growth only. A separate function for describing the dependence of senescence on temperature, $\tau_G(T(t))$, is introduced. This senescence temperature response function is also from [39]. The set of equations describing change in leaf area and temperature response for this alternate compact model is

$$\frac{dL}{dt} = \frac{dG}{dt}\frac{A(t;\vec{\theta},\vec{E}(t))}{A_{opt}} - \frac{dS}{dt}, \quad L_0 = L(0)$$

$$\frac{dG}{dt} = r_G G(\tau_G(T(t)))\left(1 - G(\tau_G(T(t)))/k_G\right)\left(\frac{d\tau_G}{dT}\frac{dT}{dt}\right), \quad G_0 = G(0)$$

$$\frac{dS}{dt} = r_S S(\tau_S(T(t)))\left(1 - S(\tau_S(T(t)))/k_S\right)\left(\frac{d\tau_S}{dT}\frac{dT}{dt}\right), \quad S_0 = S(0) \tag{1.5}$$

$$\frac{d\tau_G}{dT}\frac{dT}{dt} = \begin{cases} \frac{2(T(t)-T_{min})^\alpha(T_{opt}-T_{min})^\alpha-(T(t)-T_{min})^{2\alpha}}{(T_{opt}-T_{min})^{2\alpha}}, & T_{min} \leq T \leq T_{max} \\ 0, & T < T_{min} \quad \text{or} \quad T > T_{max} \end{cases}$$

$$\frac{d\tau_S}{dT}\frac{dT}{dt} = \begin{cases} 0, & T(t) \leq T_{min} \\ (T - T_{min})/T_{opt}, & T(t) > T_{min}. \end{cases} \tag{1.6}$$
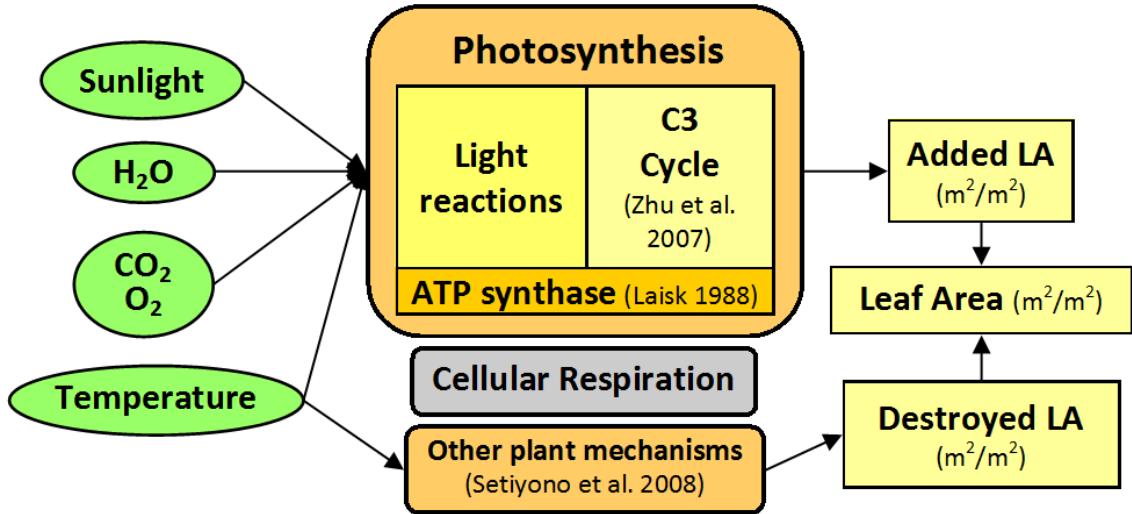
Figure 1.5: Schematic of how existing models of the Calvin Cycle and photophosphorylation are combined with a new, simple model of the light reactions and then related to environmental factors and biomass change.

The first-order reaction equations for the light reactions formulated in [28] and used in the comprehensive model are entirely replaced with four Michaelis-Menten chemical kinetic equations describing this set of chemical summary equations describing reactions numbered 1.10.3.9, 1.10.9.1, 1.97.1.12, and 1.18.1.2 in the MetaCyc Encyclopedia of Metabolic Pathways

[42]:

$$2H_2O + 2PQ \quad \rightarrow \quad O_2 + 2PQH_2 \tag{1.7}$$

$$PQH_2 + 2PC^+ + 2H_S^+ \quad \leftrightarrow \quad 2PC + PQ + 4H_L^+ \tag{1.8}$$

$$PC + Fd \quad \rightarrow \quad PC^+ + Fd^- \tag{1.9}$$

$$Fd^- + H_S^+ + NADP \quad \leftrightarrow \quad 2Fd + NADPH \tag{1.10}$$
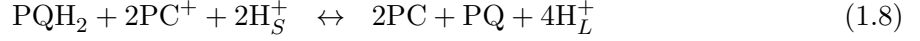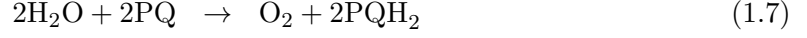
These summary equations correspond to the OEC and PS II (1.7), PC and cyt b6f (1.8), PS1 (1.9), and FNR and Fd (1.10). Reversibility of (1.8) and (1.10) were confirmed by information stored in BRENDA [37], an enzyme database. The corresponding Michaelis-Menten type equations are

$$v_{37}(x, t, \vec{\theta}, \vec{E}(t)) = \frac{V_{37} \cdot PQ \cdot r(t) \cdot w(t)}{k_{m37} + PQ} \tag{1.11}$$

$$v_{38}(x, t, \vec{\theta}, \vec{E}(t)) = \frac{V_{38} k_{m381}^{-1} k_{m382}^{-1} \left(PQH \cdot PC^+ - \frac{PC \cdot PQ}{k_{e38} \left(\frac{H_S^+}{H_L^+} \right)^2} \right)}{1 + \frac{PQH_2}{k_{m381}} + \frac{PC^+}{k_{m382}} + \frac{PC}{k_{m383}} + \frac{PQ}{k_{m384}} + \frac{PQH \cdot PC^+}{k_{m381} k_{m382}} + \frac{PC \cdot PQ}{k_{m383} k_{m384}}} \tag{1.12}$$

$$v_{39}(x, t, \vec{\theta}, \vec{E}(t)) = \frac{V_{39} \cdot PC \cdot Fd \cdot r(t)}{(k_{m391} + PC)(k_{m392} + Fd)} \tag{1.13}$$

$$v_{40}(x, t, \vec{\theta}, \vec{E}(t)) = \frac{V_{40} k_{m401}^{-1} k_{m402}^{-1} (Fd^- \cdot NADP - Fd \cdot NADPH / k_{e40})}{1 + \frac{Fd^-}{k_{m401}} + \frac{NADP}{k_{m402}} + \frac{Fd}{k_{m403}} + \frac{NADPH}{k_{m404}} + \frac{Fd^- \cdot NADP}{k_{m401} k_{m402}} + \frac{Fd \cdot NADPH}{k_{m403} k_{m404}}} \tag{1.14}$$

where V_{37}, V_{38}, V_{39}, and V_{40} are the reaction maximum velocities, k_{e38} and k_{e40} are equilibrium constants (k_{e38} is modulated with a proton gradient-type mechanism much like the mathematical description of ATP synthase [26]), and all k_m's are Michaelis Menten constants.

All other model components remain unchanged between the two models. The compact model, descriptions of all state variables and parameters, values of parameters and initial values of state variables (values at time $t = 0$), equations from [49], and equations from [28] are included in Appendices B.4, A.1, A.2, B.1, and B.2, respectively. Parameter estimation for these two models and sensitivity analysis on select parameters in these models using the methods described in Chapter 2 are performed in Chapter 3.

17

Chapter 2

Experimental Design for Vector Output Systems

2.1 Introduction

In many scientific fields where mathematical modeling is utilized, mathematical models grow increasingly complex, containing possibly more state variables and parameters, over time as the underlying governing processes of a system are better understood and refinements in mechanisms are considered. Additionally, as technology invents and improves devices to measure physical and biological phenomena, new data become available to inform mathematical modeling efforts. The world is approaching an era in which the vast amounts of information available to researchers may be overwhelming or even counterproductive to efforts. We explore a framework based on the Fisher Information Matrix (FIM) for a system of ordinary differential equations (ODEs) to determine when an experimenter should take samples and what variables to measure when collecting information on a physical or biological process that is modeled by a dynamical system.

Building on the modeling and experimental design theory described in Chapter 1 and efforts in [6], we formulate a previously unexplored optimal design problem to determine not only the optimal sampling variables out of a finite set of possible sampling variables but also the optimal sampling time distribution given a fixed final time. We compare the SE-optimal design introduced in [12] and [16] with the well-known methods of D-optimal and E-optimal design on a six-compartment HIV model [2] and a thirty-one dimensional model of the Calvin Cycle [49]. Such models where there may be a wide range of variables to possibly observe are not only ideal on which to test our proposed methodology, but also are widely encountered in applications.

2.2 Mathematical Background

2.2.1 Mathematical and statistical models

We explore our experimental design questions using a **mathematical model**

$$\frac{d\vec{x}}{dt}(t) = \vec{g}(t, \vec{x}(t; \vec{\theta}), \vec{q}), \quad t \in [t_0, t_f] \tag{2.1}$$
$$\vec{x}(t_0; \vec{\theta}) = \vec{x}_0$$

where $\vec{x}(t; \vec{\theta})$ is the vector of state variables of the system generated using a parameter vector $\vec{\theta} = (\vec{x}_0; \vec{q}) \in \mathbb{R}^p$, $p = m + r$, that contains m initial values and r system parameters listed in $\vec{q}$, $\vec{g}$ is a mapping $\mathbb{R}^{1+m+r} \to \mathbb{R}^m$, $t_0 \geq 0$ is the initial time, and $t_f < \infty$ is the final time. We define an **observation process**

$$\vec{f}(t; \vec{\theta}) = C\vec{x}(t; \vec{\theta}), \tag{2.2}$$

where C is an observation operator that maps $\mathbb{R}^m \to \mathbb{R}^N$, where N is the number of variables observed at a single sampling time. If we were able to observe all states, each measured by a different sampling technique, then $N = m$ and $C = \mathbf{I}^{m \times m}$; however, this is most often not the case because of the impossibility of or the expense in measuring all state variables. In other cases (such as the HIV example below) we may be able to directly observe only combinations of the states.

In order to discuss the amount of uncertainty in parameter estimates, we formulate a **statistical model** [11] of the form

$$\vec{Y}(t) = \vec{f}(t; \vec{\theta}_0) + \vec{\mathcal{E}}(t), \quad t \in [t_0, t_f], \tag{2.3}$$

where $\vec{\theta}_0$ is the hypothesized true values of the unknown parameters and $\vec{\mathcal{E}}$ is a vector random process that represents observation error for the measured variables. We make the standard assumptions:

$$\text{E}(\vec{\mathcal{E}}(t)) = \vec{0}, \quad t \in [t_0, t_f],$$
$$\text{Var}(\vec{\mathcal{E}}(t)) = V_0(t) = \text{diag}(\sigma_{0,1}(t)^2, \sigma_{0,2}(t)^2, \ldots, \sigma_{0,N}(t)^2), \quad t \in [t_0, t_f],$$
$$\text{Cov}(\mathcal{E}_i(t)\mathcal{E}_i(s)) = \sigma_{0,i}(t)^2 \delta(t - s), \quad s, t \in [t_0, t_f],$$
$$\text{Cov}(\mathcal{E}_i(t)\mathcal{E}_j(s)) = 0, \quad i \neq j, \quad s, t \in [t_0, t_f],$$

where $\delta(0) = 1$ and $\delta(t) = 0$ for $t \neq 0$. Realizations of the statistical model (2.3) are written

19

$$\vec{y}(t) = \vec{f}(t; \vec{\theta}_0) + \vec{\epsilon}(t), \quad t \in [t_0, t_f].$$

When collecting experimental data, it is often difficult to take continuous measurements of the observed variables. Instead, we assume that we have n observations at times t_j, $j = 1, \ldots, n$, $t_0 \leq t_1 < t_2 < \ldots < t_n \leq t_f$. We then write the observation process (2.2) as

$$\vec{f}(t_j; \vec{\theta}) = C\vec{x}(t_j; \vec{\theta}), \quad j = 1, 2, \ldots, n, \tag{2.4}$$

the discrete statistical model as

$$\vec{Y}_j = \vec{f}(t_j; \vec{\theta}_0) + \vec{\mathcal{E}}(t_j), \quad j = 1, 2, \ldots, n, \tag{2.5}$$

and a realization of the discrete statistical model as

$$\vec{y}_j = \vec{f}(t_j; \vec{\theta}_0) + \vec{\epsilon}(t_j), \quad j = 1, 2, \ldots, n.$$

If we were given $\vec{\theta}_0$, we could solve (2.1) for $\vec{x}(t; \vec{\theta}_0)$, a process known as solving the **forward problem**. Alternatively, if we had a set of data $\vec{y}_j$, $j = 1, 2, \ldots, n$, we could estimate $\vec{\theta}_0$ in a process known as solving the **inverse problem**. We will use this mathematical and statistical framework to develop a methodology to identify sampling variables that provide the most information pertinent to estimating a given set of parameters and the most informative times at which the samples should be taken.

2.2.2 Formulation of the Optimal Design Problem

Several methods exist to solve the inverse problem. A major factor [11] in determining which method to use is additional assumptions made about $\vec{\mathcal{E}}(t)$. It is common practice to make the assumption that realizations of $\vec{\mathcal{E}}(t)$ at particular time points are independent and identically distributed (i.i.d.). If, additionally, the distributions describing the behavior of the components of $\vec{\mathcal{E}}(t)$ are known, then maximum likelihood methods may be used to find an estimate of $\vec{\theta}_0$. On the other hand, if the distributions for $\vec{\mathcal{E}}(t)$ are not known but the variance $V_0(t)$ (also unknown) is assumed to vary over time, weighted least squares methods are often used. We propose an optimal design problem formulation using a generalized weighted least squares criterion.

Let $\mathcal{P}_1([t_0, t_f])$ denote the set of all bounded distributions on the interval $[t_0, t_f]$. We consider the generalized weighted least squares cost functional for systems with vector output

$$J_{WLS}(\vec{y}, \vec{\theta}) = \int_{t_0}^{t_f} [\vec{y}(t) - \vec{f}(t; \vec{\theta})]^T V_0^{-1}(t) [\vec{y}(t) - \vec{f}(t; \vec{\theta})] dP_1(t), \tag{2.6}$$

where $P_1(t) \in \mathcal{P}_1([t_0, t_f])$ is a general measure on the interval $[t_0, t_f]$. For a given continuous data set $\vec{y}(t)$, we search for a parameter $\hat{\theta}$ that minimizes $J_{WLS}(\vec{y}, \vec{\theta})$.

We next consider the case of observations collected at discrete times. If we choose a set of n time points $\tau = \{t_j\}$, $j = 1, 2, \ldots, n$, where $t_0 \leq t_1 < t_2 < \ldots < t_n \leq t_f$ and take

$$P(t) = P_\tau = \sum_{j=1}^{n} \delta_{t_j},$$ (2.7)

where δ_a represents the Dirac delta distribution with atom at a, then the weighted least squares criterion (2.6) for a finite number of observations becomes

$$J_{WLS}^n(\vec{y}, \vec{\theta}) = \sum_{j=1}^{n} [\vec{y}(t_j) - \vec{f}(t_j; \vec{\theta})]^T V_0^{-1}(t_j)[\vec{y}(t_j) - \vec{f}(t_j; \vec{\theta})].$$

To select a useful distribution of time points and set of observation variables, we introduce the N by p sensitivity matrices $\left[\frac{\partial \vec{f}(t;\vec{\theta})}{\partial \vec{\theta}}\right]$ and the m by p sensitivity matrices $\left[\frac{\partial \vec{x}(t;\vec{\theta})}{\partial \vec{\theta}}\right]$ that are determined using the differential operator in row vector form $(\partial_{\theta_1}, \partial_{\theta_2}, \ldots, \partial_{\theta_p})$ represented by $\nabla_{\vec{\theta}}$ and the observation operator defined in (2.2),

$$\begin{aligned}
\nabla_{\vec{\theta}} \vec{f}(t, \vec{\theta}) &= \frac{\partial \vec{f}(t; \vec{\theta})}{\partial \vec{\theta}} \\
&= C \frac{\partial \vec{x}(t; \vec{\theta})}{\partial \vec{\theta}} \\
&= C \begin{pmatrix}
\frac{\partial x_1(t;\vec{\theta})}{\partial \theta_1} & \frac{\partial x_1(t;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial x_1(t;\vec{\theta})}{\partial \theta_p} \\
\frac{\partial x_2(t;\vec{\theta})}{\partial \theta_1} & \frac{\partial x_2(t;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial x_2(t;\vec{\theta})}{\partial \theta_p} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_m(t;\vec{\theta})}{\partial \theta_1} & \frac{\partial x_m(t;\vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial x_m(t;\vec{\theta})}{\partial \theta_p}
\end{pmatrix} \\
&= C \nabla_{\vec{\theta}} \vec{x}(t; \vec{\theta}).
\end{aligned}$$ (2.8)

Using the sensitivity matrix $\nabla_{\vec{\theta}} f(t, \vec{\theta}_0)$, we may formulate the Generalized Fisher Information Matrix (GFIM). Consider the set $\mathcal{C} \subset \mathbb{R}^m$ of admissible observation maps and let $\mathcal{P}_2(\mathcal{C})$ represent the set of all bounded distributions $P_2(c)$ on $\mathcal{C}$. Then the GFIM may be written

$$\begin{aligned}
\mathcal{F}(P_1, P_2, \vec{\theta}_0) &\equiv \int_{t_0}^{t_f} \int_{\mathcal{C}} \frac{1}{\sigma^2(t, c)} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{f}(t; \vec{\theta}_0) \nabla_{\vec{\theta}} \vec{f}(t; \vec{\theta}_0) dP_2(c) dP_1(t) & (2.9) \\
&= \int_{t_0}^{t_f} \int_{\mathcal{C}} \frac{1}{\sigma^2(t, c)} \nabla_{\vec{\theta}}^{\mathsf{T}} \left(c\vec{x}(t; \vec{\theta}_0)\right) \nabla_{\vec{\theta}} \left(c\vec{x}(t; \vec{\theta}_0)\right) dP_2(c) dP_1(t). & (2.10)
\end{aligned}$$

Taking N different sampling maps in $\mathcal{C}$ represented by the m-dimensional row vectors c_k, $k = 1, 2, \ldots, N$, we construct the discrete distribution on $\mathcal{C}$

$$P_C = \sum_{k=1}^{N} \delta_{c_k}, \tag{2.11}$$

where δ_a represents the Dirac delta distribution with atom at a. Using P_C in (2.10), we obtain the GFIM for multiple discrete observation methods taken continuously over $[t_0, t_f]$,

$$
\begin{aligned}
\mathcal{F}(P_1, P_C, \vec{\theta}_0) &= \int_{t_0}^{t_f} \sum_{k=1}^{N} \frac{1}{\sigma^2(t, c_k)} \nabla_{\vec{\theta}}^{\mathsf{T}} \left(c_k \vec{x}(t; \vec{\theta}_0) \right) \nabla_{\vec{\theta}} \left(c_k \vec{x}(t; \vec{\theta}_0) \right) dP_1(t) \\
&= \int_{t_0}^{t_f} \sum_{k=1}^{N} \frac{1}{\sigma^2(t, c_k)} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{x}(t; \vec{\theta}_0) c_k^{\mathsf{T}} c_k \nabla_{\vec{\theta}} \vec{x}(t; \vec{\theta}_0) dP_1(t) \\
&= \int_{t_0}^{t_f} \sum_{k=1}^{N} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{x}(t; \vec{\theta}_0) c_k^{\mathsf{T}} \frac{1}{\sigma^2(t, c_k)} c_k \nabla_{\vec{\theta}} \vec{x}(t; \vec{\theta}_0) dP_1(t) \\
&= \int_{t_0}^{t_f} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{x}(t; \vec{\theta}_0) \sum_{k=1}^{N} \left(c_k^{\mathsf{T}} \frac{1}{\sigma^2(t, c_k)} c_k \right) \nabla_{\vec{\theta}} \vec{x}(t; \vec{\theta}_0) dP_1(t) \\
&= \int_{t_0}^{t_f} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{x}(t; \vec{\theta}_0) \left(C^{\mathsf{T}} V_0^{-1}(t) C \right) \nabla_{\vec{\theta}} \vec{x}(t; \vec{\theta}_0) dP_1(t), \tag{2.12}
\end{aligned}
$$

where $C = (c_1, c_2, \ldots, c_N)^{\mathsf{T}} \in \mathbb{R}^{N \times m}$ is the observation operator in (2.2) and (2.4) and $V_0(t) \in \mathbb{R}^{N \times N}$ is the covariance matrix as described in (2.3). Applying the distribution P_τ as described in (2.7) to the GFIM (2.12) for discrete observation operators measured continuously yields the discrete $p \times p$ Fisher Information Matrix (FIM) for discrete observation operators measured at discrete times

$$F(\tau, C, \vec{\theta}_0) = F(P_\tau, P_C, \vec{\theta}_0) = \sum_{j=1}^{n} \nabla_{\vec{\theta}}^{\mathsf{T}} \vec{x}(t_j; \vec{\theta}_0) C^{\mathsf{T}} V_0^{-1}(t_j) C \nabla_{\vec{\theta}} \vec{x}(t_j; \vec{\theta}_0). \tag{2.13}$$

This describes the amount of information about the p parameters of interest that is captured by the observed quantities described by the sampling maps c_k, $k = 1, 2, \ldots, N$, listed in C, when they are measured at the time points in τ.

The questions of determining the best (in some sense) C and τ are important questions in the optimal design of an experiment. Recall that the set of time points τ has an associated distribution $P_1(\tau) = P_\tau \in \mathcal{P}_1([t_0, t_f])$, where $\mathcal{P}_1([t_0, t_f])$ is the set of all bounded distributions on $[t_0, t_f]$. Similarly, the set of sampling maps $\{c_k\}$ has an associated bounded distribution $P_C \in \mathcal{P}_2(\mathcal{C})$. Define the space of bounded distributions $\mathcal{P}([t_0, t_f] \times \mathcal{C}) = \mathcal{P}_1([t_0, t_f]) \times \mathcal{P}_2(\mathcal{C})$

with elements $P = (P_\tau, P_C) \in \mathcal{P}$. Without loss of generality, assume that $\mathcal{C} \subset \mathbb{R}^m$ is closed and bounded, and assume that there exists a functional $\mathcal{J} : \mathbb{R}^{p \times p} \to \mathbb{R}^+$ of the GFIM (2.10). Then the **optimal design problem** associated with $\mathcal{J}$ is selecting a distribution $\hat{P} \in \mathcal{P}$ such that

$$\mathcal{J}\left(\mathcal{F}(\hat{P}, \vec{\theta}_0)\right) = \min_{P \in \mathcal{P}} \mathcal{J}\left(\mathcal{F}(P, \vec{\theta}_0)\right), \tag{2.14}$$

where $\mathcal{J}$ depends continuously on the elements of $\mathcal{F}(P, \vec{\theta}_0)$.

The Prohorov metric [34], which is a metric designed for use in the weak star topology, provides a basis for a general theoretical framework for the existence of $\hat{P}$ and approximation in $\mathcal{P}([t_0, t_f] \times \mathcal{C})$ (a general theoretical framework is developed in [8, 12]). The application of the Prohorov metric to optimal design problems formulated as (2.14) is explained more fully in [12]: briefly, define the Prohorov metric ρ on the space $\mathcal{P}([t_0, t_f] \times \mathcal{C})$, and consider the metric space $(\mathcal{P}([t_0, t_f] \times \mathcal{C}), \rho)$. Since $[t_0, t_f] \times \mathcal{C}$ is compact, $(\mathcal{P}([t_0, t_f] \times \mathcal{C}), \rho)$ is also compact. Additionally, by the properties of the Prohorov metric, $(\mathcal{P}([t_0, t_f] \times \mathcal{C}), \rho)$ is complete and separable. Therefore an optimal distribution $\hat{P}$ exists and may be approximated by a discrete distribution.

The formulation of the cost functional (2.14) may take many forms. We focus on the use of traditional optimal design methods, D-optimal, E-optimal, or SE-optimal design criteria, to determine the form of $\mathcal{J}$. Each of these design criteria are functions of the inverse of the FIM (assumed hereafter to be invertible) defined in (2.13).

In D-optimal design, the cost functional is written

$$\mathcal{J}_D(F) = \det\left(F(\tau, C, \vec{\theta}_0)^{-1}\right) = \frac{1}{\det\left(F(\tau, C, \vec{\theta}_0)\right)}.$$

By minimizing $\mathcal{J}_D$, we minimize the volume of the confidence interval ellipsoid describing the uncertainty in our parameter estimates. Since F is symmetric and positive semi-definite, $\mathcal{J}_D(F) \geq 0$. Additionally, since F is assumed invertible, $\mathcal{J}_D(F) \neq 0$, therefore, $\mathcal{J}_D : \mathbb{R}^{p \times p} \to \mathbb{R}^+$.

In E-optimal design, the cost functional is $\mathcal{J}_E$ is the largest eigenvalue of $\left(F(\tau, C, \vec{\theta}_0)\right)^{-1}$, or equivalently

$$\mathcal{J}_E(F) = \max \frac{1}{\operatorname{eig}\left(F(\tau, C, \vec{\theta}_0)\right)}.$$

To obtain a smaller standard error, we must reduce the length of the principal axis of the confidence interval ellipsoid. Since an eigenvalue λ solves $\det(F - \lambda I) = 0$, an eigenvalue of $\lambda = 0$ would mean $\det(F) = 0$, or that F is not invertible. Since F is positive definite, all eigenvalues are therefore positive. Thus $\mathcal{J}_E : \mathbb{R}^{p \times p} \to \mathbb{R}^+$.

In SE-optimal design, $\mathcal{J}_{SE}$ is a sum of the elements on the diagonal of $\left(F(\tau, C, \vec{\theta}_0)\right)^{-1}$

weighted by the respective parameter values [12, 16], written

$$\mathcal{J}_{SE}(F) = \sum_{i=1}^{p} \frac{\left(F(\tau, C, \vec{\theta}_0)) \right)^{-1}_{i,i}}{\theta_{0,i}^2}.$$

Thus in SE-optimal design, the goal is to minimize the sum of squared errors of the parameters normalized by the true parameter values. As the diagonal elements of F^{-1} are all positive and all parameters are assumed non-zero in $\vec{\theta} \in \mathbb{R}^p$, $\mathcal{J}_{SE} : \mathbb{R}^{p \times p} \to \mathbb{R}^+$.

In [16], it is shown that the D-, E-, and SE-optimal design criteria select different time grids and yield different standard errors. We expect that these design cost functionals will also choose different observation variables (maps) in order to minimize different dimensions of the confidence interval ellipsoid.

2.3 Standard Errors

In order to compare the ability of different optimal design criteria to minimize uncertainty in parameter estimation, we compute the standard errors associated with these parameters. We begin by selecting an ODE system, a nominal set of parameters $\vec{\theta}$ that we would estimate, the start and end times of the experiment t_0 and t_f, the number of sampling times n, and the number of observation maps we wish to use N. After an optimal τ and C are determined according to one of the three previously described optimal design methods, we compute the standard errors for the parameters in $\vec{\theta}$. There are multiple techniques [16] available to compute standard errors; here we choose to use asymptotic theory due to its ease of implementation.

2.3.1 Asymptotic Theory for Standard Errors

If we assume that the covariance matrix $V_0(t)$ is constant over time ($V_0(t) \equiv V_0 = \text{Var}(\vec{\mathcal{E}}(t_j)) = \text{diag}(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,N}^2)$), then we may use an ordinary least squares (OLS) framework to estimate standard errors. Once an optimal $\tau = \{t_j\}_{j=1}^n$, $t_0 \le t_1 < t_2 < \ldots < t_n \le t_f$, and C are determined, we obtain data from an experiment or simulate data $\{y_j\}$ as a realization of the random process $\{Y_j\}$ described in (2.5), and then we estimate the parameters in $\vec{\theta}$ by solving the inverse problem using the OLS criterion [11]. The discrete OLS estimator is defined as

$$\vec{\theta}_{\text{OLS}} = \arg \min_{\vec{\theta} \in \Theta} \sum_{j=1}^{n} \left[\vec{Y}_j - \vec{f}(t_j; \vec{\theta}) \right]^T V_0^{-1} \left[\vec{Y}_j - \vec{f}(t_j; \vec{\theta}) \right], \tag{2.15}$$

where Θ is the set of all possible values of $\vec{\theta}$, such that each element of the difference vector $\vec{Y}_j - \vec{f}(t_j; \vec{\theta})$ is weighted using the variance of its corresponding sampling maps. One realization

of this problem using data $\vec{y}_j$, $j = 1, 2, \ldots, n$, is written

$$\hat{\theta}_{\mathrm{OLS}} = \arg\min_{\vec{\theta} \in \Theta} \sum_{j=1}^{n} \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right]^T V_0^{-1} \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right]. \tag{2.16}$$

However, calculating $\hat{\theta}_{\mathrm{OLS}}$ still requires the unknown V_0. If the number of parameters p of a system is sufficiently small and number of observations n large so that $p < n$, then we may calculate the bias adjusted estimate of the variances

$$V_0 \approx \hat{V} = \mathrm{diag}\left(\frac{1}{n-p} \sum_{j=1}^{n} \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right] \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right]^T \right), \tag{2.17}$$

and find the estimate of $\vec{\theta}_0$ using

$$\vec{\theta}_0 \approx \hat{\theta}_{\mathrm{OLS}} = \arg\min_{\vec{\theta} \in \Theta} \sum_{j=1}^{n} \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right]^T \hat{V}^{-1} \left[\vec{y}_j - \vec{f}(t_j; \vec{\theta}) \right]. \tag{2.18}$$

Therefore, finding $\hat{\theta}_{\mathrm{OLS}}$ when V_0 is unknown requires solving the coupled system of equations (2.17) and (2.18).

We may utilize the asymptotic properties of the OLS minimizer (2.15) to learn about the behavior of the model (2.1) and (2.3). As the number of samples $n \to \infty$, $\vec{\theta}_{\mathrm{OLS}}$ has the following properties [11, 19, 44]

$$\vec{\theta}_{\mathrm{OLS}} \sim \mathcal{N}(\vec{\theta}_0, \Sigma_0^n) \approx \mathcal{N}(\hat{\theta}_{\mathrm{OLS}}, \hat{\Sigma}^n),$$

where

$$\Sigma_0^n \approx \left(\sum_{j=1}^{n} \chi_j^T(\vec{\theta}_0) V_0^{-1} \chi_j(\vec{\theta}_0) \right)^{-1} \tag{2.19}$$

is the $p \times p$ covariance matrix, and $\chi_j(\vec{\theta}) = \chi_j^n(\vec{\theta}) = \nabla_{\vec{\theta}} \vec{f}(t_j; \vec{\theta}_0)$ is the $N \times p$ matrix

$$\chi_j(\vec{\theta}) = \chi_j^n(\vec{\theta}) = \begin{pmatrix} \frac{\partial f_1(t_j; \vec{\theta})}{\partial \theta_1} & \frac{\partial f_1(t_j; \vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_1(t_j; \vec{\theta})}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_N(t_j; \vec{\theta})}{\partial \theta_1} & \frac{\partial f_N(t_j; \vec{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_N(t_j; \vec{\theta})}{\partial \theta_p} \end{pmatrix}. \tag{2.20}$$

The approximation $\hat{\Sigma}^n$ to the covariance matrix Σ_0^n is

$$\Sigma_0^n \approx \hat{\Sigma}^n = \left(\sum_{j=1}^{n} \chi_j^T(\hat{\theta}_{\mathrm{OLS}}) \hat{V}^{-1} \chi_j(\hat{\theta}_{\mathrm{OLS}}) \right)^{-1}. \tag{2.21}$$

We may use $\hat{\Sigma}^n$ to approximate the standard errors of each parameter in $\hat{\theta}_{\text{OLS}}$. For the kth element of $\hat{\theta}_{\text{OLS}}$, written $\hat{\theta}_{\text{OLS},k}$, the asymptotic standard error is

$$ASE(\vec{\theta}_{0,k}) = \sqrt{\hat{\Sigma}^n_{0,kk}} \approx ASE(\hat{\theta}_{\text{OLS},k}) = \sqrt{\hat{\Sigma}^n_{kk}},$$

where $\hat{\Sigma}^n_{0,kk}$ is the element in the kth row and kth column of Σ^n_0, and $\hat{\Sigma}^n_{kk}$ is the corresponding element in $\hat{\Sigma}^n$. Additionally, since the FIM is defined to be the inverse of the covariance matrix, we may approximate the FIM using (2.21) by $F(\tau, C, \vec{\theta}_0) \approx F(\hat{\tau}, \hat{C}, \hat{\theta}_{\text{OLS}}) = (\hat{\Sigma}^n)^{-1}$.

2.3.2 Clarification for plant growth model

In performing the parameter estimation problem on the candidate plant growth models, only some parameters – namely, those in (1.1) – must be estimated. Most parameters in the model have been determined experimentally or previously estimated using a variety of methods. Additionally, the models utilize five environmental forcing functions. The mathematical model would then be written

$$\begin{aligned}
\frac{d\vec{x}}{dt}(t) &= \vec{g}(t, \vec{x}(t, \vec{E}(t); \vec{\theta}), \vec{\theta}), \quad t \in [t_0, t_f] \\
\vec{x}(t_0, \vec{E}(t_0); \vec{\theta}) &= \vec{x}_0,
\end{aligned} \tag{2.22}$$

where $\vec{x}(t, \vec{E}(t); \vec{\theta})$ is the vector of m state variables of the system generated using environmental forcing functions $\vec{E}(t)$, parameter vector $\vec{\theta} \in \mathbb{R}^p$, $p = m^* + r$, that contains $m^* \leq m$ initial values from $\vec{x}_0$ and r system parameters that we will estimate listed in $\vec{q}$, $\vec{g}()$ is a mapping $\mathbb{R}^{6+m^*+r} \to \mathbb{R}^m$, $t_0 \geq 0$ is the initial time, and $t_f < \infty$ is the final time. The statistical model, observation operator, and WLS cost functional are similar to (2.3) and (2.5), (2.2) and (2.4), and (2.6), respectively, and the process of solving the parameter estimation and experimental design problems are identical.

2.4 Numerical algorithm and optimization constraints

We use MATLAB to solve the system of ODEs and conservation equations and perform parameter estimation. Due to the rapid changes in metabolite concentrations observed when radiance changes, the ODEs are solved using `ODE15s` [38], a variable-order ODE solver in MATLAB for systems of stiff differential equations. Initially, manual tuning of parameters that were estimated (that is, not experimentally determined) in [26], [28], and [49] was performed in order for the daytime steady state of the enzyme kinetics reactions to be physically acceptable (all metabolite concentrations non-negative). Further identification of these parameters is difficult due to the shortage of available metabolite concentration data in the literature. Preliminary values of

parameters and initial conditions added by the individual-level equations were estimated during the exploratory phase of model development for each data set using the MATLAB routine `fminsearch` while $A(t)$ and the environmental conditions in $\vec{E}(t)$ were set to constants that would reflect daytime in a greenhouse setting.

To estimate parameters, we use the MATLAB routines `fminsearch` or, when physical constraints on parameters are known, `fmincon` [18], which searches a specified region of a space for a point that minimizes a given cost functional (in this case (2.15)) to find the OLS minimizer of (2.18) The `fmincon` routine enables us to enforce lower and upper bounds on the parameter and solution spaces, experiment with multiple numerical stepping methods to find a local minimum, and parallelize the exploration of the parameter space via the MATLAB routine `MultiStart`. For this problem we use the interior-point algorithm, which solves a series of minimization problems that approximate the original problem such that the solution of each step in the series approaches the solution to the original problem. The boundary condition on this parameter set is $\vec{\theta} \geq \vec{0}$ so that all initial conditions and reaction constants maintain realistic values and the logistic equations $G(t; \vec{\theta}, \vec{E}(t))$ and $S(t; \vec{\theta}, \vec{E}(t))$ remaining nondecreasing functions. We additionally implement the condition that $\vec{x}(t; \vec{\theta}, \vec{E}) \geq 0$ for all $t \geq 0$ so that the model solution remains relevant to the physical phenomena it describes.

Because the only data used in these simulations is leaf area, corresponding to $L(t)$, we may neglect the calculation of (2.17) and set $V_0 = 1$ to simplify and speed the determination of the optimal parameter vector $\hat{\theta}$. Following the determination of $\hat{\theta}$, we calculate $\hat{\Sigma}^n$ and the standard errors to inform our confidence in how well $\hat{\theta}$ approximates the unknown parameters $\vec{\theta}_0$.

In most optimal design problems, there is not a continuum of measurement possibilities that may be used; rather, there are $N^* < \infty$ possible observation maps c. Denote this set as $\mathcal{C}_{N^*} \subset \mathbb{R}^m$. While we may still use the Prohorov metric-based framework to guarantee existence and convergence of (2.14), we have a stronger result first proposed in [6] that is useful in numerical implementation. Because $\mathcal{C}_{N^*}$ is finite, all probability distributions made from the elements of $\mathcal{C}_{N^*}$ have the form $P_2(\{c_k\}) = P_C = \sum_{k=1}^{N} \delta_{c_k}$ for a fixed N. Moreover, the set of all distributions that use N sampling methods $\mathcal{P}_2^N(\mathcal{C}_{N^*})$ is also finite. For a fixed distribution of time points P_τ, we may compute using (2.13) the set of all possible FIM $F(P_\tau, P_C, \vec{\theta})$ that could be formulated from $c \in \mathcal{C}_{N^*}$. By the properties of matrix multiplication and addition, this set is also finite. Then the functional (2.14) applied to all F in the set produces a finite set contained in $\mathbb{R}^+$. Because this set is finite, it is well-ordered by the relation $\leq$ and therefore has a minimal element. Thus for any distribution of time points P_τ, we may find at least one solution $\hat{P}_C \in \mathcal{P}_2^N(\mathcal{C}_{N^*})$. Moreover, $\hat{P}_C$ may be determined by a search over all matrices $C = (c_1, c_2, \ldots, c_N)^\mathsf{T}$ formed by N elements from $\mathcal{C}_{N^*}$. Therefore, for a fixed P_τ and $N \leq N^* < \infty$ a global optimal set of N sampling methods may be determined.

Due to the computational demands of performing nonlinear optimization for n time points

and N observation maps (for a total of $n + N$ dimensions), and to the difference in techniques between searching for an optimal P_C in the finite set $\mathcal{C}_{N*}$ and searching for an optimal distribution of n sampling times, we instead solve the coupled set of equations

$$\hat{C} = \arg \min_{\{C | P_C \in \mathcal{P}_2^N(\mathcal{C}_{N*})\}} \mathcal{J}\left(F(\hat{\tau}, C, \vec{\theta}_0)\right) \tag{2.23}$$

$$\hat{\tau} = \arg \min_{\{\tau | P_\tau \in \mathcal{P}_1([t_0, t_f])\}} \mathcal{J}\left(F(\tau, \hat{C}, \vec{\theta}_0)\right), \tag{2.24}$$

where $C \in \mathbb{R}^{N \times m}$ represents a set of N sampling maps and $\tau = \{t_j\}_{j=1}^n$, $t_0 \leq t_1 < t_2 < \ldots < t_n \leq t_f$, is an ordered set of n sampling times. These equations are solved iteratively as

$$\hat{C}_i = \arg \min_{\{C | P_C \in \mathcal{P}_2^N(\mathcal{C}_{N*})\}} \mathcal{J}\left(F(\hat{\tau}_{i-1}, C, \vec{\theta}_0)\right) \tag{2.25}$$

$$\hat{\tau}_i = \arg \min_{\{\tau | P_\tau \in \mathcal{P}_1([t_0, t_f])\}} \mathcal{J}\left(F(\tau, \hat{C}_i, \vec{\theta}_0)\right), \tag{2.26}$$

where $\mathcal{J}$ is the D-, E-, or SE-optimal design criterion. We begin by solving for $\hat{C}_1$ where $\hat{\tau}_0$ is specified by the user. The system (2.25)-(2.26) is solved until the requirement

$$|\mathcal{J}\left(F(\hat{\tau}_i, \hat{C}_i, \vec{\theta}_0)\right) - \mathcal{J}\left(F(\hat{\tau}_{i-1}, \hat{C}_{i-1}, \vec{\theta}_0)\right)| < \epsilon$$

is met or until $\hat{C}_i = \hat{C}_{i-1}$. For each iteration, (2.25) is solved using a global search over all possible C. Since the sensitivity equations cannot be easily solved for in the models chosen to illustrate our method, we use a modified version of *tssolve.m* by A. Attarian [5], which implements the *myAD* package developed by M. Fink [21].

Solving (2.26) requires using a nonlinear constrained iterative optimization algorithm. While MATLAB's *fmincon* is a natural choice for such problems, as reported in [16], it does not perform well in this situation. Instead, we use *SolvOpt* developed by A. Kuntsevich and F. Kappel [24] (which utilizes a modified version of Shor's r-algorithm) to search for an optimal distribution local to an initial uniformly spaced time point distribution. There exist many types of constraints that may be placed on this optimization. The four constraints used in [16] are

(C1) Optimize all time points such that $t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t_f$. This method optimizes n time points.

(C2) The initial and final time points are fixed as $t_1 = t_0$ and $t_n = t_f$. The routine then optimizes over the remaining $n - 2$ time points such that $t_j \leq t_{j+1}$.

(C3) Optimize the time steps $\nu_j \geq 0$. Fix $t_1 = t_0$ and $t_n = t_f$, and the remaining time points may be found by $t_{j+1} = t_j + \nu_j$, $j = 1, 2, \ldots, n - 2$. Additionally, $t_f - t_0 \geq \sum_{j=1}^{n-2} \nu_j$. This

routine also optimizes over $n - 2$ variables.

(C4) Optimize the time steps $\nu_j \geq 0$. Fix $t_1 = t_0$. The remaining time points may be found by $t_{j+1} = t_j + \nu_j$, $j = 1, 2, \ldots, n - 1$ such that $t_f - t_0 = \sum_{j=1}^{n-1} \nu_j$. This method optimizes over $n - 1$ variables.

We select for our use two constraints, (C2) and (C3), to reduce the complexity of the problem of time point distribution selection. Once either of the convergence requirements are met and $\hat{C}$ and $\hat{\tau}$ are determined, we compute standard errors using the asymptotic theory described in the next section.

Chapter 3

Numerical Results

We illustrate the potential of the mathematical concepts explained in Chapter 2 using a variety of models. Parameter estimation is performed using the C3 cycle model of [49] on simulated data generated using parameter values found in [49] and using the comprehensive and compact models described in Chapter 1 on an *Arabidopsis thaliana* leaf area data set from [35] and a soybean leaf area data set from [39]. Sensitivities for the maximum velocity parameters (listed in Appendix A.2) over one day of the compact model solution with parameters that best fit the *Arabidopsis thaliana* leaf area data [35] are calculated in order to explore the relationship between the model and some parameters. In order to show proof of concept, the experimental design framework is applied to a HIV model [1] and the C3 cycle model [49].

3.1 Parameter estimation

3.1.1 C3 Model of [49]

We test the feasibility of predicting a small number of parameters in the Zhu Calvin Cycle model [49]. Since we do not have a set of data, we use the simulated solution to this model, with parameter vector θ_0 with values as described in [49], to calculate eleven data points of all 38 metabolites at times $t_j = \frac{j-1}{10}5000$, $j = 1, 2, \ldots, 11$, so that $T_f = 5000$ seconds. To perform the inverse problem, we add or subtract 25% of the true value of the parameters we will estimate and then use unconstrained nonlinear optimization via a simplex search to minimize the unweighted least squares error between the predicted solution and the true solution that uses θ_0.

We perform the inverse problem for the parameter subsets

$$\theta_{0a} = [KM11, Km521, KI523, KC, Km1221, Km1241],$$

$$\theta_{0b} = [KM11, KM12, KM13, KE4, KM131, KE22, Km511, Km521,$$
$$KI523, KC, Km1221, Km1241, V58],$$

$$\theta_{0c} = [RuBP_0, SBP_0, KM11, KM12, KM13, KI13, KE4, KM9, KM131, KI135, KE22,$$
$$Km511, Km514, Km521, KI523, KC, Km1221, Km1241, V9, V58, V111],$$

record the parameter estimates $\hat{\theta}_a^{\pm}$, $\hat{\theta}_b^{\pm}$, and $\hat{\theta}_c^{\pm}$, respectively, for the initial parameter guesses at $1 \pm .25$ times the parameter subset true values, and plot the CO_2 uptake rates for the true solution and the solution generated by the parameter estimate. We also record the least squares error and the runtime for the inverse problem.

Estimating six parameters provided mixed results when only 25% away from the true value. Starting at 25% below θ_0, the minimization routine was able to estimate KM11, KI523, and KC very well; however, starting at 25% above θ_0 lead to an estimate $\hat{\theta}_a^+$ that is not near θ_0 (Table 3.1). This is reflected in the poor approximation of the carbon uptake rate pictured in Figure 3.1. The variability in results may be due to the capabilities of `fminsearch` or undetected interactions between parameters. For example, when estimating Km521, perhaps KI521 should also be estimated. When estimating the top 13, all parameters could be estimated reasonably well with the exception of KM12, which was far from its true value in both $\hat{\theta}_b^-$ and $\hat{\theta}_b^+$, KM521, KC, and KM1221. The CO_2 uptake rates of the models generated by $\hat{\theta}_b^-$ and $\hat{\theta}_b^+$, however, appear to be very close to that of the model generated by θ_0 (Figure 3.2). Adding parameter constraints would aid in producing biologically correct parameter estimates. Estimating 21 parameters (the top 13 plus 8 more parameters used in describing reactions that involve RuBP and SBP, which are seen to be related to Rubisco in [49]) was surprisingly effective, with only 6 values in $\hat{\theta}_c^-$ and 2 in $\hat{\theta}_c^+$ being far from their true values (Table 3.1). The performance of the model generated by $\hat{\theta}_c^+$ is also impressive (Figure 3.3). Since some of these poorly estimated parameters are also problematic in the 13-parameter case, we may want to refocus our efforts on other parameters or perform more refined sensitivity analysis.

It appears that if the runtime of the parameter estimation problem is a concern, it would be best to estimate no more than 10 parameters; however, the results are sometimes far from the true parameters. Having correct parameter estimates in this context is especially important, as different reaction rates would paint a different picture of how the plant is performing. It is possible to estimate a large number of parameters (over 20) if reasonable initial guesses can be provided. Additionally, the number of parameters that could be estimated - and the quality of those estimates - could be improved if constraints were imposed.

Table 3.1: θ_0 and parameter estimates for top 13 parameters

Name	θ_0 **value**	$\hat{\theta}_a^-$	$\hat{\theta}_a^+$	$\hat{\theta}_b^-$	$\hat{\theta}_b^+$	$\hat{\theta}_c^-$	$\hat{\theta}_c^+$
KM11	0.0115	0.0115	0.0108	0.009	0.0139	0.0081	0.0140
KM12	0.222	-	-	0.1447	0.4354	0.1275	0.377
KM13	0.02	-	-	0.198	0.0298	0.0170	0.0276
KE4	0.05	-	-	0.0502	0.05	0.049	0.0502
KM131	0.05	-	-	0.0492	0.0499	0.0354	0.0561
KE22	0.058	-	-	0.0593	0.0577	0.0547	0.0581
KM511	0.02	-	-	0.0194	0.02	0.0203	0.0201
KM521	0.0025	0.0201	0.0097	5.5e-4	0.0022	4e-5	0.0031
KI523	7e-5	7.1e-5	2.2e-5	2.7e-5	6.2e-5	1e-6	8.5e-5
KC	0.0115	0.0114	0.0425	0.0114	0.0116	0.0062	0.0021
KM1221	0.15	0.0813	0.0023	0.1247	0.2096	0.1007	0.1894
KM1241	0.15	0.0894	0.0236	0.1276	0.2027	0.1064	0.1848
V58	0.0168	-	-	0.0098	0.0168	0.0192	0.0174
Runtime(sec)		282.09	362.80	854.20	640.53	1940.63	2115.75
LSQ Error		0.1429	88.98	0.9895	0.02649	0.1700	0.04131
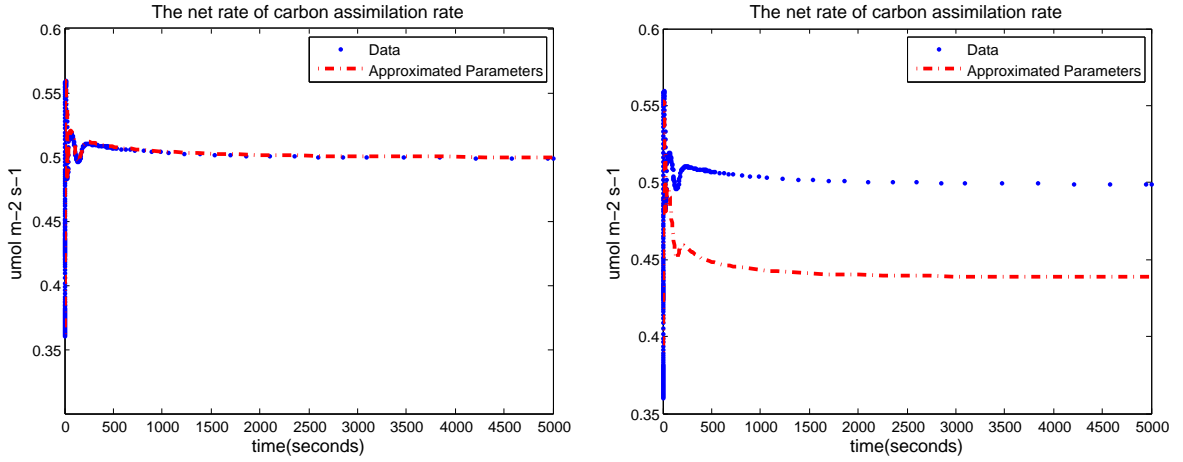
Figure 3.1: CO_2 uptake rates for θ_0(blue dots) and estimates $\hat{\theta}_a^-$ (left, red dash) and $\hat{\theta}_a^+$ (right, red dash).
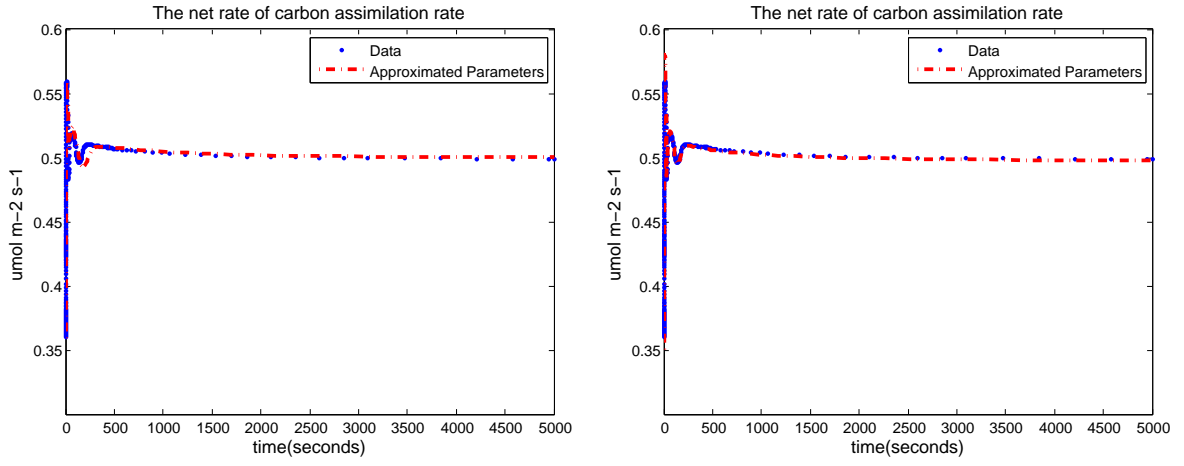
Figure 3.2: CO_2 uptake rates for θ_0(blue dots) and estimates $\hat{\theta}_b^-$ (left, red dash) and $\hat{\theta}_b^+$ (right, red dash).
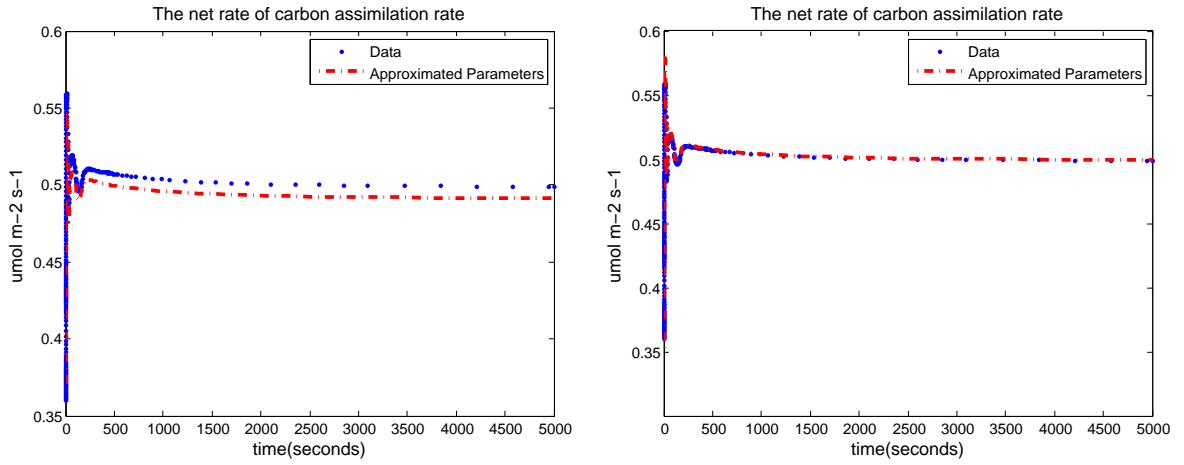
Figure 3.3: CO_2 uptake rates for θ_0(blue dots) and estimates $\hat{\theta}_c^-$ (left, red dash) and $\hat{\theta}_c^+$ (right, red dash).

3.1.2 Modeling plant growth using the proposed models

While the inverse problem experiments performed on the C3 cycle model [49] exhibit the potential of the parameter estimation method described in Chapter 2, all the parameters in the C3 cycle model [49] are known via experimentation or estimation and listed in [49]. Next we exhibit the ability of this method to identify parameters that enable the comprehensive model, which contains (1.1) and the equations of sections B.1, B.2, and B.3, and the compact model, which is constructed using (1.5), (1.11)–(1.14), and the equations of sections B.1 and B.4, to best approximate a data set. Radiance $r(t)$ is estimated using a smoothed square wave such that $r(t) = 0.5$ between 9:00 and 15:00 and $r(t) = 0$ between 21:00 and 3:00. Water availability is set to $w(t) = 1$.

Modeling *Arabidopsis thaliana* growth

The first data set used to test the performance of the model is of *Arabidopsis thaliana*, taken from [35]. The data set contains 14 time points over 22 days of growth in a greenhouse, capturing the vegetative phase and beginning of the reproductive phase of the species' growth. Temperature in the greenhouse of [35] is approximated by $T(t) = 2.5\sin(2\pi t/86400) + 17.5$, where t is time in seconds.

Using parameter estimates near the values found in [49] and [28] for all parameters held constant, we determine the value of $\vec{\theta}$ that allows the comprehensive and the compact models to best fit the data as described by the OLS minimizer (2.15). Parameter estimates are given in Table 3.2, and leaf area and carbon uptake predictions are pictured in Figures 3.4 and 3.5, respectively. While G_0, r_G, and r_S of the two models are on the same order of magnitude, S_0, k_G, and k_S differ greatly. These differences in parameter values are reflected in the shape of the model leaf area curves seen in Figure 3.4. Some of the differences may also be related to the changes in carbon uptake rates seen in Figure 3.5 due to the changed model mechanisms describing photosynthesis.

Leaf area and carbon uptake also show differences in productivity between day and night (Figures 3.4 and 3.5). While leaf area and carbon increase during the day (when sunlight is present), they exhibit no increase during the night. The daytime carbon uptake rate predicted by the comprehensive model is approximately 300 μ mol m^{2}s^{-1}, about one order of magnitude away from the estimate of approximately 30 μ mol m^{2}s^{-1} from [49], and the daytime carbon uptake rate predicted by the compact model is near 500 μ mol m^{2}s^{-1}. While it would be expected for a plant to exhibit a decrease in overall carbon absorbed during periods of low light such as night, this model does not illustrate that behavior because it does not include the reactions of cellular respiration. Thus it is reasonable that the model does not exhibit any change in total carbon uptake during the night.

Table 3.2: Estimated $\vec{\theta}$ values for the models fitting *A. thaliana* data [35].

Parameter	Comprehensive	Compact
G_0	1.88e-4	4.09e-4
S_0	9.29e-4	3.7e-12
r_G	6.6e-4	9.86e-4
k_G	21.39	1.26e4
r_S	2.7e-4	3.29e-4
k_S	0.78	8.46e-3
V_{c62}	18.24	-

The concentrations of PC^+, RuBP, T3P, PGA, T3Pc, ATP, and ADP predicted by each model are shown in Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, and 3.12 respectively. All metabolites shown exhibit two concentrations, one concentration during the day and second concentration during the night when the light reactions are occurring at a lower rate. While the concentrations predicted by the two models differ by an order of magnitude - and often less - they do not exactly agree.

The metabolites that are predicted to have a higher concentration during the day by the comprehensive model (PC^+, RuBP, T3P, PGA, T3Pc, and ADP) also are predicted to have higher daytime concentrations by the compact model, while those predicted to have a higher concentration at night by the comprehensive model (ATP) also are predicted to have a higher concentration at night by the compact model. The daytime concentrations of T3P (Figure 3.8) and T3Pc (Figure 3.10) are predicted by the compact model to be an order of magnitude higher than the comprehensive model.

At night, both models exhibit similar concentrations for RuBP, T3P, PGA, T3Pc, ADP, and ATP, metabolites that are heavily influenced by the dark reactions, but the concentration of PC^+ predicted by the compact model remains above 1 mmol l^{-1}, but the prediction by the comprehensive model approaches 0 mmol l^{-1} (Figure 3.6). These differences indicate that replacing the light reaction model of [28] with a simpler model does impact the behavior of other model compartments, especially during the day. At night, the differences between the two models as exhibited by metabolite concentrations are less pronounced.
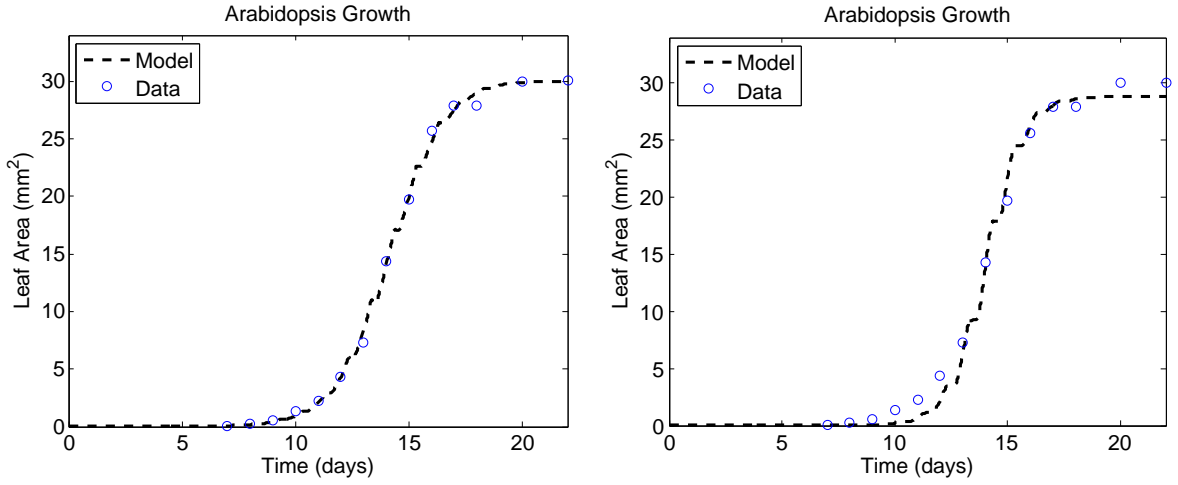
Figure 3.4: **Left:** Leaf area for *A. thaliana* as predicted by the comprehensive model (dashed line) and seen in [35](open circles). **Right:** Leaf area for *A. thaliana* as predicted by the compact model (dashed line) and seen in [35](open circles). Error bars not available for data.

Figure 3.5: **Left:** Carbon uptake as predicted by the comprehensive model (dashed) and by the constant rate of 28 μmol m$^-$2s$^-$1 (solid) as predicted by [49]. **Right:** Carbon uptake as predicted by the compact model (dashed).

Figure 3.6: PC$^+$ concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.7: RuBP concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.8: T3P concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.9: PGA concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.10: Cytosolic T3P concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.11: ATP concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Figure 3.12: ADP concentration in *A. thaliana* as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [35].

Table 3.3: Estimated $\vec{\theta}$ values for models fitting Mead, NE, 2004 soybean data [39]

Parameter	Comprehensive	Compact
G_0	0.0296	0.4101
S_0	0.0061	0.0211
r_G	7.08e-5	1.15e-4
k_G	4.2014	2.82e3
r_S	5.1e-5	5.1747e-5
k_S	6.3624	110.76
V_{c62}	15.0203	-

Modeling soybean growth

The second data set used to evaluate the model is soybean leaf area data taken from the Mead, NE, 2004 sample in [39]. Functions describing the temperature during soybean growth were constructed using interpolation via the `interp1` of weather data (including the high, low, and average temperature) collected from stations near the locations listed in [39]. Cubic spline interpolation is used with the comprehensive model, while linear spline interpolation is used with the compact model to reduce computation times with little impact on the calculated least squares error of the model.

Using parameter estimates near the values found in [49] and [28] for all parameters held constant, we determine the value of $\vec{\theta}$ that allows the models to best fit the data as described by the OLS minimizer (2.15). Parameter estimates are given in Table 3.3, and leaf area for the duration of the experiment and carbon uptake predictions for the first 20 days of the experiment are pictured in Figures 3.13 and 3.14, respectively. The values of S_0, r_G, and r_S vary by less than an order of magnitude between the two models, but G_0, k_G, and k_S vary greatly. These differences in parameter values are reflected in the leaf area predicted by each model, shown with the leaf area data from [39], in Figure 3.13. While the comprehensive model more accurately describes leaf area during the early vegetative phases of soybean growth, the compact model better reflects the reproductive and senescent phases of the soybean plant's development. Some of the differences seen in predicted leaf area may also be related to the changes in carbon uptake rates seen in Figure 3.14 due to the changed model mechanisms describing photosynthesis.

While not apparent Figure 3.13, soybean leaf area and carbon uptake also show differences in productivity between day and night similar to the trends seen in *A. thaliana* leaf area (Figure 3.4). Carbon absorbed by the plant increases during the day (when sunlight is present) but exhibits no increase during the night (Figure 3.14). The daytime carbon uptake rate predicted by the comprehensive model is approximately 300 μ mol m^{2}s^{-1}, about one order of magnitude

away from the estimate of approximately 30 μ mol m^{2}s^{-1} from [49], and the daytime carbon uptake rate predicted by the compact model is near 500 μ mol m^{2}s^{-1}. Again, the model does not exhibit any change in total carbon uptake during the night due to the lack of mechanisms describing cellular respiration.
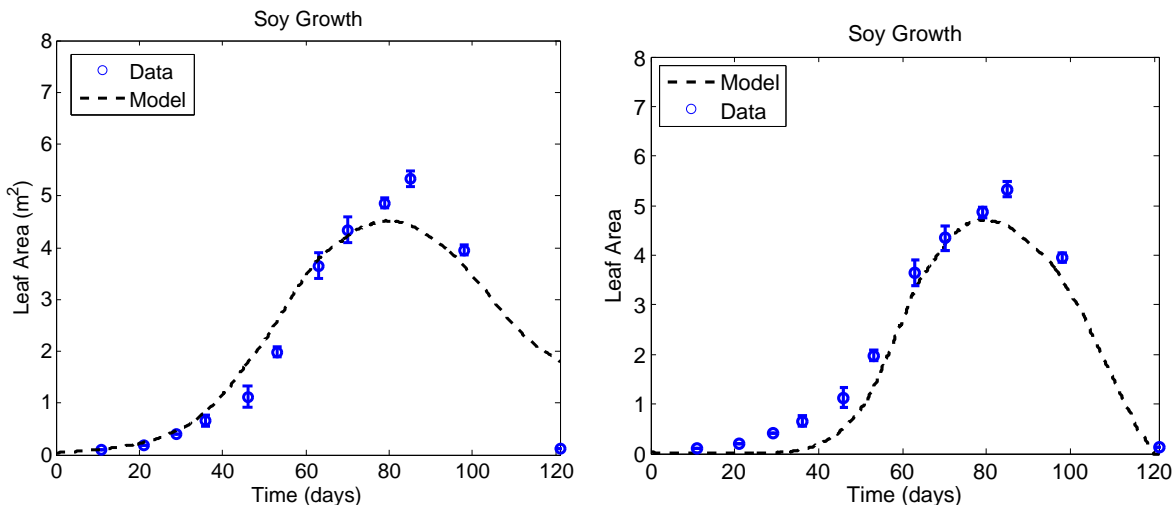
Figure 3.13: **Left:** Leaf area for soybean as predicted by the comprehensive model (dashed line) grown under conditions described in [39] and measured at Gretna, NE. (open circles with error bars). **Right:** Leaf area for soybean as predicted by the compact model (dashed line)grown under conditions described in [39] and measured at Gretna, NE (open circles with error bars).

The concentrations of PC$^+$, RuBP, T3P, PGA, T3Pc, ATP, and ADP predicted by each model for the first 20 days of growth are shown in Figures 3.15, 3.16, 3.17, 3.18, 3.19, 3.20, and 3.21 respectively. All metabolites shown exhibit concentrations that gravitate toward two levels, one during the day and a second during the night when the light reactions are occurring at a lower rate. The concentrations fluctuate, likely due to the environment's effect on how much O$_2$ and CO$_2$ the leaf absorbs as well as the dependence of ATP and NADPH production on available light. These fluctuations are most visible in the plot for RUBP concentration as predicted by the comprehensive model and the PGA, T3P, ADP, and ATP concentrations predicted by the compact model. While the concentrations predicted by the two models differ by an order of magnitude - and often less - they do not exactly agree.

The metabolites that are predicted to have a higher concentration during the day by the comprehensive model (PC$^+$, RuBP, T3P, PGA, T3Pc, and ADP) also are predicted to have
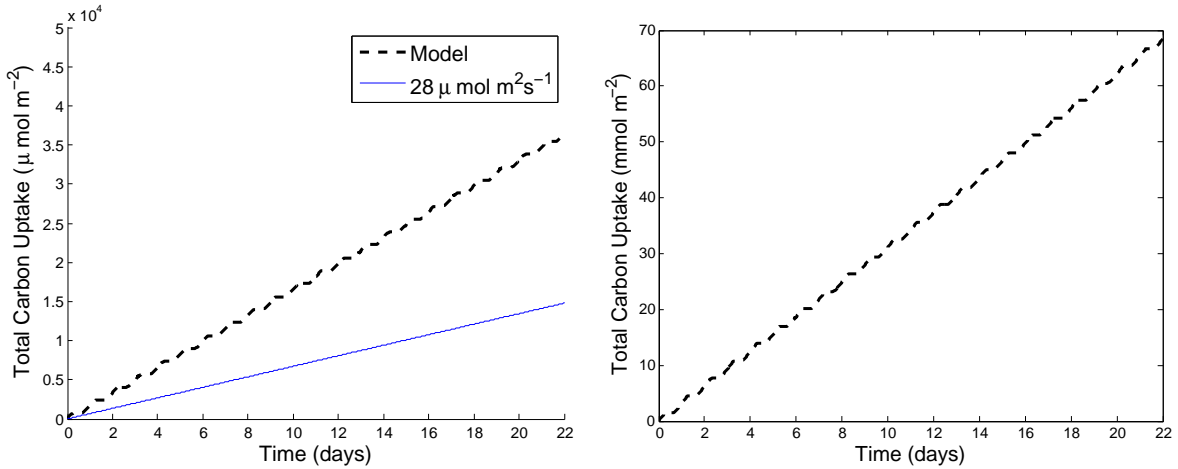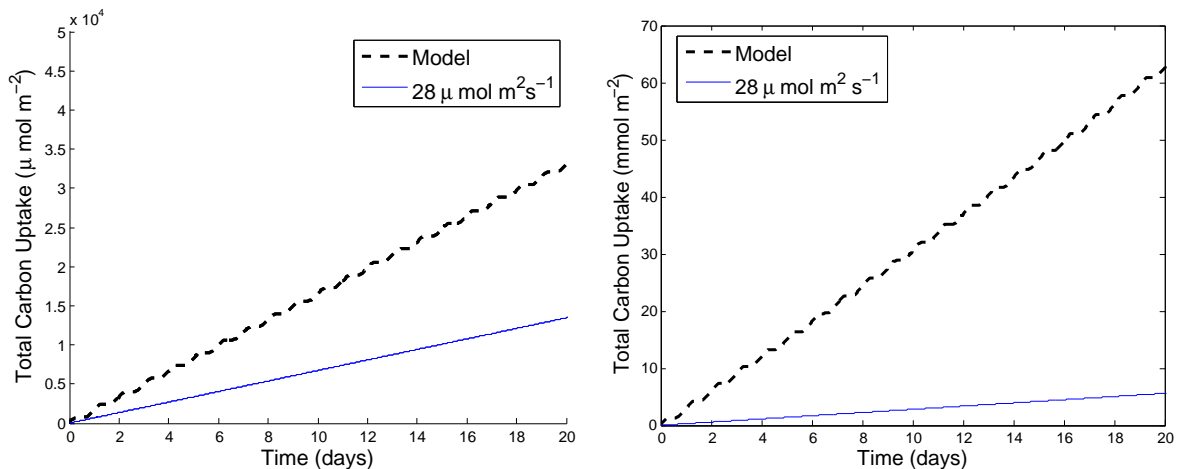
Figure 3.14: **Left:** First 20 days of carbon uptake as predicted by the comprehensive model (dashed) and by the constant rate of 28 μmol m^{-2}s^{-1} (solid) as predicted by [49]. **Right:** First 20 days of carbon uptake as predicted by the compact model (dashed) when grown under conditions described in [39] and measured at Gretna, NE.

higher daytime concentrations by the compact model, while those predicted to have a higher concentration at night by the comprehensive model (ATP) also are predicted to have a higher concentration at night by the compact model. The daytime concentrations of T3P (Figure 3.17) and T3Pc (Figure 3.19) are predicted by the compact model to be an order of magnitude higher than the comprehensive model.

At night, both models exhibit similar concentrations for RuBP, T3P, PGA, T3Pc, ADP, and ATP, metabolites that are heavily influenced by the dark reactions, but the concentration of PC$^+$ predicted by the compact model remains above 1 mmol l^{-1} whereas the prediction by the comprehensive model approaches 0 mmol l^{-1} (Figure 3.15). These differences indicate that replacing the light reaction model of [28] with a simpler model does impact the behavior of other model compartments, especially during the day. At night, the differences between the two models as exhibited by metabolite concentrations are less pronounced. The patterns seen in the soybean metabolites are similar to those seen in the *A. thaliana* metabolites because the parameters describing the enzyme-catalyzed reaction mechanics remain unchanged between the models used for the two species.

In order to determine the possible accuracy of our models to existing research, we compare the concentrations of RuBP, T3P, T3Pc, and PGA during the day predicted by the models to the concentrations predicted by the model in [49] when assuming a high capacity for cytosolic FBPase, and we compare the daytime concentrations of PC$^+$ predicted by the models and
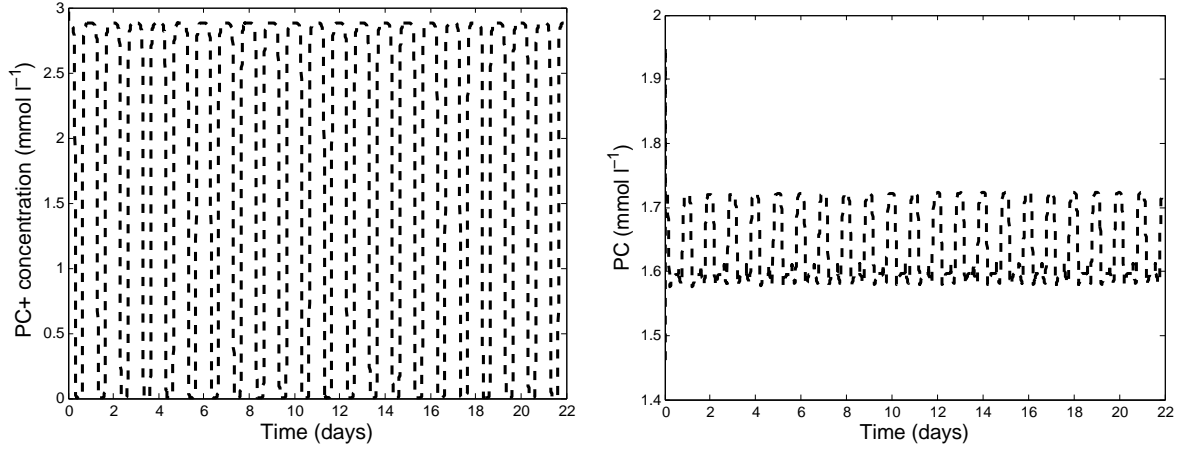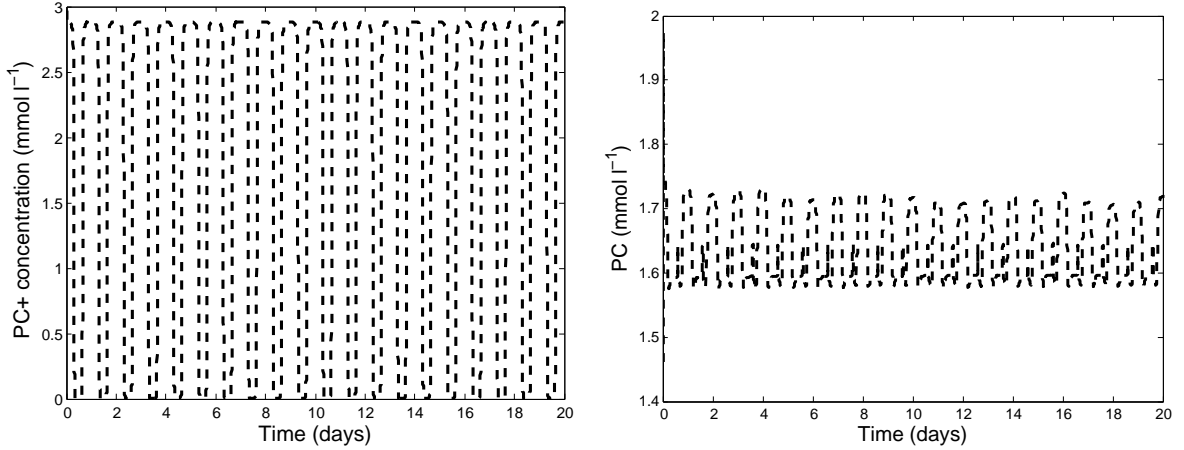
Figure 3.15: First 20 days of PC^+ concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.
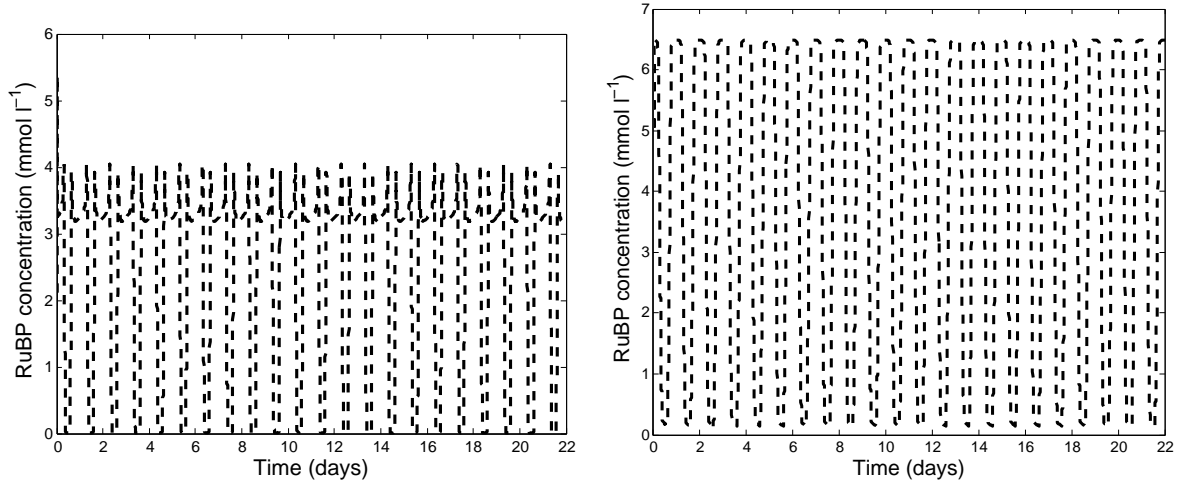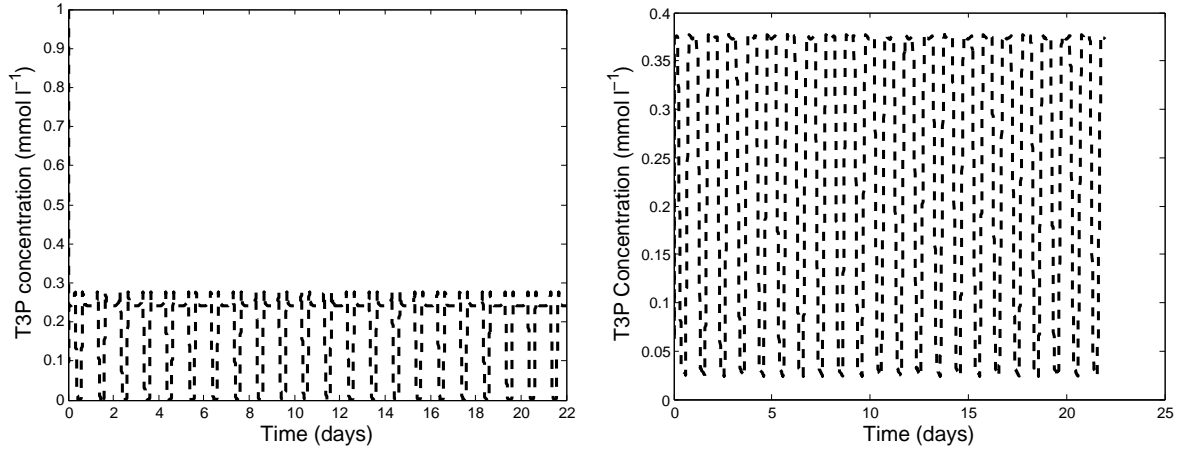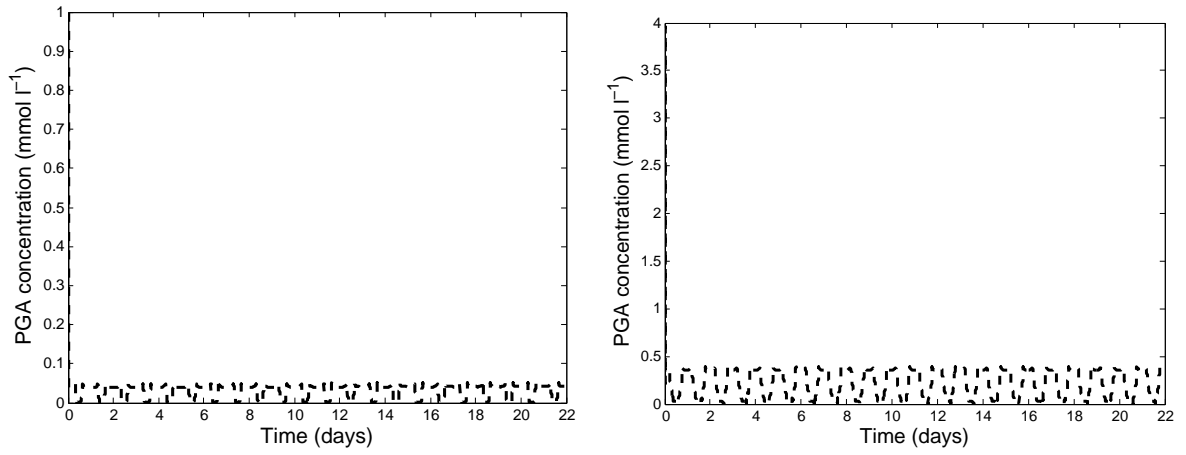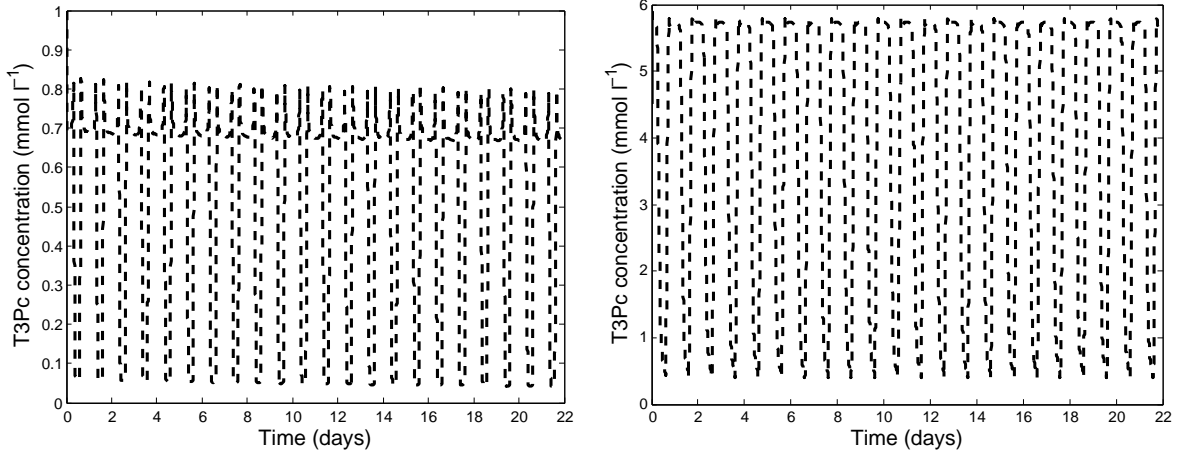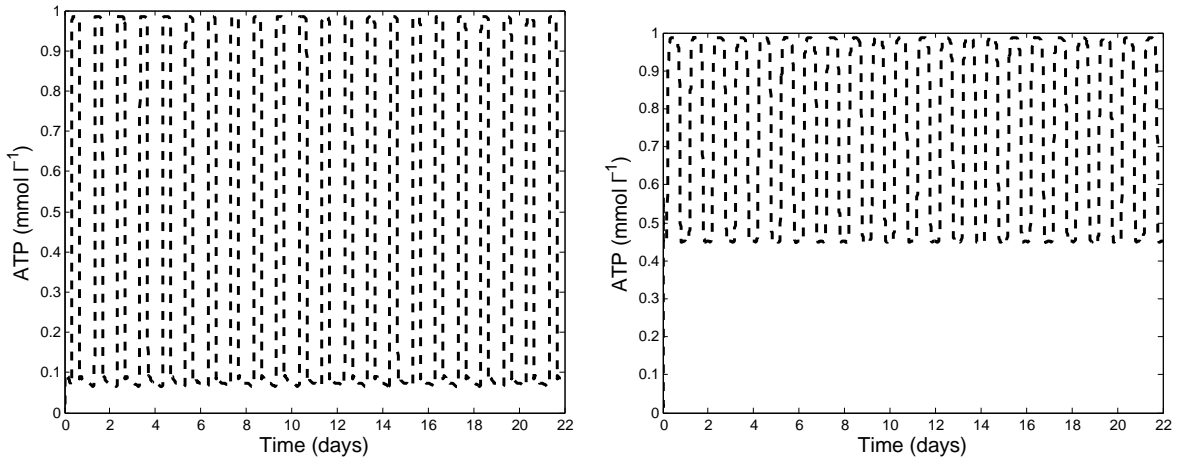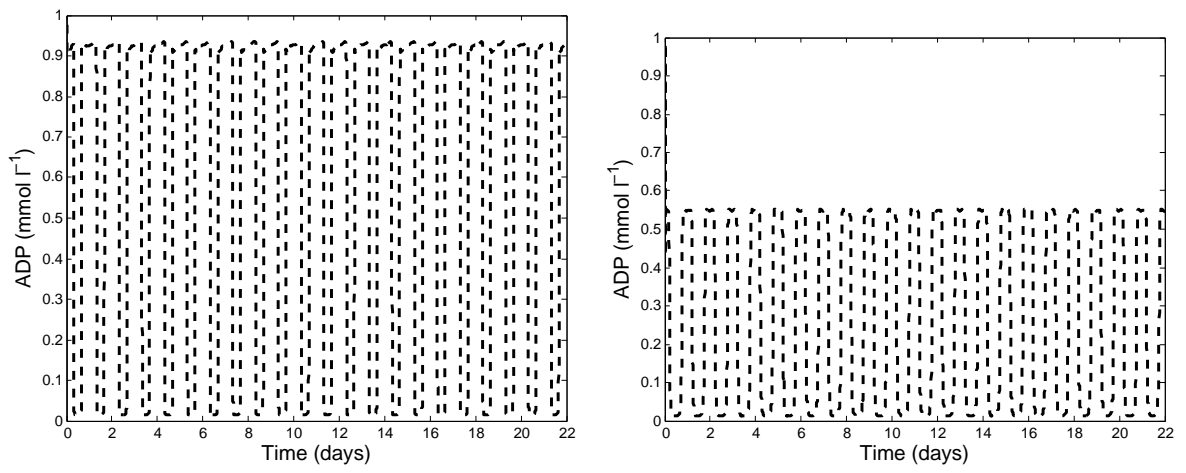
$\sum P_{700}^+$ predicted by the comprehensive model to the relative concentrations predicted by [28].

The concentrations of metabolites involved in the C3 cycle predicted by the model fit to the *A. thaliana* data [35] or the soybean data [39] are not near those predicted by [49] but are typically within an order of magnitude. Except for the concentration of RuBP, the metabolite concentrations predicted by the compact model are closer to the concentrations predicted by the model in [49]. The ratio of the concentrations are not similar. The model of [49] predicts that of the four listed metabolites involved in the C3 cycle, T3Pc has the highest concentration, folled by PGA, T3P, and finally RuBP. The proposed comprehensive and compact models predict that RuBP will have the highest concentration, followed by T3Pc, and finally T3P and PGA.

Because the Michaelis-Menten parameters remained unchanged between the models for each set of data, it is expected that the resulting solutions of the model to fit to the data sets of [35] and [39] would be similar - that is, to create different metabolite concentrations in the model, we must determine parameters that better describe the enzyme kinetics occurring in each species. These metabolite concentrations are listed in Table 3.4 and pictured in Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, and 3.12 for *A. thaliana* and Figures 3.15, 3.16, 3.17, 3.18, 3.19, 3.20, and 3.21 for soybean. The differences seen in metabolite concentrations predicted by [28] and [49] from the proposed models may be related to the incorporation of additional metabolic processes into the model or the effects of environmental factors: [49] shows that changes in atmospheric composition have a large impact on metabolite concentrations. Additionally, the numerical experiments in [28], reach only the P phase of chlorophyll α fluorescence and do not
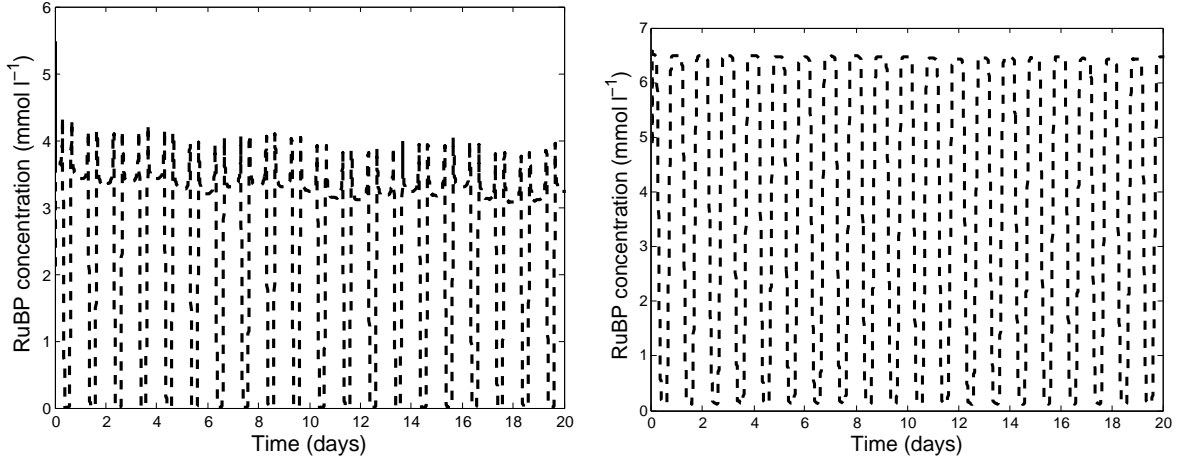
Figure 3.16: First 20 days of RuBP concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.

Table 3.4: Predicted daytime metabolite concentrations

Metabolite	Literature [28], [49]	Comprehensive Model	Compact Model
RuBP	1.8	3.2	6.3
T3P	3.3	0.25	0.36
PGA	6	0.05	0.4
T3Pc	9.5	0.7	5.7
PC^+	0.15	2.9	1.72
$\sum P_{700}^+$	≈ 0	0.82	-

investigate long-term dynamics.

Both the comprehensive and the compact models may be used in a least-squares minimization parameter estimation framework to approximate leaf area data. Both models exhibit changes in behavior between daytime and nighttime conditions through metabolite concentrations, leaf area rate of change, and carbon uptake rate. While the comprehensive model appears to better approximate leaf area during the vegetative phases of plant growth (Figures 3.4 and 3.13), there is little indication in the model formulations as to why this phenomenon may occur.
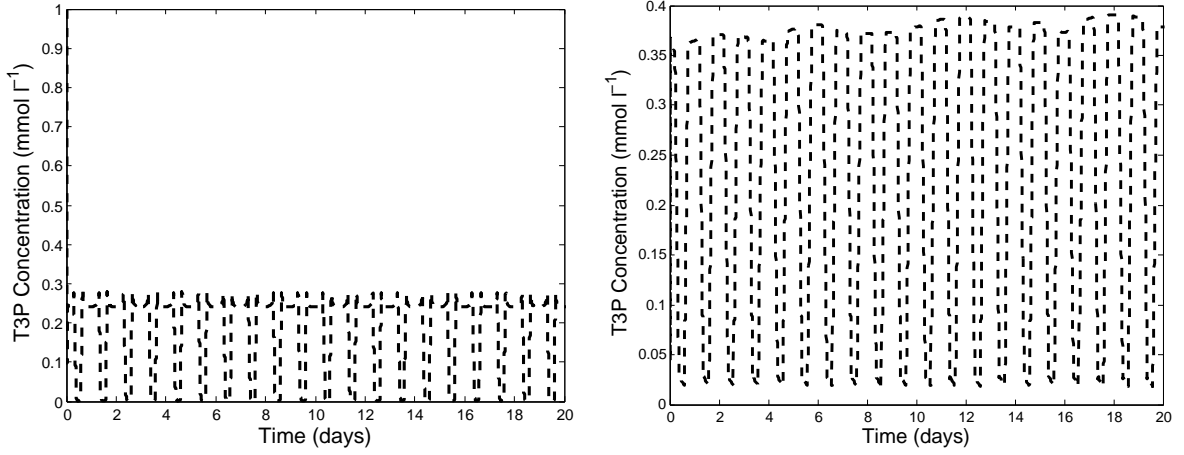
Figure 3.17: T3P concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.
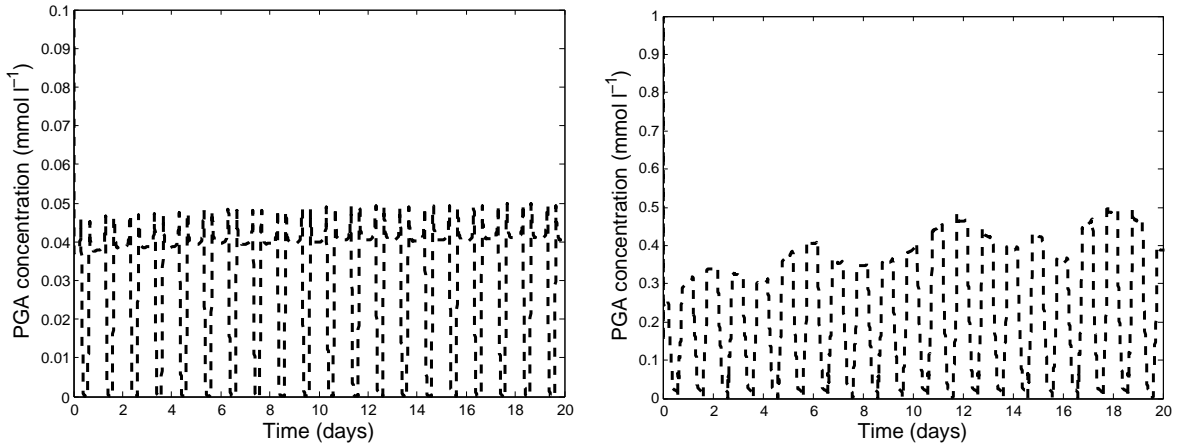
Figure 3.18: PGA concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.
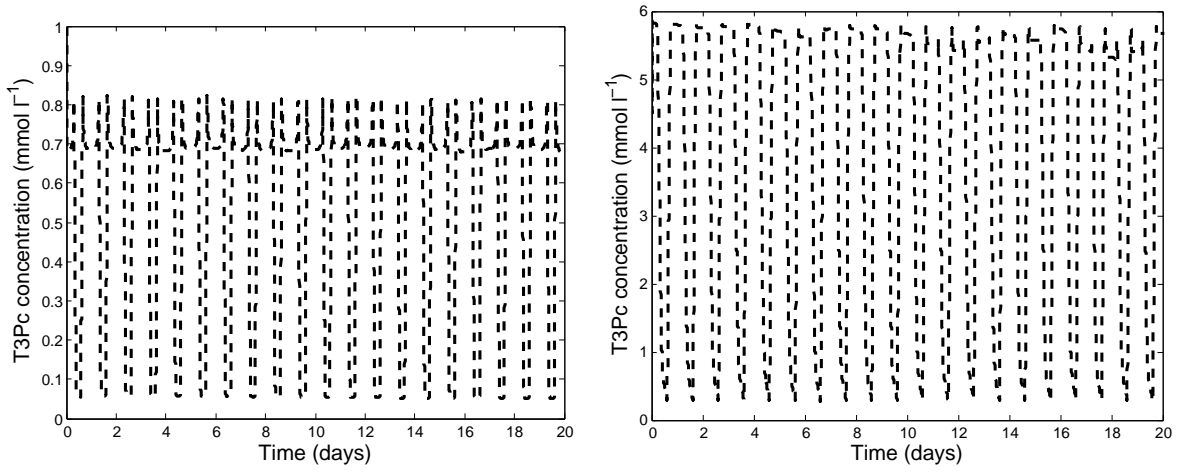
Figure 3.19: First 20 days of Cytosolic T3P concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.
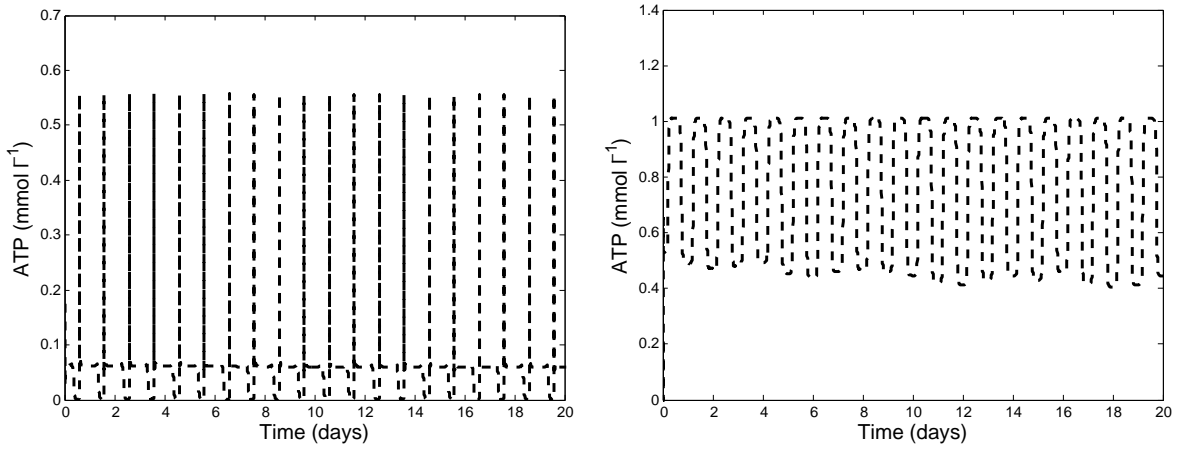
Figure 3.20: First 20 days of ATP concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.
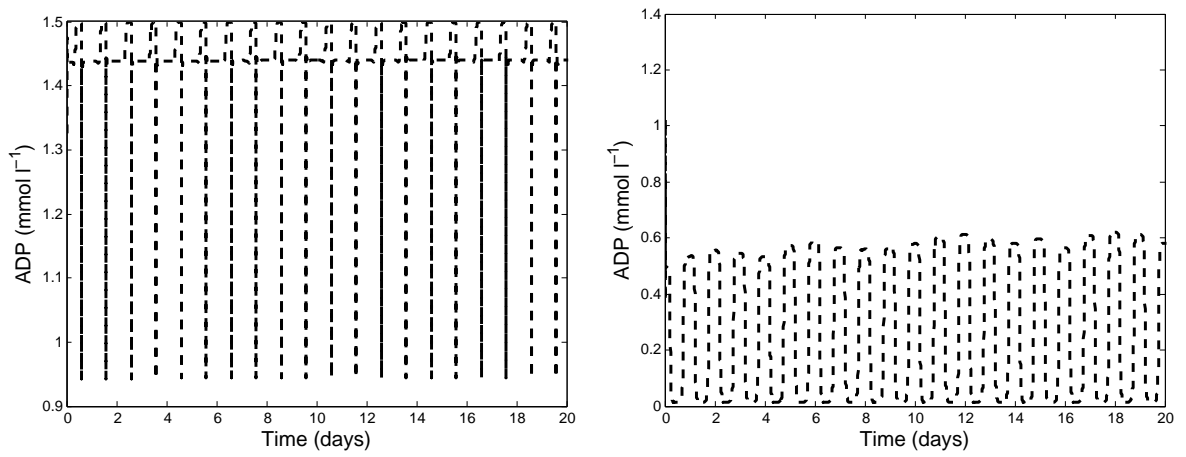
Figure 3.21: First 20 days of ADP concentration in soybean as predicted by the comprehensive model (left) and compact model(right) when grown under conditions described in [39] and measured at Gretna, NE.

3.2 Sensitivity analysis of C3 model

To better understand the influence of parameters and initial conditions on the reactions utilized in the proposed models, we perform sensitivity analysis as described in Chapter 2 on the compact model with respect to the maximum velocity terms, growth and senescence parameters, and initial conditions for the first 2000 minutes (approximately 1.4 days) of growth using the parameters estimated to best describe *Arabidopsis thaliana* growth in the experimental setting used in [35]. Temperature is again modeled using a sinusoid curve, and radiance is at its maximum between $t = 0$ and $t = 180$ minutes as well as $t = 1260$ and $t = 1620$ minutes and at its minimum of 0 between $t = 540$ and $t = 900$ minutes as well as after $t = 1980$ minutes. Initial sensitivity analysis on the C3 cycle model [49] is performed in [6] to identify parameters that may potentially be best candidates to estimate in an inverse problem; here we examine which model equations (and hence potentially observable state variables) are changed when parameter values are varied.

The state variables primarily associated with the light reactions, PQ, PQH_2, PC, PC^+, Fd, and Fd^-, are sensitive to changes in many maximum velocity terms and approximately half of the initial conditions, but not to the growth and senescence parameters. This indicates that the concentrations of these metabolites are affected by the properties of reactions that are not light dependent. All six state variables exhibit similar sensitivities to parameters. The sensitivity of PQ with respect to several parameters are pictured in Figure 3.22. Many of the sensitivity plots exhibit extreme changes in slope at $t = 280$, $t = 875$, $t = 1100$, and $t = 1720$. The periods of change centered around $t = 280$, $t = 1100$, and $t = 1720$ correspond to changes between day and night, which is when the dynamical system is showing the most change. After $t = 0$ the sensitivity of PQ to PQ_0 and Fd_0^- appear to follow the same trends, indicating that the effects of PQ_0 and Fd_0^- on PQ are similar as time progresses.

The state variables primarily associated with the C3 cycle are also sensitive to changes in many maximum velocity terms and approximately one quarter of the initial conditions, but not to the growth and senescence parameters. Representative of the sensitivities are the sensitivity of RuBP to parameters and initial conditions, of which a selection is shown in Figure 3.23. While the metabolites used in the C3 cycle are often impacted by changes to the properties of the reactions described by the parameters (in this case the maximum velocity terms), they are not sensitive to the initial concentrations of most other metabolites – even their own initial concentrations – once the model progresses past $t = 1$. Many of the sensitivity plots exhibit extreme changes in slope at $t = 280$, $t = 1100$, and $t = 1720$, and with lower frequency, $t = 875$. The periods of change centered around $t = 280$, $t = 1100$, and $t = 1720$ correspond to changes between day and night, an indication that as radiation intensity and temperature change, a parameter's effect on model compartments may also change.

Carbon uptake $\int A(t, \vec{E}(t); \vec{\theta})dt$ is sensitive to many parameters directly utilized in the C3 cycle and light-dependent reaction equations. A small sample of the sensitivity plots are pictured in Figure 3.24. Like other state variables related to the light-dependent reactions and the C3 cycle, the greatest changes in curvature of the sensitivity plots occur at approximately $t = 280$, $t = 1100$, and $t = 1720$. While the sensitivity of carbon uptake varies over time for some parameters, its sensitivity to others, such as RuBP_0 and A_0, remains constant after a short period of time. The influence of parameters on the value of $\int A(t, \vec{E}(t); \vec{\theta})dt$ is thus unchanged for the duration of the model solution after the first few minutes of the model solution are calculated.

The growth and senescence equations $G(t, \vec{E}(t); \vec{\theta})$ and $S(t, \vec{E}(t); \vec{\theta})$, respectively, are not affected by changes to most parameters or initial conditions. As seen in prior sensitivity analysis performed on logistic equations [12] and in the calculated sensitivities pictured in Figure 3.25, the sensitivity of these two equations to their initial condition decreases over time while their sensitivity to their growth rates r_G and r_S increase over time. These equations are not observed to be very sensitive to k_G and k_S, however, 2000 minutes is too short of a time frame for these two equations to approach their carrying capacity.

As the culmination of activity in the light-dependent reactions, C3 cycle, and whole-plant growth and senescence, leaf area $L(t, \vec{E}(t); \vec{\theta})$ is sensitive to many maximum velocity terms, the growth and senescence parameters, and initial conditions of several other equations. A small sample of the sensitivity plots are pictured in Figure 3.26. As leaf area is directly related to the carbon uptake rate $A(t, \vec{E}(t); \vec{\theta})$, it is expected that $L(t, \vec{E}(t); \vec{\theta})$ and $\int A(t, \vec{E}(t); \vec{\theta})dt$ are sensitive to the many of the same parameters. The sensitivity curves pictured in Figure 3.24 are very similar in shape to those in Figure 3.26 that depict each state variables' dependence on the same parameter.

The relationships between the state variables of the compact model and parameters including the maximum velocity terms, growth and senescence parameters, and initial conditions are much more varied than those seen in the C3 cycle model [49] when sensitivity analysis was performed in [6]. This may be due in part to the light-dependent reaction equations causing previously unrelated metabolites to interact with each other or to the use of environmental factors to modulate reaction velocities. Exploring which state variables are sensitive to particular parameters is important when interpreting and understanding why particular observable state variables are selected in a Fisher information matrix-based experimental design scheme.
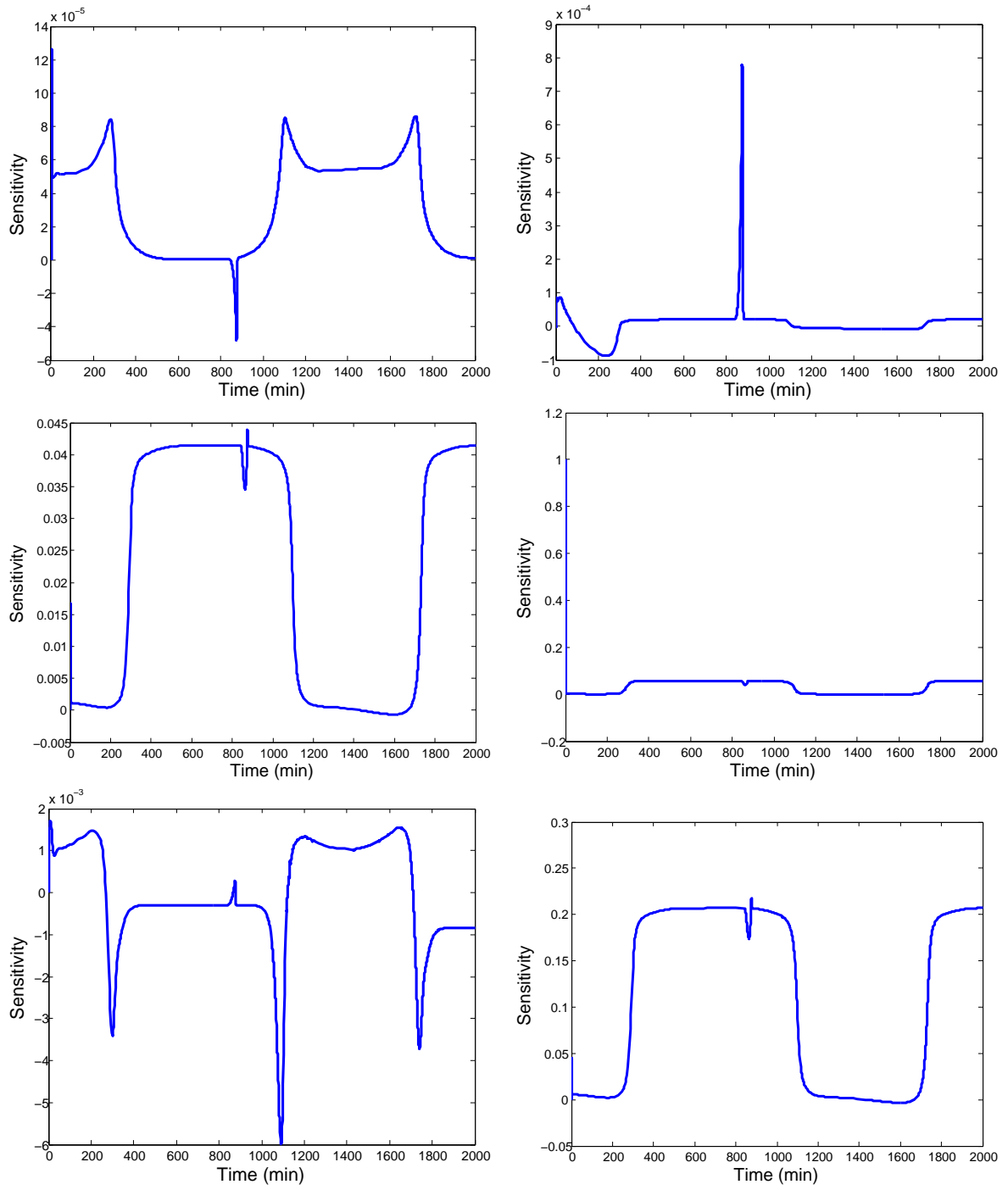
Figure 3.22: **Left:** Sensitivity of PQ to the maximum velocity parameters V_3, V_{38},and V_{52} (top to bottom). **Right:** Sensitivity of PQ to the initial conditions $SERc_0$, PQ_0, and Fd_0^-.
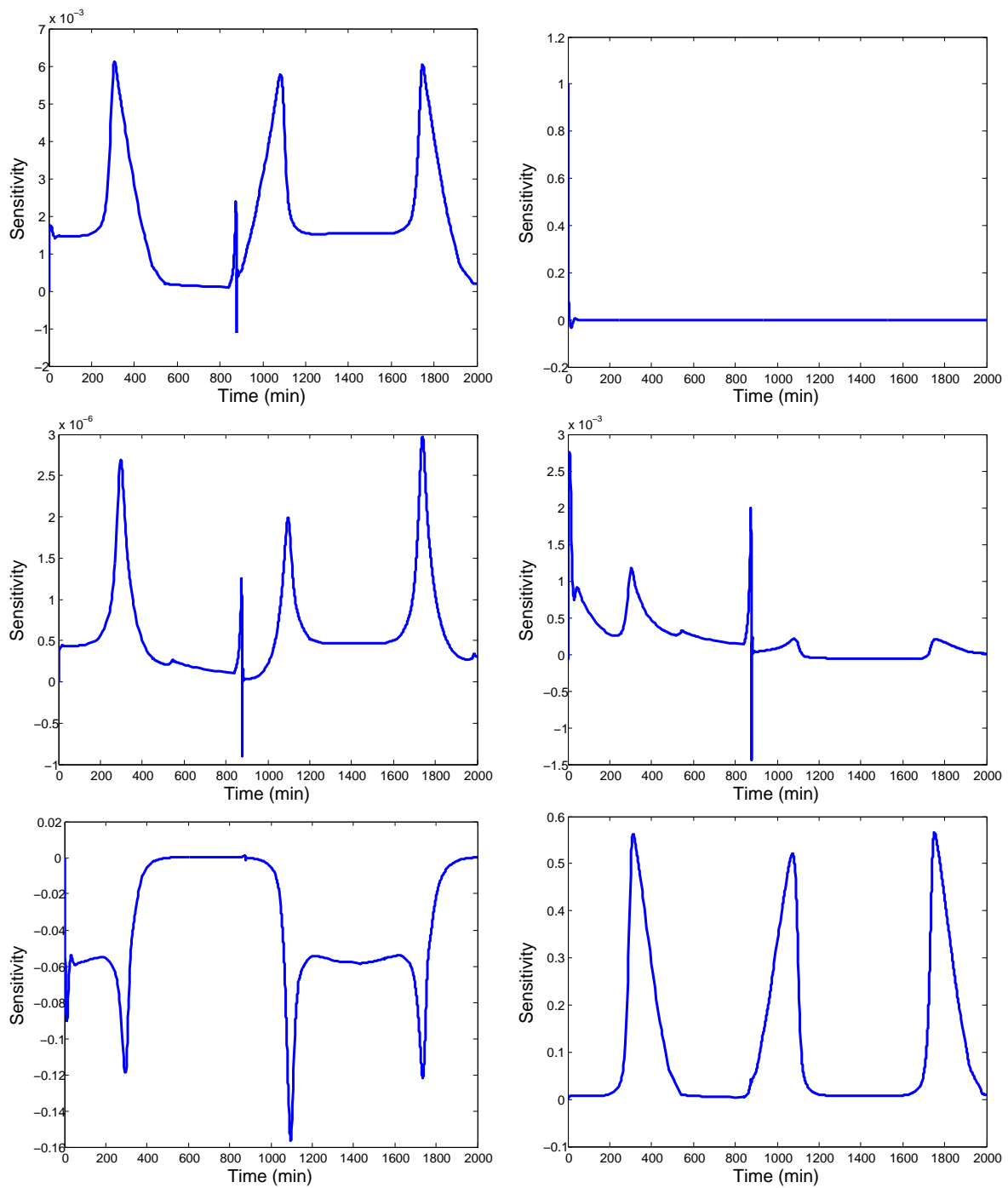
Figure 3.23: **Left:** Sensitivity of RuBP to the maximum velocity parameters V_1, V_{38}, and V_{52} (top to bottom). **Right:** Sensitivity of PQ to the initial conditions $RuBP_0$, $SERc_0$, and PQ_0.
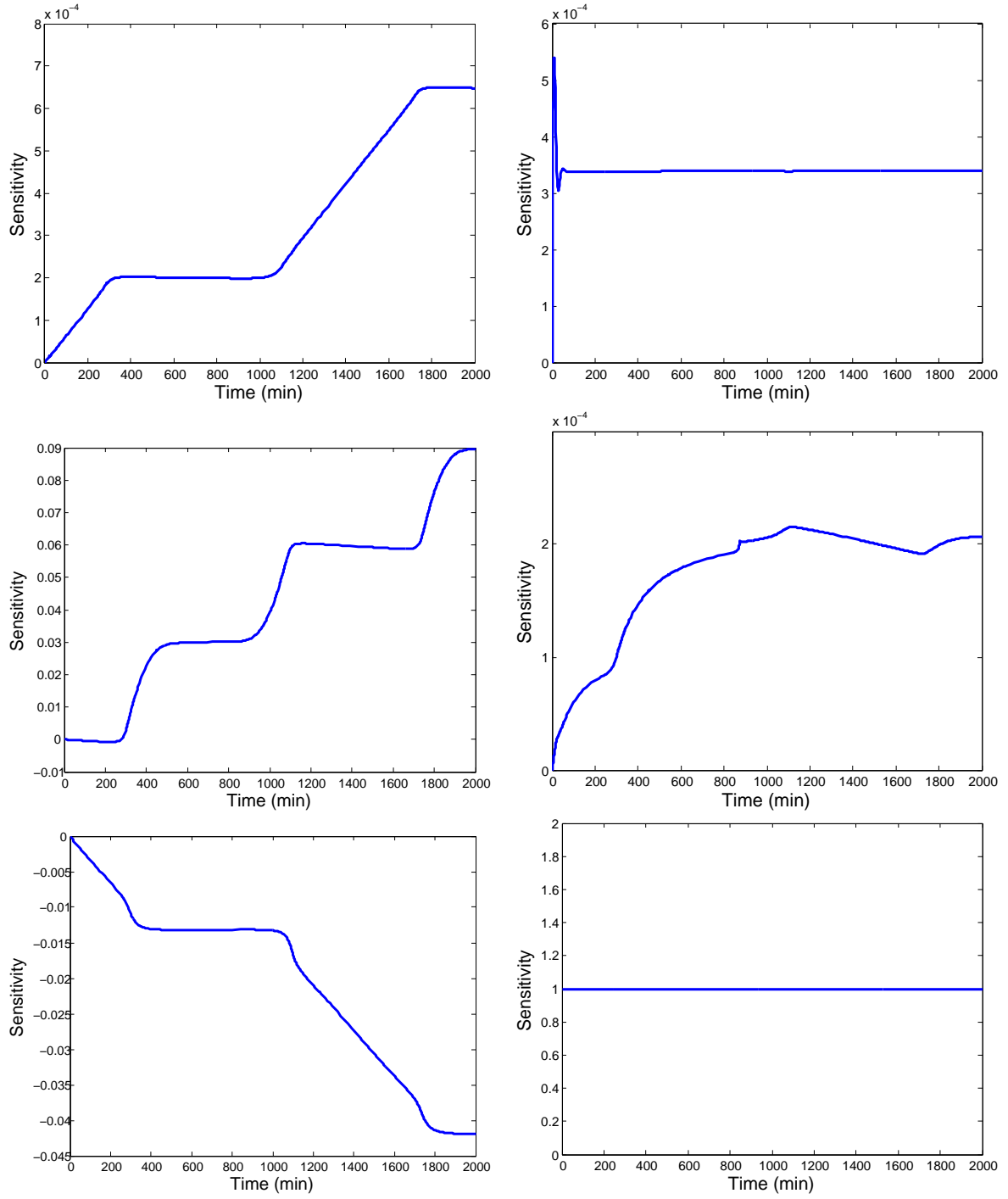
Figure 3.24: **Left:** Sensitivity of $\int A(t)dt$ to the maximum velocity parameters V_1, V_{39},and V_{52} (top to bottom). **Right:** Sensitivity of $\int A(t)dt$ to the initial conditions $RuBP_0$, $SERc_0$, and A_0.
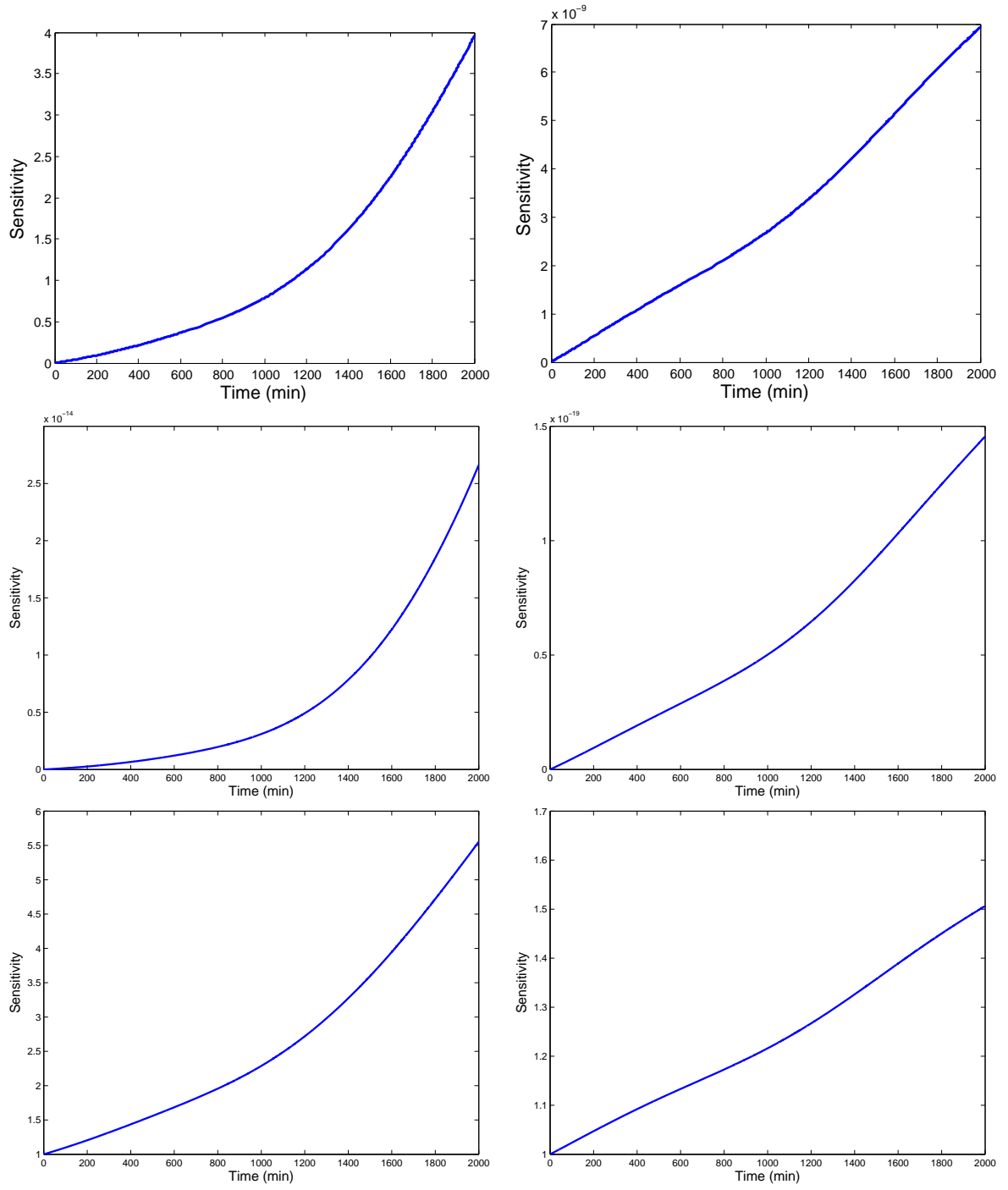
Figure 3.25: **Left:** Sensitivity of $G(t, \vec{E}(t); \vec{\theta})$ to r_G, k_G, and G_0 (top to bottom). **Right:** Sensitivity of $S(t, \vec{E}(t); \vec{\theta})$ to r_S, k_S, and S_0 (top to bottom).
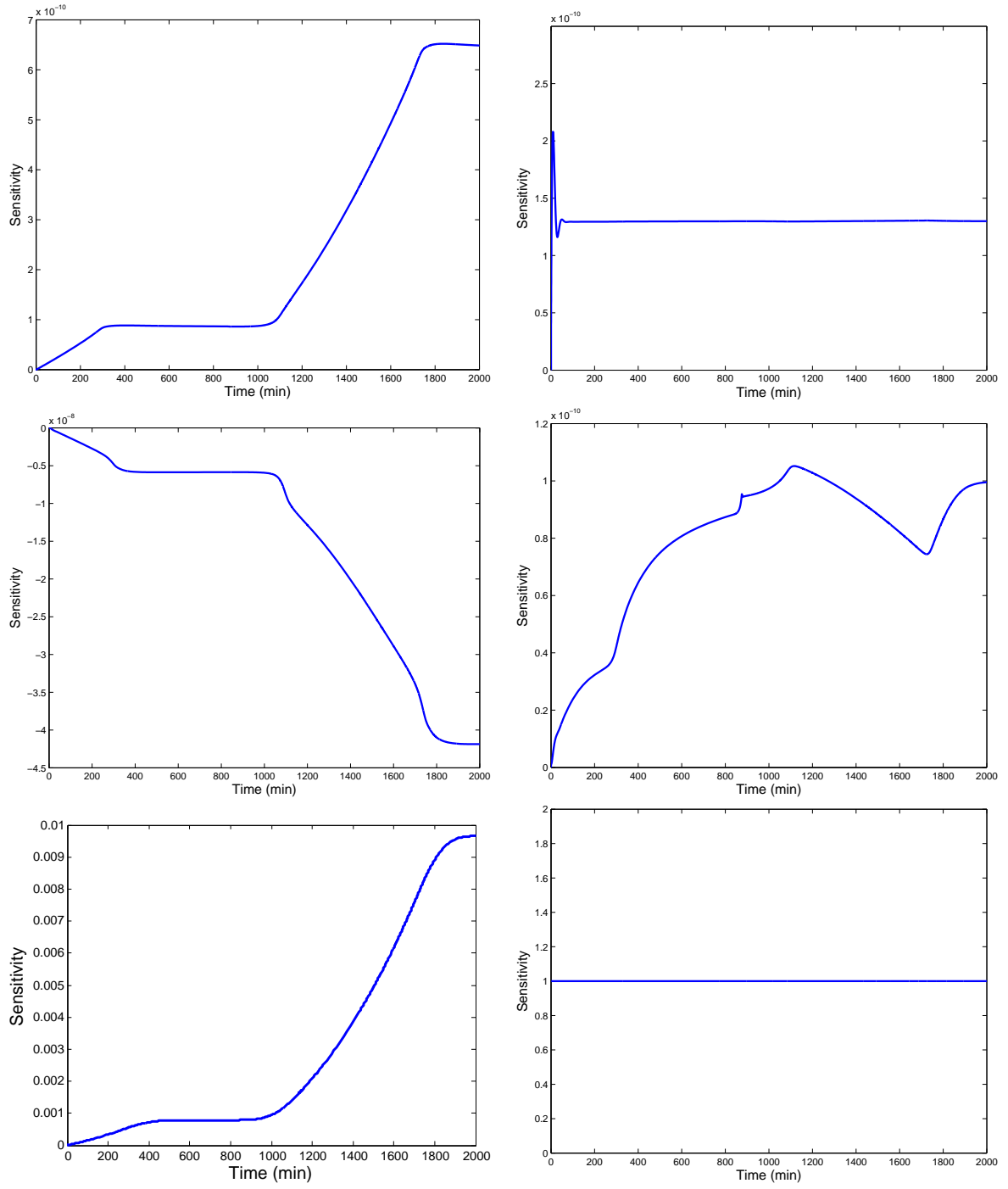
Figure 3.26: **Left:** Sensitivity of leaf area $L(t, \vec{E}(t); \vec{\theta})$ to the parameters V_3, V_{52}, and r_g (top to bottom). **Right:** Sensitivity of $A(t)$ to the initial conditions $RuBP_0$, $SERc_0$, and L_0.

3.3 Experimental design

Since using either the comprehensive plant growth model or compact plant growth model to perform the experimental design procedure described in Chapter 2 for even 20 days of plant growth would require months of computations on current personal computer machines, we demonstrate the ability of the experimental design techniques on two simpler examples: an HIV model [1] used to model basic immune response over an extended period of time and a C3 cycle model [49] describing short term reaction rates. Code accuracy for time point selection was validated by replicating the results found in [16]. After determining optimal observables, we illustrate how an optimized set of observation time points affects calculated asymptotic standard errors for parameters of interest.

3.3.1 HIV model

We examine the performance of both the time point and observable operator selection algorithms on the log-scaled version of the HIV model developed in [1]. While the analytic solution of this system cannot easily be found, this model has been studied, improved, and successfully fit to and indeed validated with several sets of longitudinal data using parameter estimation techniques [2, 3]. Additionally, the sampling or observation operators used to collect data in a clinical setting as well as the relative usefulness of these sampling techniques are known [2]. The model includes uninfected and infected CD4$^+$ T-cells, called type 1 target cells (T_1 and T_1^*, respectively), uninfected and infected macrophages (subsequently determined to more likely be resting or inactive CD4$^+$ T-cells), called type 2 target cells (T_2 and T_2^*), infectious free virus V_I, and immune response E produced by cytotoxic T-lymphocytes CD8$^+$. The HIV model with treatment function $u(t)$ is given by

$$
\begin{aligned}
\dot{T_1} &= \lambda_1 - d_1 T_1 - (1 - \epsilon_1 u(t))k_1 V_I T_1 \\
\dot{T_2} &= \lambda_2 - d_2 T_2 - (1 - f\epsilon_1 u(t))k_2 V_I T_2 \\
\dot{T_1^*} &= (1 - \epsilon_1 u(t))k_1 V_I T_1 - \delta T_1^* - m_1 E T_1^* \\
\dot{T_2^*} &= (1 - f\epsilon_1 u(t))k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V} &= N_T \delta (T_1^* + T_2^*) - [c + (1 - \epsilon_2 u(t))10^3 k_1 T_1 + (1 - f\epsilon_1 u(t))10^3 k_2 T_2]V_I \\
\dot{E} &= \lambda_E + b_E \frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_b} E - d_E \frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_d} E - \delta_E E.
\end{aligned}
\tag{3.1}
$$

The log-scaled HIV model, which is used to alleviate difficulties due to the large differences

56

in scales of the variables and the parameters, is

$$\dot{x}_1 = \frac{10^{-x_1}}{\ln(10)}\left(\lambda_1 - d_1 10^{x_1} - (1 - \epsilon_1 u(t))k_1 10^{x_1+x_5}\right)$$

$$\dot{x}_2 = \frac{10^{-x_2}}{\ln(10)}\left(\lambda_2 - d_2 10^{x_2} - (1 - f\epsilon_1 u(t))k_2 10^{x_5+x_2}\right)$$

$$\dot{x}_3 = \frac{10^{-x_3}}{\ln(10)}\left((1 - \epsilon_1 u(t))k_1 10^{x_5+x_1} - \delta 10^{x_3} - m_1 10^{x_6+x_3}\right)$$

$$\dot{x}_4 = \frac{10^{-x_4}}{\ln(10)}\left((1 - f\epsilon_1 u(t))k_2 10^{x_5+x_2} - \delta 10^{x_4} - m_2 10^{x_6+x_4}\right) \tag{3.2}$$

$$\dot{x}_5 = \frac{10^{-x_5}}{\ln(10)}\left((1 - \epsilon_2 u(t))10^3 N_T \delta(10^{x_3} + 10^{x_4}) - c 10^{x_5}\right.$$
$$\left. -(1 - \epsilon_1 u(t))\rho_1 10^3 k_1 10^{x_1+x_5} - (1 - f\epsilon_1 u(t))\rho_2 10^3 k_2 10^{x_2+x_5}\right)$$

$$\dot{x}_6 = \frac{10^{-x_6}}{\ln(10)}\left(\lambda_E + \frac{b_E(10^{x_3} + 10^{x_4})}{10^{x_3} + 10^{x_4} + K_b}10^{x_6} - \frac{d_E(10^{x_3} + 10^{x_4})}{10^{x_3} + 10^{x_4} + K_d}10^{x_6} - \delta_E 10^{x_6}\right), \tag{3.3}$$

where $T_1 = 10^{x_1}$, $T_2 = 10^{x_2}$, $T_1^* = 10^{x_3}$, $T_2^* = 10^{x_4}$, $V_I = 10^{x_5}$, and $E = 10^{x_6}$.

In [2] this model's parameters are estimated for each of 45 different patients who were in a treatment program for HIV at Massachusetts General Hospital (MGH). These individuals experienced interrupted treatment schedules, in which the patient did not take any medication for viral load control. Seven of these patients adhered to a structured treatment interruption schedule planned by the clinician. We use the parameters estimated to fit the data of one of these patients, Patient 4 in [2], as our "true" parameters $\vec{\theta}_0$ for this model. This patient was chosen because the patient continued treatment for an extended period of time (1919 days) and the corresponding data set contains 158 measurements of viral load and 107 measurements of CD4 cell count $(T_1 + T_1^*)$ that exhibit a response to interruption in treatment, and the estimated parameters yield a model exhibiting trends similar to that in the data.

Sensitivity analysis performed in [1] and parameter subset selection carried out in [9] identified subsets of the 20 parameters and six initial conditions that would likely be reliably estimated when solving the corresponding inverse problems. Based on their results, we treat the subset of parameters $\vec{\theta} = (\lambda_1, d_1, \epsilon_1, k_1, \epsilon_2, N_T, c, b_E, x_1^0, x_2^0, x_5^0)$ as unknown and fix all other parameters. The values for the fixed parameters and $\vec{\theta}_0$ are computed in [2] and given in Table 3.5. It is important to choose parameters to which the model is sensitive; a poor choice for the components of $\vec{\theta}$ will negatively affect sensitivity matrices and may lead to a near-singular FIM. Based on currently available measurement methods, we allow the possible types of observations including (1) infectious virus x_5, (2) immune response x_6, (3) total activated CD4 cells $x_1 + x_3$, and (4) type 2 (resting or unactivated) target cells $x_2 + x_4$, each with an assumed error variance of 10% of the initial variable values given by $\vec{x}^0 = (3.0799, 1.2443, 1.7899, -0.2150, 5.9984, -0.7251)^{\mathsf{T}}$.

Table 3.5: Values of parameters in the HIV model (3.1).

parameter	value	unit	parameter	value	unit
λ_1	4.633	$\frac{\text{cells}}{\mu\text{ml-blood·day}}$	λ_2	0.1001	$\frac{\text{cells}}{\mu\text{ml-blood·day}}$
d_1	0.004533	$\frac{1}{\text{day}}$	d_2	0.02211	$\frac{1}{\text{day}}$
k_1	1.976e-6	$\frac{\text{ml}}{\text{virions·day}}$	k_2	5.529e-4	$\frac{\text{ml}}{\text{virions·day}}$
ϵ_1	$10^{0.6017}$	-	ϵ_2	$10^{0.5403}$	-
m_1	0.02439	$\frac{\mu\text{ml}}{\text{cells·day}}$	m_2	0.013099	$\frac{\mu\text{ml}}{\text{cells·day}}$
δ	0.1865	$\frac{1}{\text{day}}$	c	19.36	$\frac{1}{\text{day}}$
f	$10^{0.53915}$	-	N_T	19.04	$\frac{\text{virions}}{\text{cell}}$
λ_E	0.009909	$\frac{\text{cells}}{\mu\text{ml-blood·day}}$	δ_E	0.0703	$\frac{1}{\text{day}}$
b_E	0.09785	$\frac{1}{\text{day}}$	d_E	0.1021	$\frac{1}{\text{day}}$
K_b	0.3909	$\frac{\text{cells}}{\mu\text{ml-blood}}$	K_d	.8379	$\frac{\text{cells}}{\mu\text{ml-blood}}$

3.3.2 HIV observable selection results, times fixed

To simulate the experience of a clinician gathering data as a patient regularly returns for scheduled testing, we fix the sampling times and choose the optimal observables. We consider choices of $N = 1, 2, 3$ sampling operators out of the four possible observables, all N of which will be measured at $n = 51, 101, 201, 401$ evenly spaced times over 2000 days, corresponding to measurements every 40, 20, 10, and 5 days, respectively. The N optimal sampling maps are determined for each of the three optimal design criteria, and the asymptotic standard errors (ASE) are calculated after carrying out the corresponding inverse problem calcualtions.

Tables 3.6, 3.7, and 3.8 display the optimal observation operators determined by the D-, E-, and SE-optimal design cost functions, respectively, as well as the lowest and highest ASE. In all three optimal design criteria, there was a distinct best choice of observables (listed in Tables 3.6, 3.7, and 3.8) for each pair of n and N. When only $N = 1$ observable could be measured, each design criterion consistently picked the same observable for all n; similarly, at $N = 2$, both the D-optimal and SE-optimal design criteria were consistent in their selection over all n, and E-optimal only differed at $n = 401$. Even at $N = 3$, each optimal design method specified at least two of the same observables at all n.

The observables that were rated best changed between criteria, affirming the fact that each optimal design methods minimizes different aspects of the standard error ellipsoid. At $N = 1$

observable, D-optimal selects the CD4 cell count while E-optimal and SE-optimal choose the infectious virus count. As a result, the min(ASE) calculated for a parameter estimation problem using the D-optimal observables is approximately 1/3 lower than the min(ASE) of E- and SE-optimal for all tested time point distributions. Similarly, the max(ASE) calculated for E- and SE-optimal designed parameter estimation problems is approximately 1/3 lower than that of D-optimal. Thus at $N = 1$, based on minimum and maximum asymptotic standard errors, there is no clear best choice of an optimal design cost function.

At $N = 2$ allowed observables, both the D- and SE-optimal cost functions are minimized by an observation operator containing both activated CD4 cells (type 1 target cells) and type 2 target cells. The E-optimal cost function still favors infectious virus count in addition to type 2 target cells (at $n = 51, 101, 201$) or type 1 target cells (at $n = 401$). As a result, max(ASE) in the E-optimal design parameter estimation problem is 0% – 20% lower than that of D- or SE-optimal; however, the D- and SE-optimal designs reduce min(ASE) by 20% – 30% of the min(ASE) of E-optimal. Therefore at $N = 2$, the E-optimal cost functional would be preferable if the largest ASE's were to be reduced, but for the best overall improvement (as measured by percent reduction from the E-optimal ASE), D- and SE-optimal are recommended.

When selecting $N = 3$ observables, each of the three design criteria select many of the same observables. This is to be expected as $N^* = 4$ in this simulation. For $n = 51, 101, 201$, both total CD4 cell count and immune response $E(t)$ are selected by all design criteria. The D-optimal criterion also chooses type 2 cell count, so the lack of information on virus count as measured by $x_5(t)$ leads to its high max(ASE), $ASE(x_5^0)$. E-optimal (and at larger n, SE-optimal) choose to measure infectious virus count, reducing $\mathrm{ASE}(x_5^0)$ and thus reducing the max(ASE) by more than 50%. While at low n, E-optimal has the lowest min(ASE) and max(ASE), SE-optimal performs better at high n, so when selecting $N = 3$ observables, the number of time points n may affect which optimal design cost function performs best.

Table 3.6: Number of observables, number of time points, observables selected by D-optimal cost functional, and the minimum and maximum standard error and associated parameter for the parameter subset in the HIV model (3.2).

N	n	Observables	min(ASE)	max(ASE)
1	51	$x_1 + x_3$	$\text{ASE}(\lambda_1) = 0.18$	$\text{ASE}(\lambda_E) = 6.35$
1	101	$x_1 + x_3$	$\text{ASE}(\lambda_1) = 0.13$	$\text{ASE}(\lambda_E) = 4.40$
1	201	$x_1 + x_3$	$\text{ASE}(\lambda_1) = 0.091$	$\text{ASE}(\lambda_E) = 3.03$
1	401	$x_1 + x_3$	$\text{ASE}(\lambda_1) = 0.065$	$\text{ASE}(\lambda_E) = 2.13$
2	51	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.095$	$\text{ASE}(x_5^0) = 2.31$
2	101	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.068$	$\text{ASE}(x_5^0) = 1.52$
2	201	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.047$	$\text{ASE}(x_5^0) = 1.26$
2	401	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.033$	$\text{ASE}(x_5^0) = 1.13$
3	51	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.045$	$\text{ASE}(x_5^0) = 2.26$
3	101	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.032$	$\text{ASE}(x_5^0) = 1.43$
3	201	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.022$	$\text{ASE}(x_5^0) = 1.23$
3	401	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.015$	$\text{ASE}(x_5^0) = 1.12$

Table 3.7: Number of observables, number of time points, observables selected by E-optimal cost functional, and the minimum and maximum standard error and associated parameter for the parameter subset in the HIV model (3.2).

N	n	Observables	min(ASE)	max(ASE)
1	51	x_5	$\text{ASE}(d_1) = 0.27$	$\text{ASE}(\lambda_E) = 4.27$
1	101	x_5	$\text{ASE}(d_1) = 0.19$	$\text{ASE}(\lambda_E) = 3.11$
1	201	x_5	$\text{ASE}(d_1) = 0.13$	$\text{ASE}(\lambda_E) = 2.27$
1	401	x_5	$\text{ASE}(d_1) = 0.094$	$\text{ASE}(\lambda_E) = 1.63$
2	51	$x_5,\ x_2 + x_4$	$\text{ASE}(d_1) = 0.12$	$\text{ASE}(\lambda_E) = 2.18$
2	101	$x_5,\ x_2 + x_4$	$\text{ASE}(d_1) = 0.095$	$\text{ASE}(\lambda_E) = 1.52$
2	201	$x_5,\ x_2 + x_4$	$\text{ASE}(d_1) = 0.065$	$\text{ASE}(\lambda_E) = 1.10$
2	401	$x_5,\ x_1 + x_3$	$\text{ASE}(\lambda_1) = 0.042$	$\text{ASE}(\lambda_E) = 0.86$
3	51	$x_5,\ x_6,\ x_1 + x_3$	$\text{ASE}(\lambda_E) = 0.045$	$\text{ASE}(x_5^0) = 0.77$
3	101	$x_5,\ x_6,\ x_1 + x_3$	$\text{ASE}(\lambda_E) = 0.032$	$\text{ASE}(x_5^0) = 0.73$
3	201	$x_5,\ x_6,\ x_1 + x_3$	$\text{ASE}(\lambda_E) = 0.022$	$\text{ASE}(x_5^0) = 0.69$
3	401	$x_5,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.032$	$\text{ASE}(x_5^0) = 0.65$

Table 3.8: Number of observables, number of time points, observables selected by SE-optimal cost functional, and the minimum and maximum standard error and associated parameter for the parameter subset in the HIV model (3.2).

N	n	Observables	min(ASE)	max(ASE)
1	51	x_5	$\text{ASE}(d_1) = 0.27$	$\text{ASE}(\lambda_E) = 4.27$
1	101	x_5	$\text{ASE}(d_1) = 0.19$	$\text{ASE}(\lambda_E) = 3.11$
1	201	x_5	$\text{ASE}(d_1) = 0.13$	$\text{ASE}(\lambda_E) = 2.27$
1	401	x_5	$\text{ASE}(d_1) = 0.094$	$\text{ASE}(\lambda_E) = 1.63$
2	51	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.095$	$\text{ASE}(x_5^0) = 2.31$
2	101	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.068$	$\text{ASE}(x_5^0) = 1.52$
2	201	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.047$	$\text{ASE}(x_5^0) = 1.26$
2	401	$x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_1) = 0.033$	$\text{ASE}(x_5^0) = 1.13$
3	51	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.045$	$\text{ASE}(x_5^0) = 2.26$
3	101	$x_6,\ x_1 + x_3,\ x_2 + x_4$	$\text{ASE}(\lambda_E) = 0.032$	$\text{ASE}(x_5^0) = 1.49$
3	201	$x_5,\ x_6,\ x_1 + x_3$	$\text{ASE}(\lambda_E) = 0.022$	$\text{ASE}(x_5^0) = 0.69$
3	401	$x_5,\ x_6,\ x_1 + x_3$	$\text{ASE}(\lambda_E) = 0.015$	$\text{ASE}(x_5^0) = 0.67$

3.3.3 HIV observables and time point selection results

When taking samples on a uniform time grid, D-, E-, and SE-optimal design criteria all choose observation operators that yield favorable ASE's, with some criteria performing best under certain circumstances. For example, the SE-optimal observables at $n = 401$, $N = 3$ yield the smallest standard errors; however, for all other values of n at $N = 3$, E-optimal performs best. At $N = 2$, E-optimal is a slightly weaker scheme. The examples in [16] also reveal that D-, E-, and SE-optimal designs are all competitive when only selecting time points for several different models. Now we wish to investigate the performance of these three criteria when selecting both an observation operator and a sampling time distribution using the algorithm described by equations (2.25) and (2.26).

To maintain consistency across trials while slightly simplifying the parameter estimation problem, we allow the set of six parameters and three initial conditions $\vec{\theta} = (\lambda_1,\ d_1,\ k_1,\ N_T,\ c,\ b_E,\ x_1^0,\ x_2^0,\ x_5^0)$ to be treated as unknowns and fix all other parameters. We again allow the possible observations of (1) infectious virus x_5, (2) immune response x_6, (3) CD4 cells $x_1 + x_3$, and (4) type 2 target cells $x_2 + x_4$, each with an assumed error variance of 5% of the initial variable values given by $\vec{x}^0$. To reflect the often limited resources of a clinical trial, we allow $N = 2$ observation maps to be included in the observation operators C and examine the distribution of time points if $n = 35$ or $n = 105$ samples (consisting of all observables in the observation operator) which may be taken from $t_0 = 0$ through $t_f = 1460$. We begin all simulations with uniformly spaced sample times, and use either time grid constraint C2 or C3. Both constraints assume that samples are taken at t_0 and t_f, so we in effect are optimizing the remaining 33 or 103 observation times.

When choosing $N = 2$ observables and distributing $n = 35$ time points using constraint C2, the D-optimal cost function yields the lowest ASE for the parameters $(\lambda_1,\ d_1,\ N_T,\ c,\ b_E,\ x_1^0)$, but the SE-optimals ASE are on the same order of magnitude (Table 3.9). Both the D- and SE-optimal cost functions selected the same observables, and they both were minimized by similar distributions of time points (Figure 3.27). The E-optimal design leads in the smallest ASE for $(k_1,\ x_2^0,\ x_5^0)$, with an $\text{ASE}(x_5^0)$ that is half that of D-optimal and SE-optimal. E-optimal, however, trails D- and SE-optimal in the accuracy of an estimate for b_E, with an $\text{ASE}(b_E)$ that is larger by one order of magnitude. The large difference in ASE's for these two parameters may be related to the observables selected by each design criterion: E-optimal design chooses x_5 (the variable most closely related to x_5^0) as an observable when D- and SE-optimal choose x_6 (whose right hand side contains b_E). All three also chose $x_2 + x_4$. The distribution of time points selected under each optimal design criterion focuses on regions very soon before or after a change in behavior of the selected observables (Figure 3.27). This may indicate that fewer sampling points may be needed for parameter estimates of similar accuracy. For the experimental design

Table 3.9: Approximate asymptotic standard errors calculated by asymptotic theory (2.21) using optimally spaced $n = 35$ time points under constraints C2 (left set of 3) and C3 (right set of 3) and optimally selected $N = 2$ observables for parameters of interest $\vec{\theta}$ of the HIV model (3.2). Smallest ASE per parameter is highlighted using bold font.

Time	C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**
Observe	$x_6,$	$x_5,$	$x_6,$	$x_6,$	$x_5,$	$x_6,$
	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$
ASE(λ_1)	**0.1077**	0.2501	0.1164	**0.1048**	0.1297	0.1215
ASE(d_1)	**0.0737**	0.1523	0.0972	**0.0683**	0.0759	0.0817
ASE(k_1)	0.0967	**0.0851**	0.1404	0.1085	**0.0920**	0.1174
ASE(N_T)	**0.2200**	0.2722	0.2263	**0.2257**	0.2784	0.2398
ASE(c)	**0.2455**	0.2916	0.2505	**0.2517**	0.3098	0.2650
ASE(b_E)	**0.0480**	0.7788	0.0500	**0.0476**	0.8182	0.0494
ASE(x_1^0)	**0.1148**	0.1275	0.1366	0.1201	**0.1081**	0.1292
ASE(x_2^0)	0.2657	**0.2456**	0.2597	0.2635	**0.2469**	0.2654
ASE(x_5^0)	0.9450	**0.4798**	0.9234	0.9303	**0.4629**	0.9473

constraints of 35 time point allocated using constraint C2, the D-optimal cost functional is best for most parameters, but E-optimal is best for some parameters.

The ASE calculated for choosing $N = 2$ observables and distributing $n = 35$ time points using constraint C3 are very similar (Table 3.9), and the selected observables remain unchanged; however, the optimal distribution of time points under each design criterion (Figure 3.28) is more uniform than when using constraint C2 (Figure 3.27). For small n with $N = 2$ observables under constraint C3, D-optimal design is best for most parameters, but E-optimal is best for the remaining few. While the choice of a particular time point distribution constraint does not greatly impact the calculated ASEs for the paramters of interest, the constraint C3 yields a measurement schedule that may be more feasible for a patient to follow in that it does not call for as many days of multiple observations.

Figure 3.27: Solution of the log-scaled HIV model (3.2) plotted using the observables $log(10^{x_1} + 10^{x_3})$ (upper left), $log(10^{x_2} + 10^{x_4})$ (upper right), x_5 (lower left), and x_6 (lower right). Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 35$ times under constraint C2.
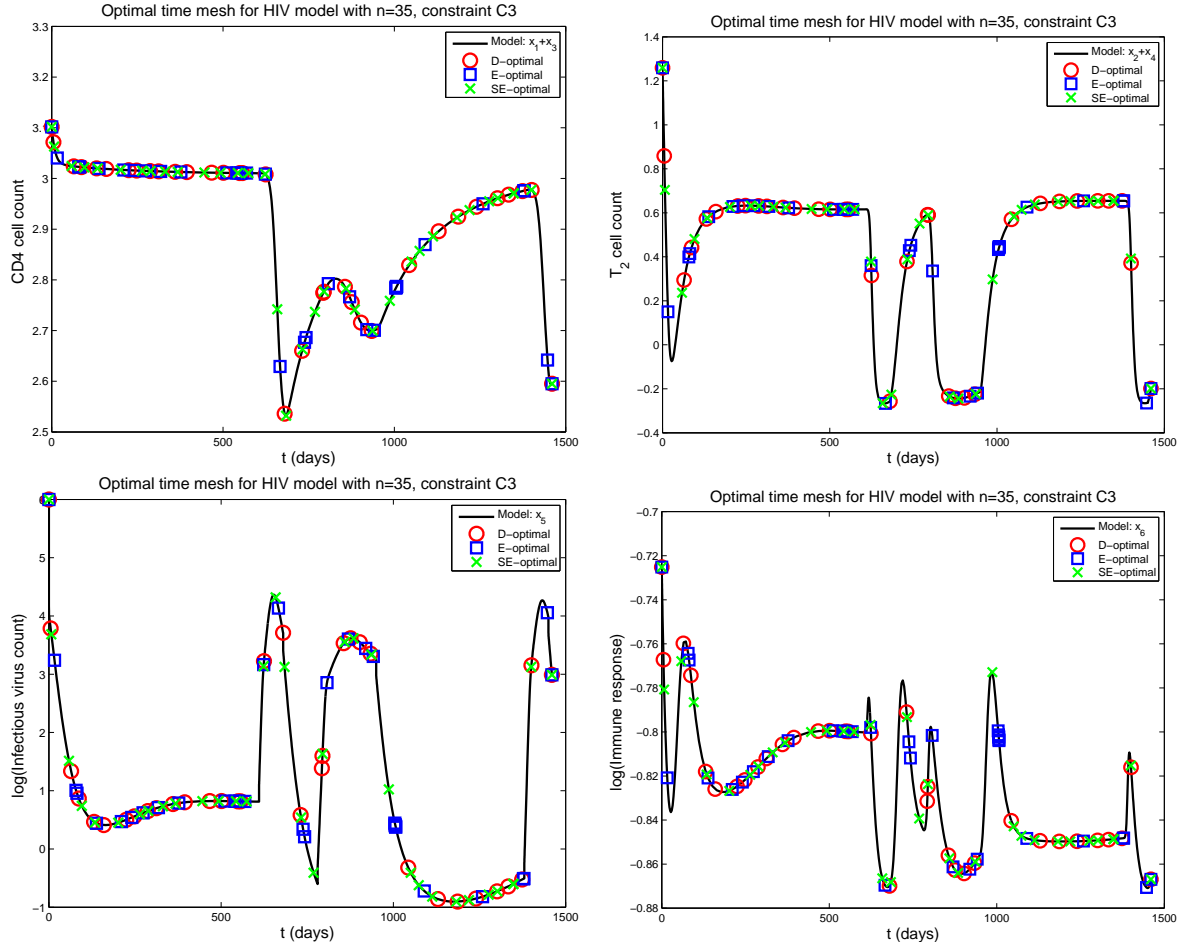
Figure 3.28: Solution of the log-scaled HIV model (3.2) plotted using the observables $log(10^{x_1} + 10^{x_3})$ (upper left), $log(10^{x_2} + 10^{x_4})$ (upper right), x_5 (lower left), and x_6 (lower right). Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 35$ times under constraint C3.

Choosing $N = 2$ observables and distributing $n = 105$ time points using constraint C2 leads to a different pattern in which optimal design criterion is best for which parameter. For the more dense time point distribution, SE-optimal design is a much stronger candidate against D- and E-optimal (Table 3.10). It yields the lowest ASE's for (λ_1, N_T, c, b_E), while E-optimal is best for $(k_1, x_1^0, x_2^0, x_5^0)$ and D-optimal is best for (d_1, b_E). The D- and SE-optimal cost functions again choose the same observables of x_6 and $x_2 + x_4$, so their ASE's for all parameters are similar. In this scenario, the E-optimal design has the largest percent reduction in ASE from those of D- and SE-optimal for the parameters x_1^0 and x_5^0. E-optimal also changed its selected observables to $x_1 + x_3$ and $x_2 + x_4$. The distribution of time points for the D-optimal design criteria appear very close to uniform (Figure 3.29), and the distributions for E- and SE-optimal are clustered near local maxima, local minima, and other large changes in behavior of the observables; however, even slight differences between the distributions of the E- and SE-optimal costs functions are visible. Consider the graph of infectious virus count in Figure 3.29. There is a cluster of SE-optimal times between $t = 1100$ and $t = 1250$, but no E-optimal times occur during that interval. Very soon after, between $t = 1250$ and $t = 1400$, there is a cluster of E-optimal times but no SE-optimal times. These clusters may indicate that either these periods in the patient's treatment history are key to characterizing the parameters or that fewer time points may be adequate to obtain sufficiently accurate parameter estimates.

The strength of SE-optimal design for large n does not hold for time grid constraint C3 (Table 3.10). E-optimal design provides the lowest ASE for the parameters $(\lambda_1, d_1, k_1, x_1^0)$, D-optimal is best for $(N_T, c, b_E, x_2^0, x_5^0)$, and the ASE calculated using the SE-optimal designed experiment is often the largest. The observables selected in this case are the same as the $n = 105$, constraint C2 case. The optimal time point distributions determined under all three design criterion are near uniform (Figure 3.30). Small differences between the distributions may be observed when the functions have a slope of high magnitude, indicating that the distributions are not exactly uniform. As the time point optimization routine is started with a uniform time point distribution, this may indicate that when samples are taken at many times during an experiment, a uniform distribution is somewhat near optimal for model (3.2) or, more significantly, that a uniform distribution is not an appropriate initial distribution to use to obtain the optimal time point distribution.

Between D- and E-optimal, there is no clear leader for optimal time point and observable selection in these examples for system (3.2). The SE-optimal criterion is useful in estimating some parameters under the constraint C2, and in most cases yields standard errors on the same order of magnitude of the leading optimal design criterion. The optimal design algorithm successfully identifies observables and a time distribution that optimizes an aspect of the Fisher Information Matrix as determined by the choice of optimal design cost functional; however, the time point distributions exhibit tendencies to cluster near times when the slopes of function

Table 3.10: Approximate asymptotic standard errors calculated by asymptotic theory (2.21) using optimally spaced $n = 105$ time points under constraints C2 (left set of 3) and C3 (right set of 3) and optimally selected $N = 2$ observables for parameters of interest $\vec{\theta}$ of the HIV model (3.2). Smallest ASE per parameter is highlighted using bold font.

Time	C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**
Observe	$x_6,$	$x_1 + x_3,$	$x_6,$	$x_6,$	$x_1 + x_3,$	$x_6,$
	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$	$x_2 + x_4$
ASE(λ_1)	0.0687	0.0618	**0.0607**	0.0674	**0.0475**	0.0729
ASE(d_1)	**0.0429**	0.0541	0.0470	0.0441	**0.0395**	0.0475
ASE(k_1)	0.0658	**0.0360**	0.0681	0.0629	**0.0388**	0.0739
ASE(N_T)	0.1599	0.1437	**0.1394**	**0.1192**	0.1351	0.1423
ASE(c)	0.1785	0.1591	**0.1551**	**0.1325**	0.1500	0.1578
ASE(b_E)	**0.0281**	0.4560	**0.0281**	0.0266	0.4637	0.0299
ASE(x_1^0)	0.0810	**0.0451**	0.0793	0.0572	**0.0490**	0.0818
ASE(x_2^0)	0.2537	**0.2004**	0.2296	**0.1823**	0.2244	0.2330
ASE(x_5^0)	0.7042	**0.4384**	0.6825	**0.0528**	0.4491	0.6571

solutions are changing. We proceed to examine the algorithm's performance when presented with a system that allows many possible observables and does not require as many sampling times.
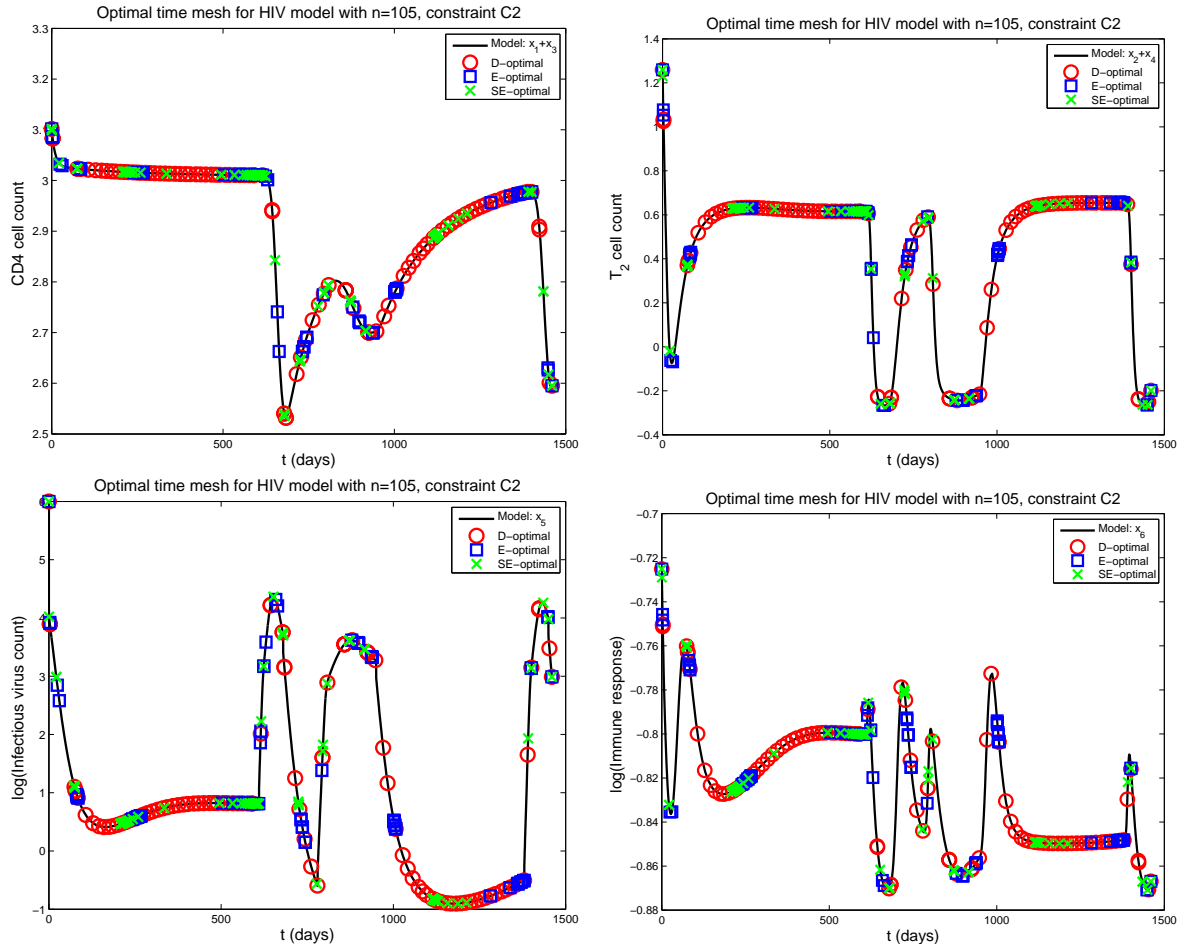
Figure 3.29: Solution of the log-scaled HIV model (3.2) plotted using the observables $log(10^{x_1} + 10^{x_3})$ (upper left), $log(10^{x_2} + 10^{x_4})$ (upper right), x_5 (lower left), and x_6 (lower right). Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 105$ times under constraint C2.
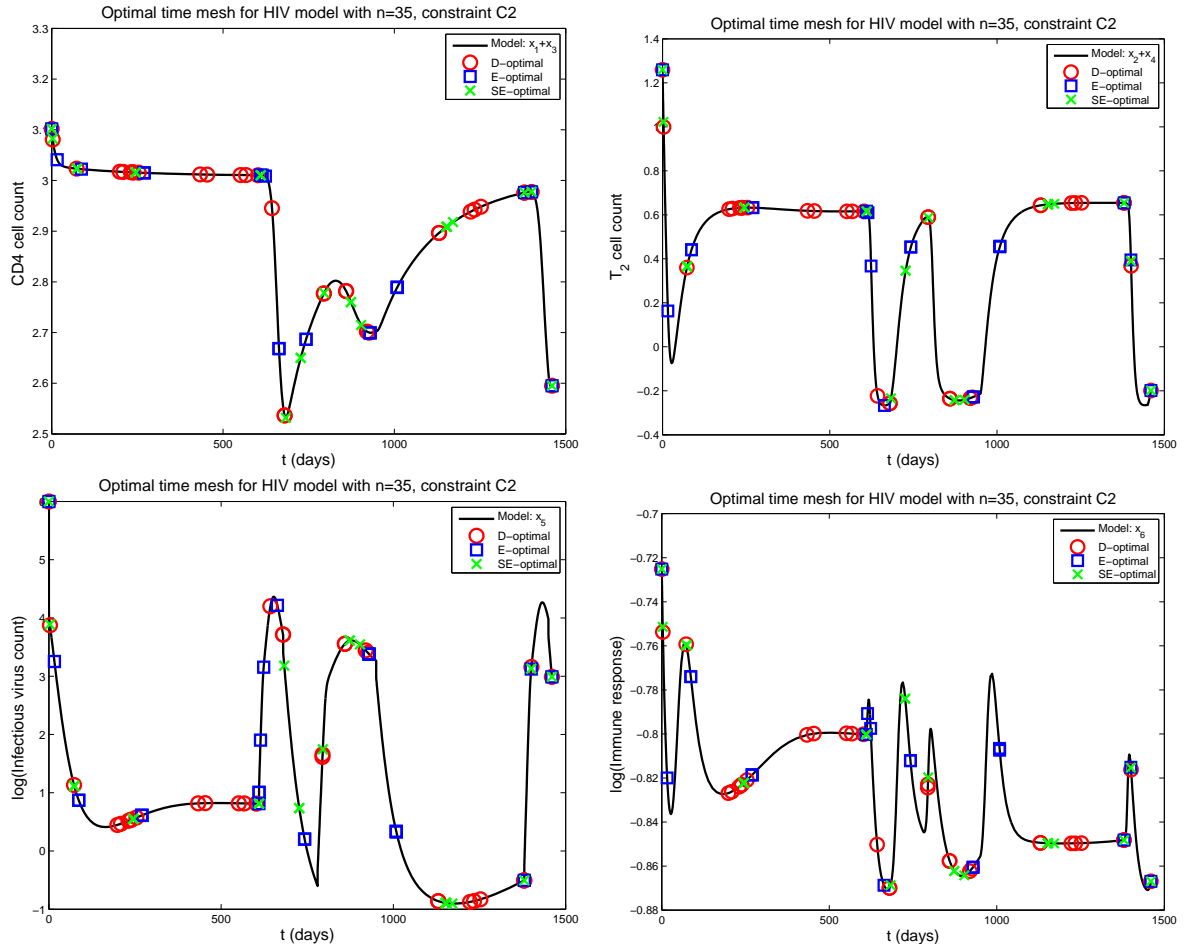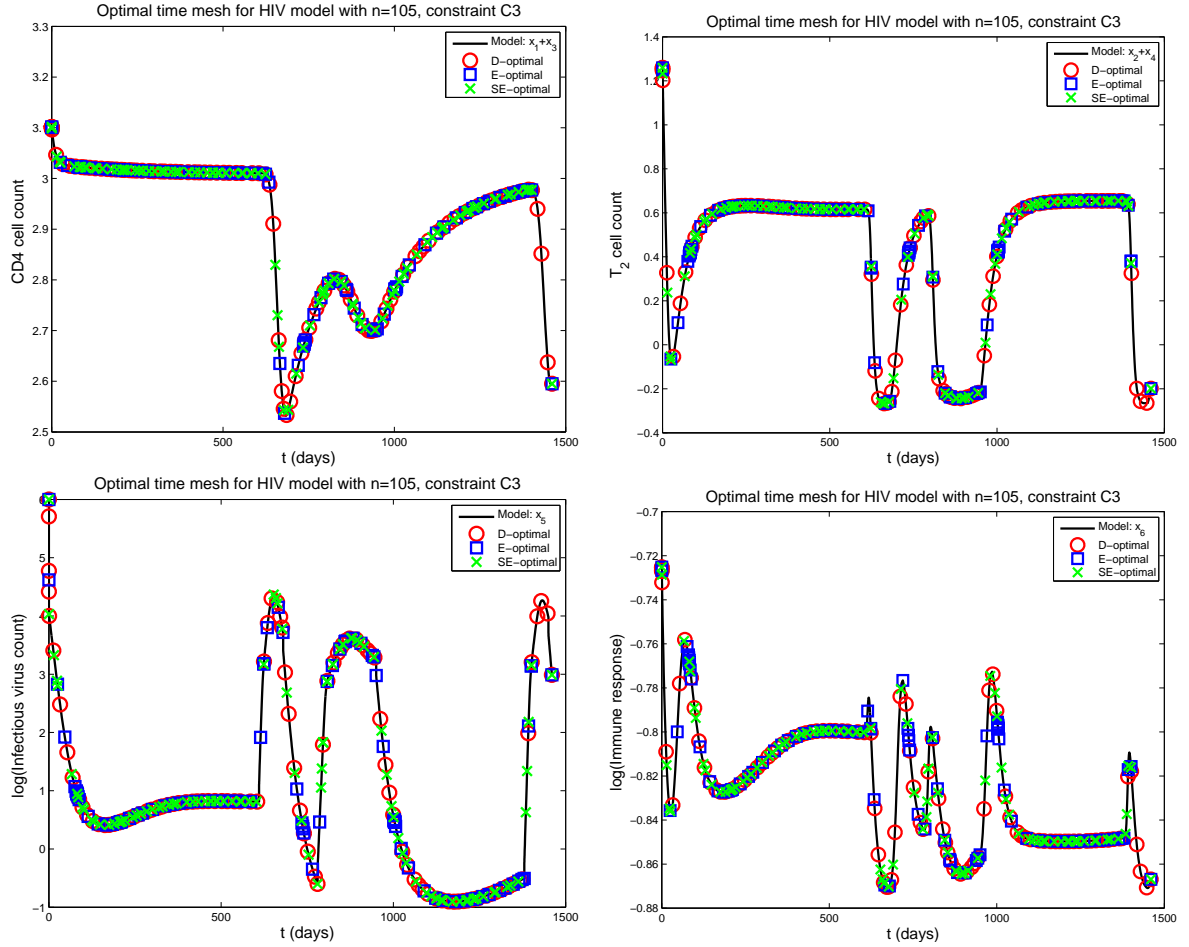
Figure 3.30: Solution of the log-scaled HIV model (3.2) plotted using the observables $log(10^{x_1} + 10^{x_3})$ (upper left), $log(10^{x_2} + 10^{x_4})$ (upper right), x_5 (lower left), and x_6 (lower right). Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 105$ times under constraint C3.

3.3.4 C3 Cycle model [49]

The second model we use as an example is characteristic of the large differential equation systems that often appear in industrial problems as well as biological systems. In [49], Zhu et al., present an ODE model for the Calvin Cycle in fully grown spinach. This model contains 165 parameters and initial conditions, 31 ODEs, and 7 concentration balance laws and involves 38 state variables (metabolite concentrations in different parts of the cell) as well as a calculation for photosynthetic CO_2 uptake rate. The metabolites used in the model are denoted by RuBP, PGA, DPGA, T3P, FBP, E4P, S7P, ATP, SBP, NADPH, HexP, PenP, NADHc, NADc, ADPc, ATPc, GLUc, KGc, ADP in photorespiration, ATP in photorespiration, GCEA, GCA, PGCA, GCAc, GOAc, SERc, GLYc, HPRc, GCEAc, T3Pc, FBPc, HexPc, F26BPc, UDPGc, UTPc, SUCP, SUCc, and PGAc, and the parameters are mainly initial conditions, maximum reaction velocities, and Michaelis-Menten constants for reaction substrates, products, activators, and inhibitors. The 'c' following some metabolite names indicates that the model compartment corresponds to the metabolite concentration in cytosol; compartment names lacking a 'c' are the metabolite in the chloroplast stroma. The full system of equations and parameter values may be found in the appendices of [49]. While the model has not been validated with data as completely as the family of HIV models discussed above, it is representative of the models used to describe plant enzyme kinetics and utilizes well-documented Michaelis-Menten enzyme kinetic model formulations [7].

In [6], sets of the optimal 3, 5, 10, and 15 metabolites are identified as the most useful to measure in an experiment in order to estimate a subset of 6 parameters, $\vec{\theta}_a = [\text{KM}_{11}, \text{KM}_{521}, \text{KI}_{523}, \text{KC}, \text{KM}_{1221}, \text{KM}_{1241}]^\mathsf{T}$ with true values $\vec{\theta}_{a0} = [0.0115, 0.0025, 0.00007, 0.0115, 0.15, 0.15]^\mathsf{T}$, and a subset of 18 parameters, $\vec{\theta}_c = [\text{RuBP}_0, \text{SBP}_0, \text{KM}_{11}, \text{KM}_{13}, \text{KI}_{13}, \text{KE}_4, \text{KM}_9, \text{KM}_{131}, \text{KI}_{135}, \text{KE}_{22}, \text{KM}_{511}, \text{KM}_{521}, \text{KI}_{523}, \text{KC}, \text{KM}_{1221}, \text{KM}_{1241}, \text{V}_9, \text{V}_{58}]^\mathsf{T}$ with true values $\vec{\theta}_{c0} = [2, 0.3, 0.0115, 0.02, 0.075, 0.05, 0.05, 0.05, 0.4, 0.058, 0.02, 0.0025, 0.00007, 0.0115, 0.15, 0.15, 0.3242, 0.0168]^\mathsf{T}$. The simulation was set in the framework of an experiment run for 3000 seconds over which 11 samples are taken at evenly spaced times for the optimization of estimates for $\vec{\theta}_a$ and 21 samples for estimation of $\vec{\theta}_c$.

We use a similar set up for our simulations in order to judge the ability of the time and variable selections to minimize the asymptotic standard errors of both $\vec{\theta}_a$ and $\vec{\theta}_c$ using $N = 5$ and $N = 10$ observables and $n = 11$ time points over the time interval $t \in [0, 3000]$. Each metabolite is assigned a variance of 5% of its initial value, and to reduce computation time, all metabolites that do not change in concentration over time ($dx/dt = 0$) are excluded from the search algorithm. As it is often possible to measure a particular metabolite's concentration in plant tissue, we allow $\mathcal{C}_{N*}$ to be composed of 28 vectors in $\mathbb{R}^{1 \times 40}$ that are composed of a one in the element corresponding to the position of the differential equation describing the dynamics

of the metabolite in the vector of model ODEs and zero elsewhere.

We compare the ASE calculated for $n = 11$ evenly spaced measurements over the interval $t \in [0, 3000]$ against the optimally spaced $n = 11$ time points according to constraints C2 and C3 and plot the D-, E- and SE-optimal time point distributions on the solutions of selected compartments in the Zhu model of [49]. The four compartments chosen for graphing are carbon uptake rate $A(t)$ (units μmol m^{-2}s^{-1}), which is a rate calculated from the output of the model that describes the plant's efficiency in using the available environmental resources to develop; adenosine triphosphate (ATP) in chloroplast stroma, which is a well-known metabolite active in photosynthesis; sucrose, a sugar, in cytosol (SUCc); and Ribulose-1,5-bisphosphate (RuBP), a metabolite connected to the enzyme RuBisCO, which is essential to carbon fixation.

The simplest scenario we test is the selection of the optimal $N = 5$ observables (metabolites) and $n = 11$ time points to use when estimating the six parameters in $\vec{\theta}_a$. For all three optimal design methods, the optimal observables determined under the uniform grid were also determined to be optimal after the time point distribution is optimized under constraints C2 and C3. These observables are listed in the upper portion of Table 3.11. All three optimal design methods identify the observables PGA, SERc, and F26BPc, indicating that these three metabolites may be central to accurate estimates of the parameters in $\vec{\theta}_a$; moreover, the E-optimal and SE-optimal cost functions selected the same set of five observables.

The similarity in results, however, does not continue through the selected time point distributions. Under constraint C2 (Figure 3.31), the D-optimal distribution is loosely clustered about the center of the time interval, the E-optimal time points are clustered about $t = 250$ seconds, and SE-optimal chooses a small cluster of time points around the initial bump and allows a few samples after the function reaches a steady state. Using constraint C3 (Figure 3.32), all three optimal design criterion chose a majority of their sampling times before $t = 600$ and allow only a few sampling times after the system reaches a steady state. The optimization of time point distributions yields improved asymptotic standard errors from those of the uniform distribution - sometimes by an order of magnitude or more. For all three time point constraints, D-optimal yields the smallest ASE's for the most number of parameters, and SE-optimal yields the smallest ASE's for the second most number of parameters. Both optimal design criteria perform better with the C3 optimal times than the C2 optimal times. Therefore, in this simple case, using either the D- or SE-optimal design criterion with time point distribution constraint C3 would yield the best results.

The next scenario is the selection of the optimal $N = 10$ observables and $n = 11$ time points to use when estimating the six parameters in $\vec{\theta}_a$. For all three optimal design methods, the optimal observables determined under the uniform grid were also determined to be optimal after the time point distribution is optimized under constraints C2 and C3. These observables are listed in the upper portion of Table 3.12. Of the 28 possible observables, 20 were selected by

Table 3.11: Top: Optimal 5 observables chosen by each optimal design criterion when estimating the parameters $\vec{\theta}_a$ of the model in [49]. Bottom: Approximate asymptotic standard errors calculated using asymptotic theory (2.21) for each parameter in $\vec{\theta}_a$ using the optimal 5 observables with time point constraints of uniform spacing, constraint C2, and constraint C3. Smallest ASE for each parameter per time point constraint is highlighted in bold font.

Method	Observables
D-opt	PGA, T3P, GOAc, SERc, F26BPc
E-opt	PGA, SERc, T3Pc, FBPc, F26BPc
SE-opt	PGA, SERc, T3Pc, FBPc, F26BPc

Time	Uniform			C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**	**D**	**E**	**SE**
ASE(KM_{11})	**0.001**	0.036	0.036	0.002	0.003	**8.7e-4**	**5.8e-4**	0.002	0.002
ASE(KM_{521})	4.181	**4.009**	**4.009**	0.132	**0.108**	0.681	0.075	0.135	**0.053**
ASE(KI_{523})	0.117	**0.113**	**0.113**	0.004	0.003	**0.019**	**0.001**	0.004	0.002
ASE(KC)	**0.001**	0.035	0.036	**0.001**	0.003	0.002	**0.001**	0.002	0.002
ASE(KM_{1221})	**0.281**	1.307	1.307	**0.274**	0.433	0.580	**0.146**	0.513	0.514
ASE(KM_{1241})	**0.251**	1.176	1.176	**0.245**	0.387	0.518	**0.131**	0.458	0.459

at least one optimal design criterion, 8 of of which (DPGA, T3P, E4P, S7P, ATP, GCA, SERc, T3Pc) were selected by two criteria and one, GOAc, was selected by all three. The effect of adding five observables does not have a large effect on the estimated ASE for the six parameters of interest. The ASE's listed in Table 3.12 for the $N = 10$ observable case are on the same order as those in Table 3.11 for the $N = 5$ observable case for each time point constraint.

The optimal time point distributions also differ from those selected under the five observable case. Under constraint C2 (Figure 3.33), the D-optimal distribution is loosely clustered near the boundaries of the time interval while the E- and SE-optimal time points are scattered over the whole interval. Using constraint C3 (Figure 3.34), the D-optimal times are spread throughout the middle half of the time interval, E-optimal over the first half, and SE-optimal over the second half. While there are no easily discernible patterns between the distributions chosen under the C2 and C3 constraints, the optimization of time point distributions again yield improved asymptotic standard errors from those of a uniform time distribution. For all three time point constraints, D-optimal yields the smallest ASE's for the most number of parameters, and SE-optimal yields the smallest ASE's for the second most number of parameters. Other than for the parameters it estimates best, SE-optimal is often the worst criterion. The time point distribution under constraint C2 allows smaller ASE's for the $N = 10$ observable variables case.

Table 3.12: Top: Optimal ten observables chosen by each optimal design criterion when estimating the parameters $\vec{\theta}_a$ of the model in [49]. Bottom: Approximate asymptotic standard errors calculated using asymptotic theory (2.21) for each parameter in $\vec{\theta}_a$ using the optimal 10 observables with time point constraints of uniform spacing, constraint C2, and constraint C3. Smallest ASE for each parameter per time point constraint is highlighted in bold font.

Method	Observables
D-opt	RuBP, PGA, T3P, E4P, SBP, GOAc, SERc, GLYc, T3Pc, F26BPc
E-opt	DPGA, T3P, S7P, ATP, NADPH, NADHc, GCEA, GCA, GOAc, HexPc
SE-opt	DPGA, FBP, E4P, S7P, ATP, HexP, GCA, GOAc, SERc, T3Pc

Time	Uniform			C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**	**D**	**E**	**SE**
ASE(KM_{11})	**0.001**	0.001	0.001	**5.9e-4**	0.001	0.001	**0.001**	0.001	0.001
ASE(KM_{521})	4.043	**3.971**	**3.971**	0.110	0.782	**0.059**	0.147	0.272	**0.059**
ASE(KI_{523})	0.113	**0.111**	**0.111**	0.003	0.022	**0.002**	0.004	0.008	**0.002**
ASE(KC)	**0.001**	0.001	0.001	**9.1e-4**	0.001	0.001	**0.001**	0.001	0.001
ASE(KM_{1221})	**0.280**	0.674	0.674	**0.253**	0.621	0.581	**0.305**	0.614	0.696
ASE(KM_{1241})	**0.250**	0.603	0.603	**0.226**	0.555	0.519	**0.272**	0.548	0.623

Therefore using the D-optimal design criterion with time point distribution constraint C2 would yield the best results.

The addition of five observables for $\vec{\theta}_a$ does not greatly impact most of the calculated ASE's. For the time point constraints C2 and C3, all but one of the minimum ASE's remain on the same order of magnitude, and some of the ASE's increase when $N = 10$ observables are allowed from when $N = 5$ observables are allowed. This may indicate that for a small parameter set such as $\vec{\theta}_a$, only a small amount of information is needed to obtain the best possible results; adding extra information (through additional observables) does not further improve the ASE's.
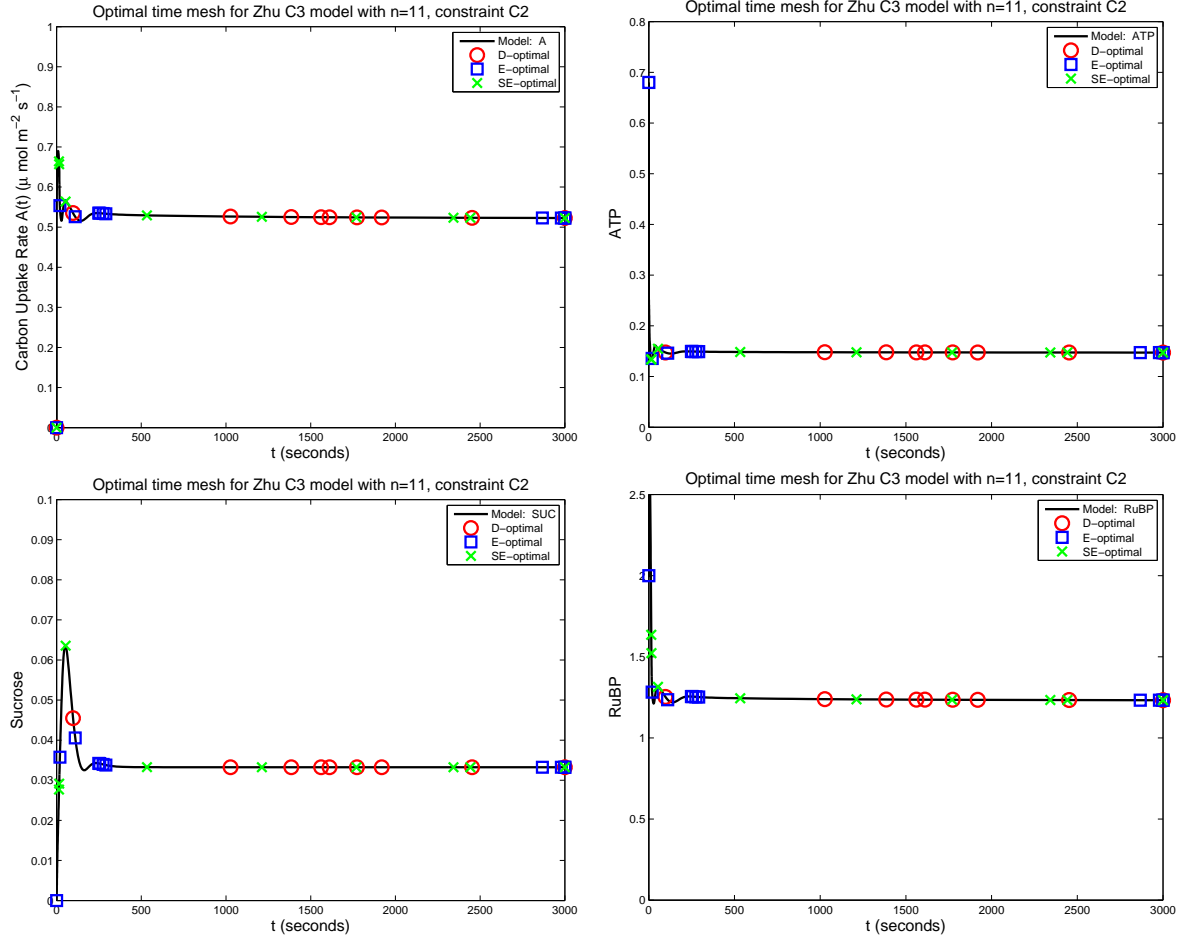
Figure 3.31: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C2** when sampling the optimal five observables to estimate $\vec{\theta}_a$ (Table 3.11). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
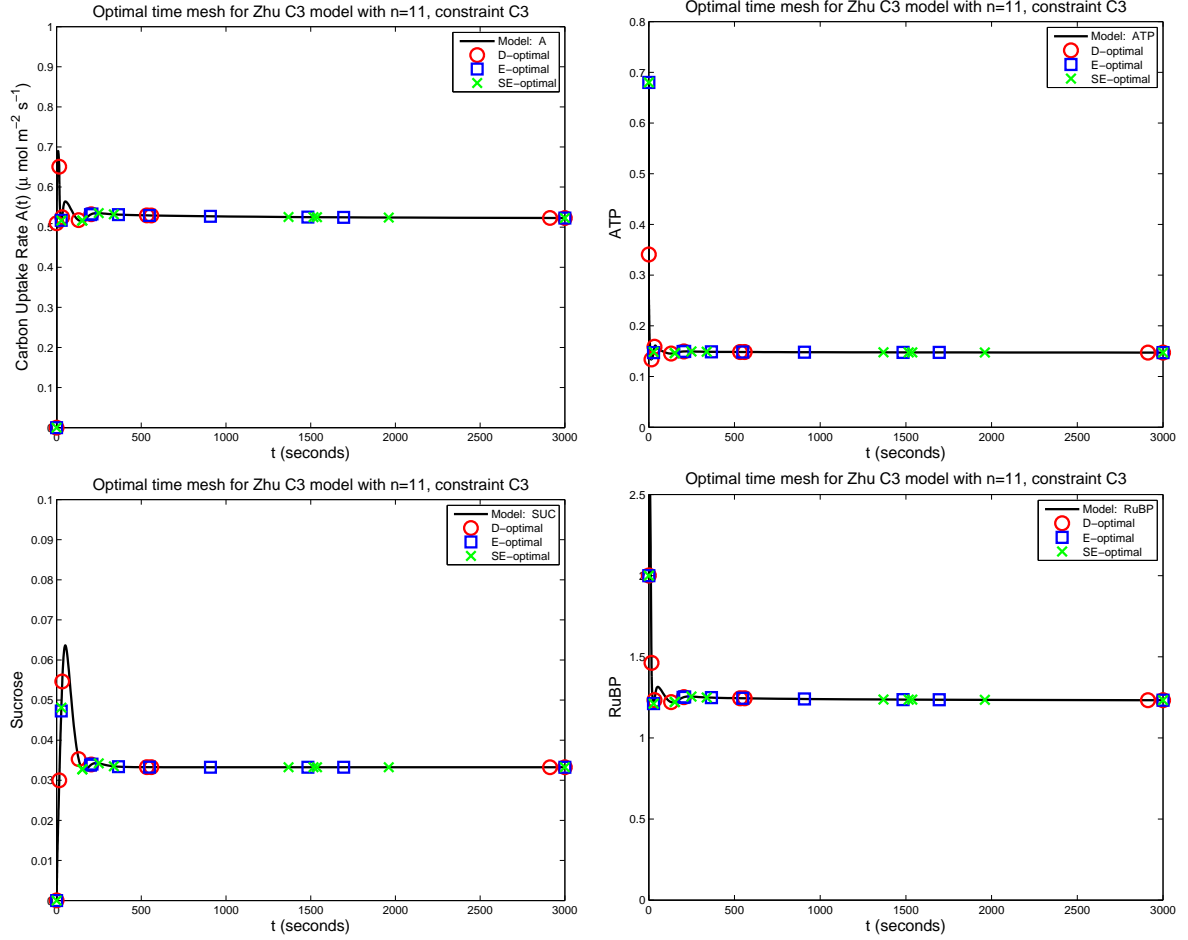
Figure 3.32: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C3** when sampling the optimal five observables to estimate $\vec{\theta}_a$ (Table 3.11). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
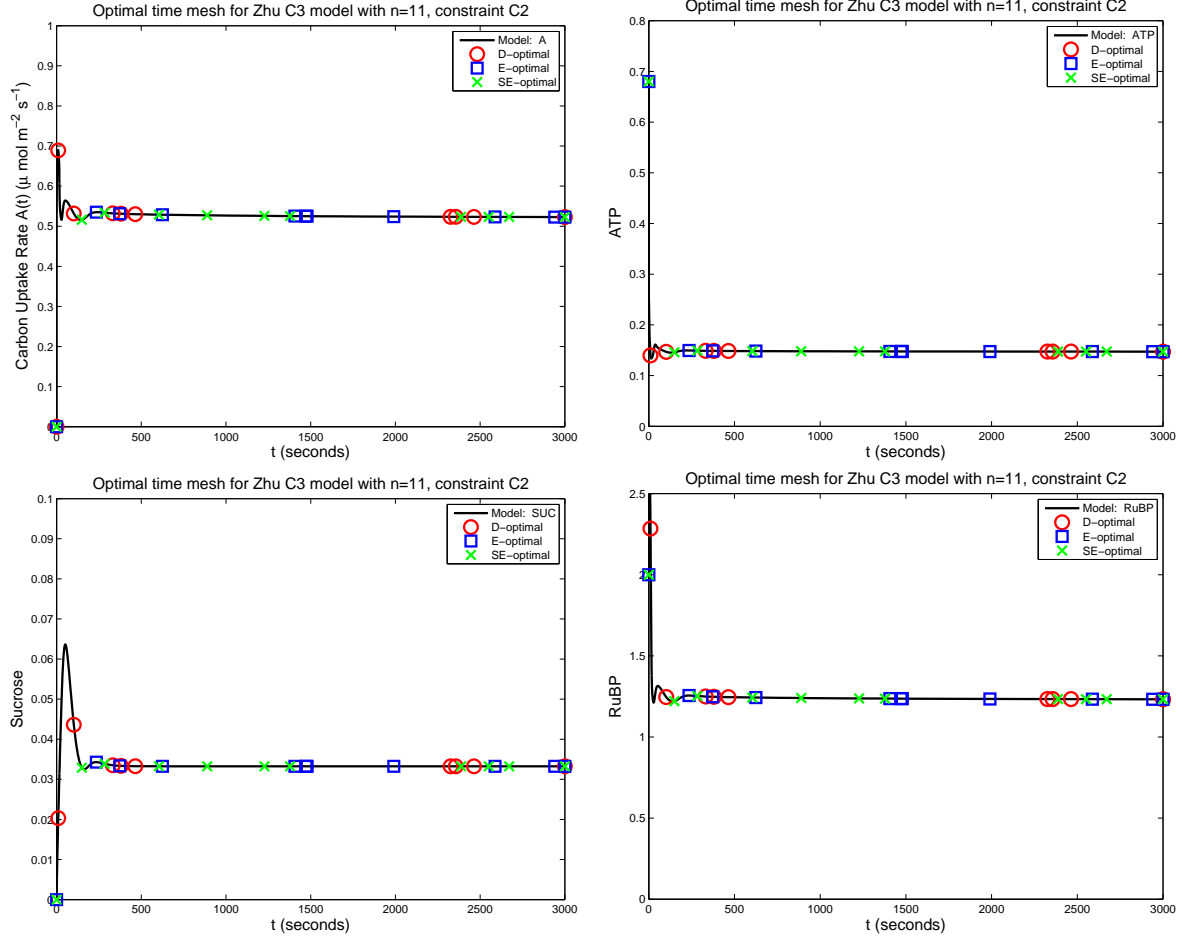
Figure 3.33: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C2** when sampling the optimal ten observables to estimate $\vec{\theta}_a$ (Table 3.12). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
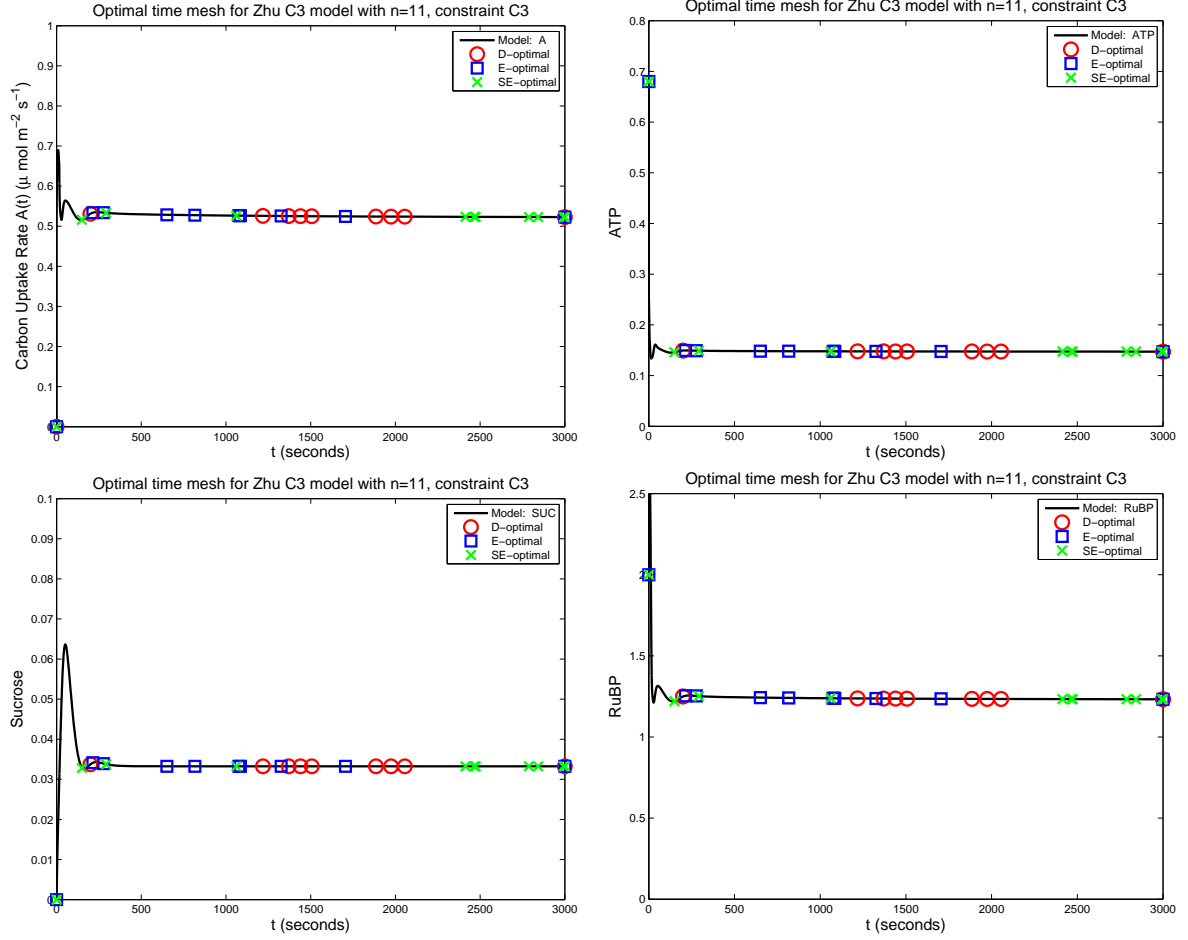
Figure 3.34: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C3** when sampling the optimal ten observables to estimate $\vec{\theta}_a$ (Table 3.12). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.

Now consider the parameter vector $\vec{\theta}_c$. We identify the optimal $N = 5$ observables and the optimal distribution of $n = 11$ time points using the D-, E-, and SE-optimal cost functionals. Three metabolites, RuBP, SBP, and F26BPc, are selected by all three cost functionals, indicating that these three observables may be important to measure to accurately estimate a large number of the parameters in $\vec{\theta}_c$. The inclusion of $RuBP_0$ and SBP_0 in $\vec{\theta}_c$ may also heavily influence the selection of observables. Many of the other metabolites selected by at least one of the cost functionals, namely, PGA, SERc, and T3Pc, were also selected for $\vec{\theta}_a$ in the experimental setup of $N = 5$ observables and $n = 11$ time points, indicating that they are still relevant to some of the parameters in $\vec{\theta}_c$.

Using the uniform time point distribution yields standard errors that are typically two or three orders of magnitude larger than the parameter value (Table 3.13). For example, the value of the inhibition constant KI_{523} is 0.00007 (from Section A.2), but the calculated asymptotic standard error is 9.295, which is six orders of magnitude larger than the parameter. Constraint C2 also leads to large standard errors when using the D- and E-optimal cost function. The E-optimal standard errors, in fact, do not change from those calculated using the uniform time point distribution, and as shown in Figure 3.35, the E-optimal time distribution under constraint C2 is uniform. The SE-optimal design criterion performs the best when using constraint C2 and produces much smaller standard errors, though the standard errors are still larger than the parameter values by orders of magnitude for most parameters. Both D- and SE-optimal perform favorably under constraint C3: the D-optimal, C3 design produces smaller ASE's than the SE-optimal, C2 design, and the SE-optimal, C3 design yields the smallest ASE's for almost all parameters. Under the constraint C3 and the SE-optimal cost functional, the calculated ASE's are zero to one orders of magnitude larger than the parameter values.

The favorable performance of the SE-optimal design criterion under constraint C2 may be attributed to its selected time point distribution (Figure 3.35). The SE-optimal distribution is the only one that places a large number of time points (5 out of 11) before the solution approaches a steady state at $t = 250$. The E-optimal design, as mentioned previously, selects the uniform distribution, and the D-optimal design suggests sampling after the solution has achieved a steady state. When using constraint C3, the E-optimal design selects a distribution that maintains the uniform spacing between points in the interior of the time interval (0,3000) (Figure 3.36), again leading to poor ASE's. Both the D- and SE-optimal designs place the majority of the 11 time points in the first 1/3 of the time interval, thus capturing more of the dynamical system's behavior before it achieves a steady state. These front-heavy distributions greatly reduce the estimated ASE's from those calculated using a uniform time distribution.

The last case tested is designing experiments using each of the three cost functionals to estimate $\vec{\theta}_c$ when allowed $N = 10$ observables and $n = 11$ time points. As in the case of only allowing five observables for estimation of $\vec{\theta}_a$, the metabolites RuBP, SBP, and F26BPc are

Table 3.13: Top: Optimal five observables chosen by each optimal design criterion when estimating the parameters $\vec{\theta}_c$ of the model in [49]. Bottom: Approximate asymptotic standard errors calculated using asymptotic theory (2.21) for each parameter in $\vec{\theta}_c$ using the optimal 5 observables with time point constraints of uniform spacing, constraint C2, and constraint C3. Smallest ASE for each parameter per time point constraint is highlighted in bold font.

Method	Observables
D-opt	RuBP, PGA, SBP, SERc, F26BPc
E-opt	RuBP, SBP, PenP, SERc, F26BPc
SE-opt	RuBP, SBP, GCEAc, T3Pc, F26BPc

Time	Uniform			C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**	**D**	**E**	**SE**
$ASE(RuBP_0)$	0.316	0.316	0.316	0.316	0.316	**0.224**	0.224	0.316	**0.222**
$ASE(SBP_0)$	0.123	0.123	0.123	0.123	0.123	**0.087**	0.087	0.123	**0.086**
$ASE(KM_{11})$	**10.8**	17.1	15.7	12.9	17.1	**0.066**	0.037	20.6	**0.010**
$ASE(KM_{13})$	**69.7**	116	88.1	68.5	116.0	**1.184**	0.108	147.7	**0.070**
$ASE(KI_{13})$	**39.7**	45.0	71.9	53.5	45.0	**4.827**	2.939	56.7	**0.974**
$ASE(KE_4)$	23.3	24.0	**15.1**	12.9	24.0	**0.984**	0.178	48.8	**0.063**
$ASE(KM_9)$	11.4	19.7	**8.765**	3.566	19.7	**1.528**	0.321	25.6	**0.110**
$ASE(KM_{131})$	109.6	**7.831**	153.8	100.1	7.83	**1.118**	0.368	7.985	**0.140**
$ASE(KI_{135})$	1310	**147.8**	2470	1140	147.9	**13.8**	4.491	115.4	**1.793**
$ASE(KE_{22})$	**26.2**	47.4	27.4	21.9	47.4	**0.179**	0.094	40.9	**0.045**
$ASE(KM_{511})$	4.40	4.937	**0.159**	1.710	4.937	**0.005**	0.163	9.898	**0.002**
$ASE(KM_{521})$	331	402.7	**18.6**	396.0	402.7	**8.415**	6.177	377.7	**0.120**
$ASE(KI_{523})$	9.295	11.3	**0.523**	11.1	11.3	**0.236**	0.174	10.6	**0.003**
$ASE(KC)$	**1.095**	1.520	4.941	0.957	1.520	**0.556**	0.079	2.445	**0.029**
$ASE(KM_{1221})$	**40.1**	61.5	621.2	49.8	61.5	**14.2**	**0.738**	84.0	2.069
$ASE(KM_{1241})$	**36.0**	55.3	557.9	44.7	55.3	**12.9**	**0.664**	75.5	1.829
$ASE(V_9)$	**28.3**	48.0	38.8	7.979	48.0	**1.120**	0.149	71.9	**0.137**
$ASE(V_{58})$	0.588	1.025	**0.159**	0.831	1.025	**0.087**	0.012	1.784	**7.8e-4**

selected by all three optimal design criteria. Additionally, PenP and SERc are selected by all three. A total of 15 different metabolites are chosen across the three optimal design criteria, of which only five are chosen by a single criterion and five selected by two criteria (Table 3.14). The three optimal design criteria show more agreement on which 10 observables are important to $\vec{\theta}_c$ than which 10 are important to $\vec{\theta}_a$ (in that case, a total of 20 different metabolites are chosen). This indicates that the three design criterion agree upon observables that are central to understanding the behavior of the model but also select observables that help minimize their associated cost functionals.

Similar to the 5 observable case for $\vec{\theta}_c$, the ASE's calculated when using a time point distribution optimized under the C2 or C3 constraints are smaller than those of the uniform time point distribution (though the difference is much less pronounced). While the ASE's from the D- and SE-optimal criteria are typically smallest under the uniform time point distribution, the E-optimal designed experiment yields the smallest ASE's for the most number of parameters under constraint C2 (Table 3.14). The ASE's for the D-optimal experiment under constraint C2 are typically larger than those for the D-optimal experiment with a uniform time distribution, and the SE-optimal ASE's see only marginal improvements. For parameters when the E-optimal ASE's are not as small as those for the D- and SE-optimal designs, the ASE's are often still on the same order. These ASE's are still orders of magnitude larger than the values of some parameters. The minimum ASE's under constraint C3 are smaller than those under constraint C2 (and are notable improvements over those from the uniform time point distribution), but most are improved by only one order of magnitude or less. The D- and SE-optimal cost functions yield the smallest standard errors for more parameters under C3 than C2, as well.

The performance improvements seen in the D- and SE-optimal experimental designs from the C2-optimal time point distributions to the distributions optimal under constraint C3 can be easily explained. Under constraint C2, neither criterion produced clusters of time points (Figure 3.37). The distributions generated by both criteria are more similar to the uniform distribution that was used as the initial seed for the time point distribution optimization than to the highly clustered distributions seen for other selections of n and N (such as the one shown in Figure 3.36). The D-optimal distribution only contains one time point before $t = 500$, as well, thus ignoring the early dynamics of the system. The E-optimal time point distribution contains clusters at the very beginning ($t < 200$) and very end ($t > 2900$) of the time interval as well as a loose cluster between $t = 1500$ and $t = 2000$.

All three optimal design criteria produce non-uniform distributions under constraint C3 (Figure 3.38). The time point distributions determined using the E- and SE-optimal cost functions place a majority of their time points before $t = 1000$, capturing the dynamics early in the system solution. The D-optimal distribution also concentrates its sampling times before $t = 1000$, but does not as tightly cluster the time points. All three cost functionals under time

80

point constraint C3 produce smaller ASE's than those calculated with a uniform time point distribution (Table 3.14). The distributions generated under both constraints C2 and C3 indicate that heavily weighting time periods when the system is changing its pattern of behavior (such as transitioning from oscillations to a steady state) produce smaller standard errors than uniform or near-uniform distributions.

Our computations using the Zhu model of [49] for the Calvin cycle indicate that both the selection of observables and the distribution of sampling times across the experiment affect the estimated asymptotic standard errors – sometimes by several orders of magnitude. While the ASE's calculated for all exercises were often larger than the parameter values, this may be related to the model's tendency to quickly reach a steady state.

Overall, the optimal design cost functions performed best under time point constraint C3, followed by C2 and finally the uniform distribution. While in some examples a particular cost functional would do particularly well (such as D-optimal in the $\vec{\theta}_a$, 10 observable case and SE-optimal in the $\vec{\theta}_c$, 5 observable case), it is hard to determine apriori – or even using the results of a uniform time distribution – which cost functional will perform best for a particular set of parameters and observables. The reduction in ASE's gained from adding additional observables is also difficult to predict. Adding five observables to better estimate $\vec{\theta}_a$ did not result in noticeably smaller estimated asymptotic standard errors as calculated by asymptotic theory. The improvement may be better quantified using a different method to calculate asymptotic standard errors such as Monte Carlo simulations or bootstrapping, techniques that may possibly be implemented as future work.

Table 3.14: Top: Optimal ten observables chosen by each optimal design criterion when estimating the parameters $\vec{\theta}_c$ of the model in [49]. Bottom: Approximate asymptotic standard errors calculated using asymptotic theory (2.21) for each parameter in $\vec{\theta}_c$ using the optimal 10 observables with time point constraints of uniform spacing, constraint C2, and constraint C3. Smallest ASE for each parameter per time point constraint is highlighted in bold font.

Method	Observables
D-opt	RuBP, T3P, SBP, PenP, GCEA, GOAc, SERc, GLYc, T3Pc, F26BPc
E-opt	RuBP, PGA, T3P, E4P, S7P, SBP, PenP, GCEA, SERc, F26BPc
SE-opt	RuBP, PGA, SBP, PenP, GOAc, SERc, T3Pc, FBPc, HexPc, F26BPc

Time	Uniform			C2 Optimal			C3 Optimal		
Method	**D**	**E**	**SE**	**D**	**E**	**SE**	**D**	**E**	**SE**
$ASE(RuBP_0)$	0.316	0.316	0.316	**0.224**	0.315	**0.224**	**0.224**	0.313	0.313
$ASE(SBP_0)$	0.123	0.123	0.123	**0.087**	0.122	**0.087**	**0.087**	0.122	0.122
$ASE(KM_{11})$	**0.375**	0.864	0.641	0.526	**0.025**	0.625	0.091	0.014	**0.012**
$ASE(KM_{13})$	**2.064**	4.639	3.972	2.931	**0.058**	3.911	0.500	0.069	**0.062**
$ASE(KI_{13})$	**2.161**	2.668	9.785	3.241	**1.696**	9.444	**0.876**	1.202	1.116
$ASE(KE_4)$	**0.011**	0.014	1.509	0.177	**0.012**	1.407	**0.008**	0.011	0.036
$ASE(KM_9)$	**0.249**	0.253	1.576	0.540	**0.176**	1.555	0.173	0.141	**0.126**
$ASE(KM_{131})$	2.686	**2.299**	3.028	4.294	**0.125**	2.981	0.218	**0.094**	0.260
$ASE(KI_{135})$	39.0	**33.4**	46.4	62.4	**1.795**	45.7	3.154	**1.332**	3.683
$ASE(KE_{22})$	**0.054**	0.093	2.284	1.151	**0.034**	2.132	0.014	**0.014**	0.050
$ASE(KM_{511})$	0.020	1.626	**0.004**	0.427	0.074	**0.004**	0.005	0.018	**0.002**
$ASE(KM_{521})$	8.695	17.6	**4.397**	133.9	**1.382**	4.739	2.686	0.507	**0.168**
$ASE(KI_{523})$	0.244	0.494	**0.123**	3.757	**0.036**	0.133	0.075	0.014	**0.005**
$ASE(KC)$	**0.051**	0.063	0.263	0.077	**0.040**	0.254	**0.022**	0.029	0.030
$ASE(KM_{1221})$	**0.305**	2.041	0.336	**0.323**	0.730	0.349	0.288	0.578	**0.258**
$ASE(KM_{1241})$	**0.273**	1.827	0.295	**0.289**	0.654	0.307	0.258	0.517	**0.231**
$ASE(V_9)$	**0.240**	0.247	2.458	0.496	**0.169**	2.210	0.164	**0.139**	0.141
$ASE(V_{58})$	0.024	0.151	**0.003**	0.510	0.008	**0.003**	0.005	0.001	**6.9e-4**
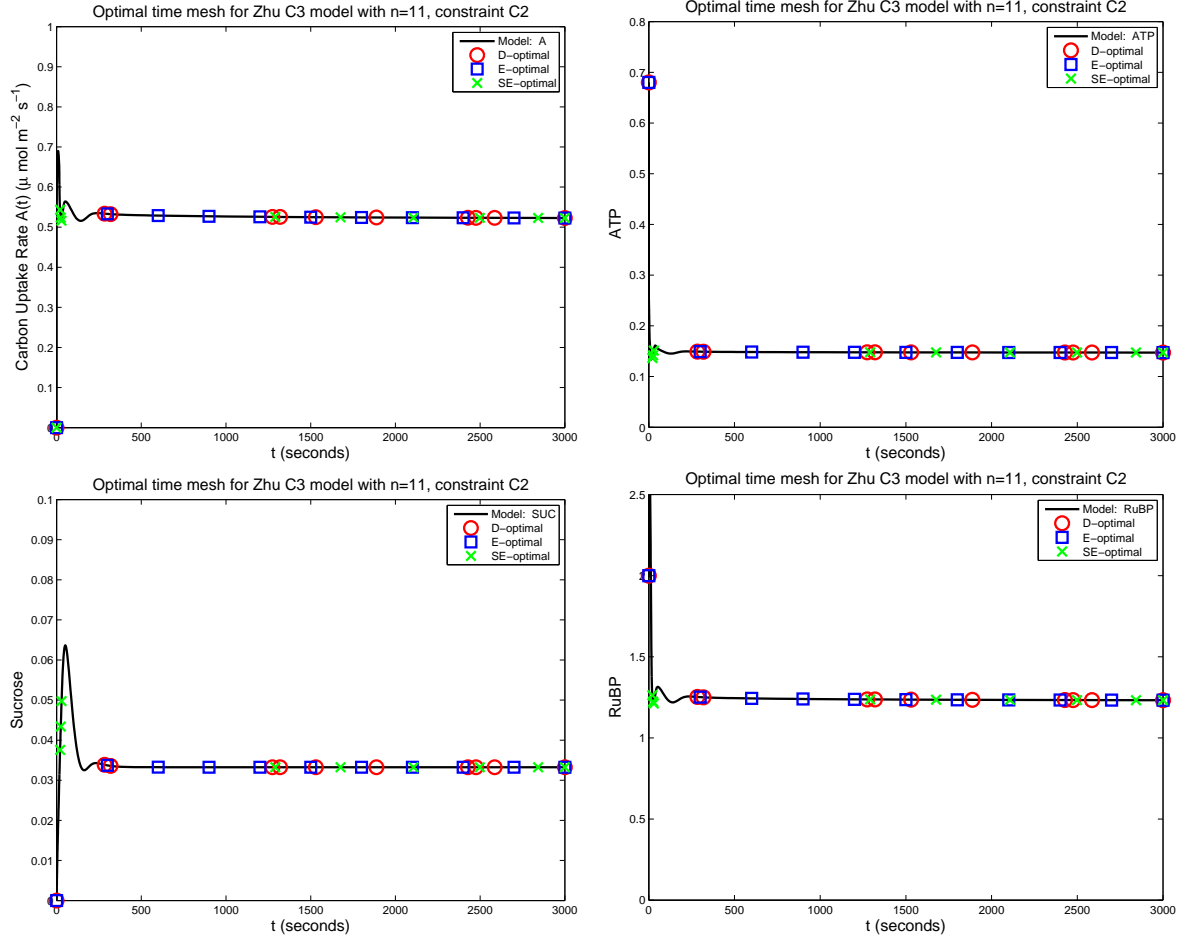
Figure 3.35: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C2** when sampling the optimal five observables to estimate $\vec{\theta}_c$ (Table 3.13). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
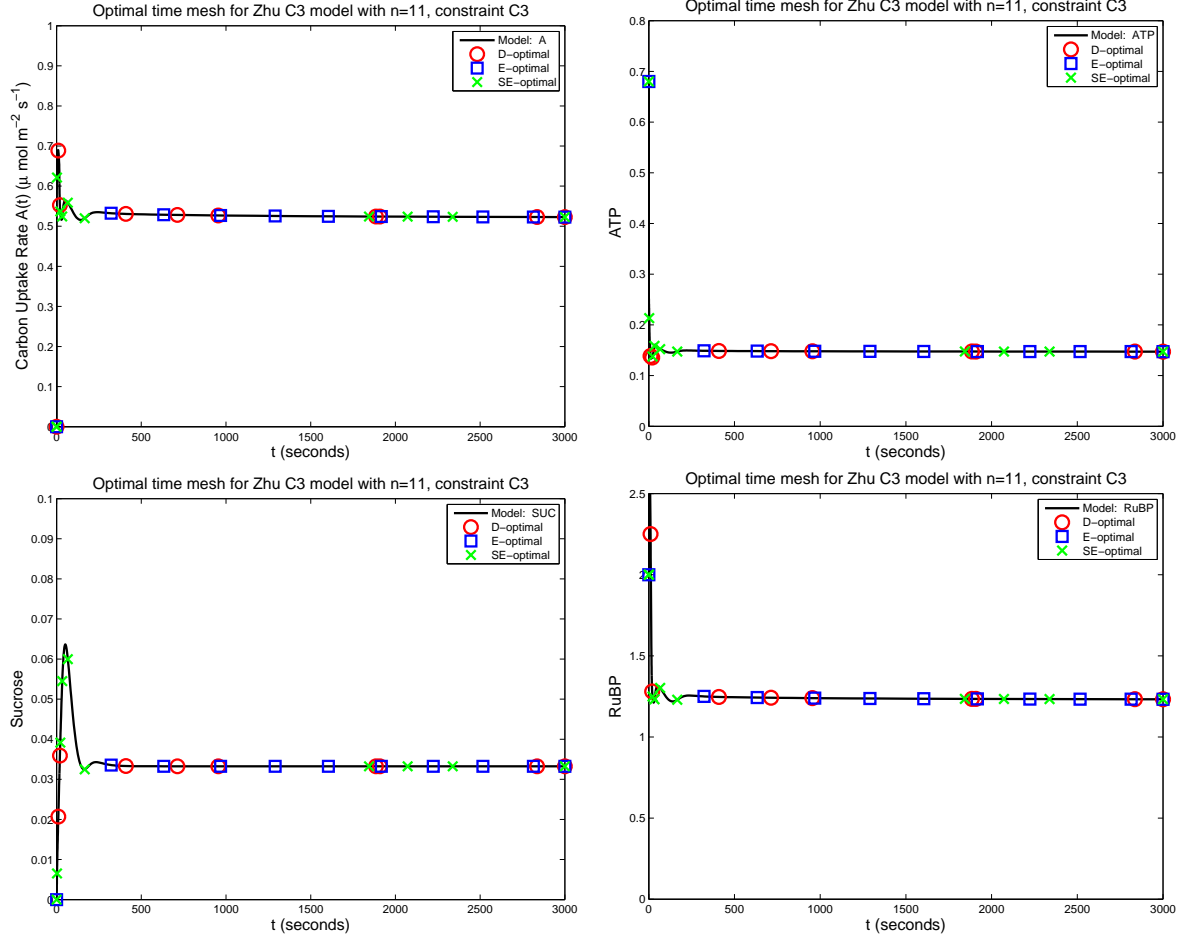
Figure 3.36: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ times under constraint **C3** when sampling the optimal five observables to estimate $\vec{\theta}_c$ (Table 3.13). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
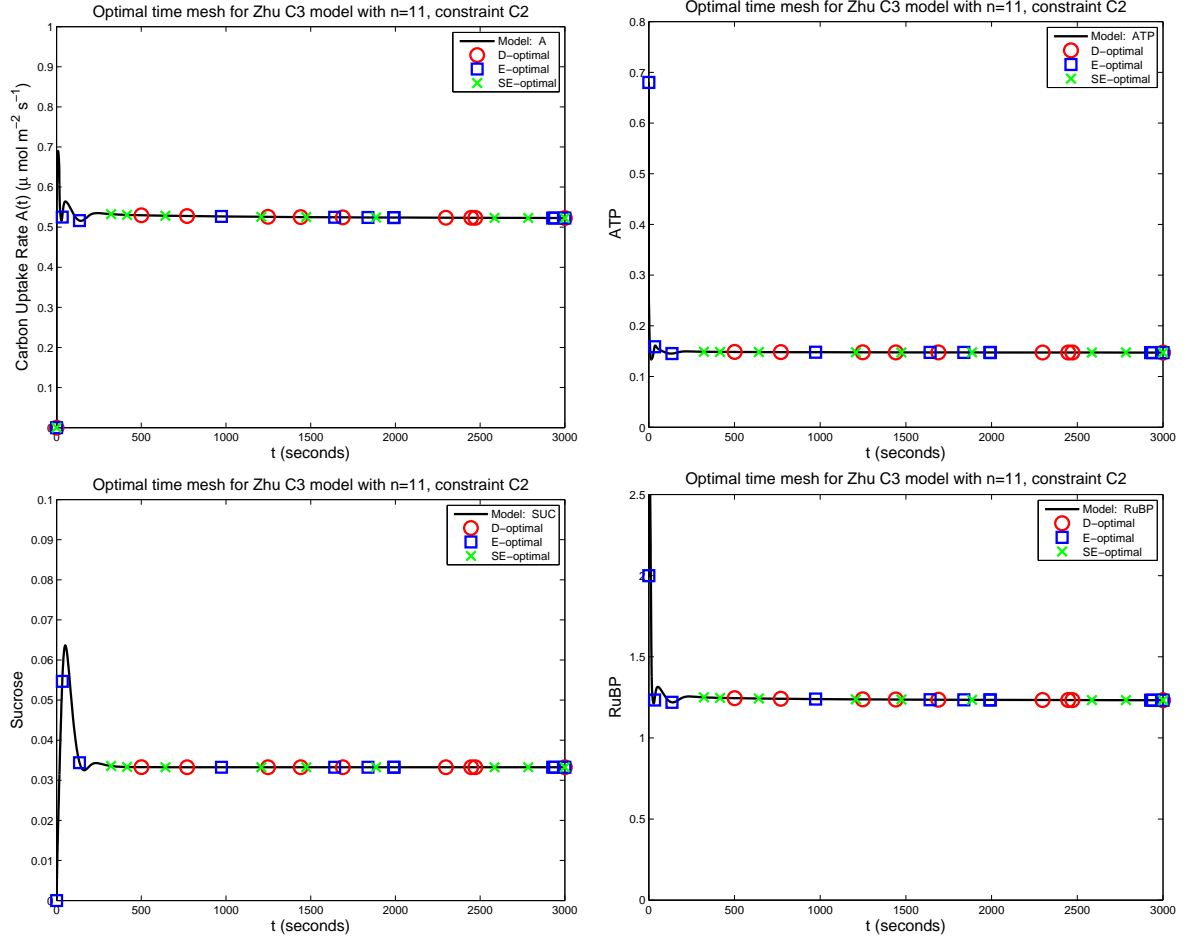
Figure 3.37: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ sampling times under constraint **C2** when sampling the optimal ten observables to estimate $\vec{\theta}_c$ (Table 3.14). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.
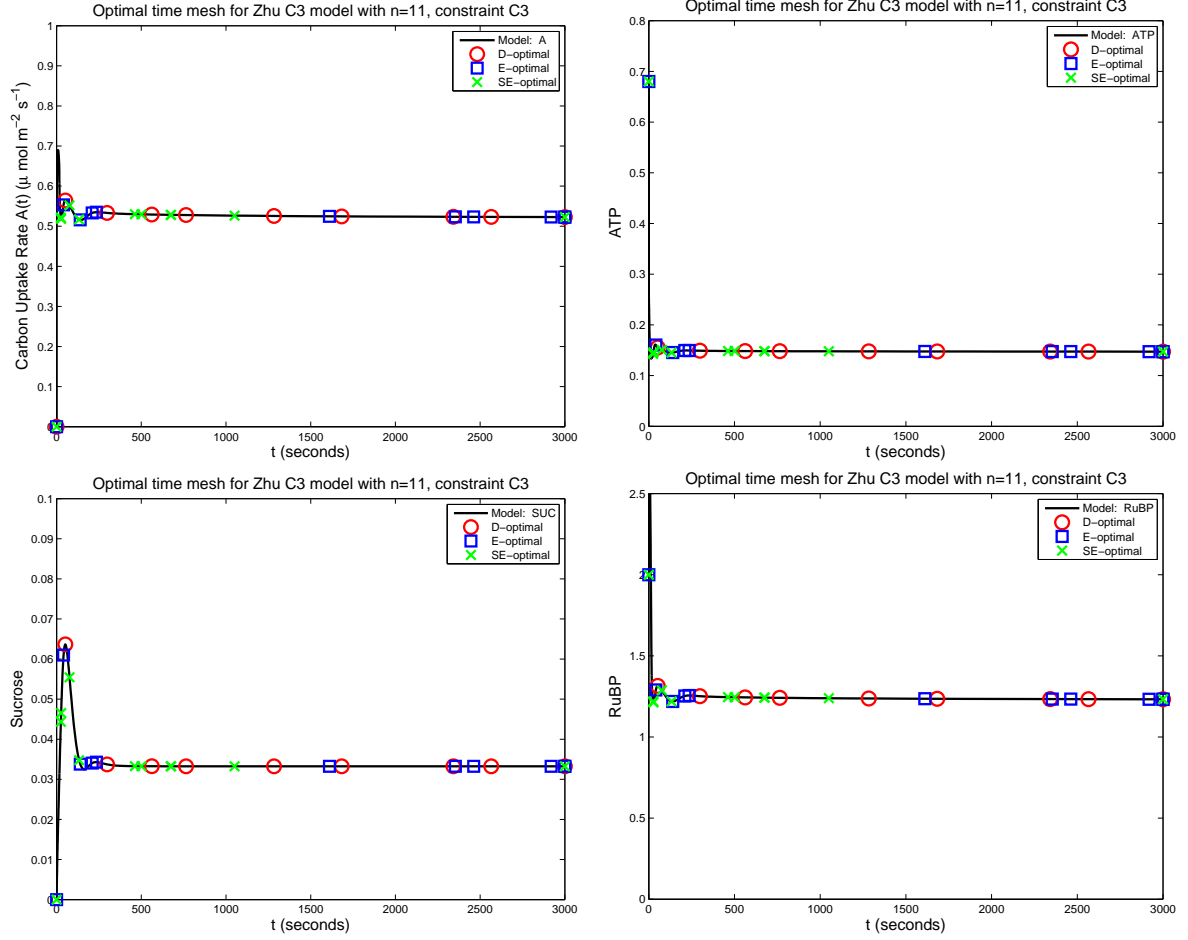
Figure 3.38: Solutions of selected state variables in the Zhu model [49]. Plotted on top of each curve are the D-optimal (circle), E-optimal (square), and SE-optimal (x) $n = 11$ times under constraint **C3** when sampling the optimal ten observables to estimate $\vec{\theta}_c$ (Table 3.14). Top Left: Carbon uptake rate $A(t)$; Top Right: ATP; Bottom Left: SUCc; Bottom Right: RuBP.

Chapter 4

Conclusion and Future Work

In this research project, we improve the ability to use in silico experimentation to inform in-field experiments by constructing multiscale models of plant growth and by developing an experimental design algorithm that determines which measurable state variables should be observed and when they should be observed. We explore the potential of these models to describe observed plant growth, examine how compartments of the model are affected by parameters, and show the potential of the experimental design algorithm using two simpler models.

4.1 Multiscale models

Drawing from the modeling and investigative work of [4], [26], [28], [42], [39], and [49], we construct two candidate dynamical systems models that relate environmental conditions, cellular level metabolic processes, and whole plant growth and senescence dynamics. The comprehensive model is listed in Appendix B.3, and the compact model is listed in Appendix B.4. Descriptions of all state variables and parameters, values of parameters and initial values of state variables (values at time $t = 0$), equations from [49], and equations from [28] are included in Appendices A.1, A.2, B.1, and B.2, respectively.

The current models improve upon past carbon metabolism modeling efforts by mathematically describing the reactions of PSI, PSII, FNR, OEC, the C3 cycle, and the most prominent links between the metabolic processes of the light reactions and C3 cycle. This cellular level is linked to whole plant productivity through the use of logistic functions and to the environment via forcing functions that represent the temperature, radiance, water availability, and atmospheric availability of CO_2 and O_2. The comprehensive model uses the C3 cycle model of [49] and the light reaction model of [28] to describe the majority of the metabolic processes while the compact model replaces the light reaction model with four equations based on the primary chemical reactions of photosynthesis as described in [42].

4.2 Parameter estimation and sensitivity analysis

The ability of these models to describe leaf area over time as observed in *Arabidopsis thaliana* [35] and soybean [39] are tested using a least squares optimization algorithm to identify unknown parameters such as growth and senescence rates. Both candidate models successfully describe overall trends in plant leaf area data (Figures 3.4 and 3.13) when a least-squares algorithm is used to estimate some of the model parameters. Further improvements to each model's fit may be possible if the initial points used in the minimization algorithm are changed.

The candidate models also exhibit changes in metabolite concentrations and carbon uptake rates, reflecting the changes in plant growth behavior between day and night. The two models exhibit metabolite concentrations that not only differ from each other but also metabolite concentrations reported in [28] and [49], all of which are not verified with data. The runtime of one forward solution of the compact model is less than 50% of the runtime of the comprehensive model with little increase in the minimal least-squares cost when fitting a set of data, indicating that of the two models, the compact model is more useful for computations.

Extending the efforts of [6] to understand which compartments of the C3 cycle model [49] are affected by particular parameters, we use automatic differentiation [21] to calculate the sensitivity equations of the compact model with respect to the maximum velocity terms, growth and senescence parameters, and initial conditions. The state variables of the compact model are sensitive to many parameters that are not used in the expression of that variable's rate of change (such as other variables' initial conditions), indicating that the addition of the light reactions and environmental factors has caused new relationships between variables and parameters to become apparent. The calculation of these sensitivity equations, however, is very time intensive, making even the compact model a poor candidate for use in a Fisher information matrix-based experimental design problem.

4.3 Experimental design

Expanding the efforts reported in [16] on time point distribution selection using the D-, E-, and SE-optimal design criteria, we introduce a new methodology and algorithm for selecting both optimal observables and an optimal time point distribution. While the D- and E-optimal cost functions are well established in the literature, the SE-optimal design method is relatively new but competitive. We compare the abilities of these three design criteria to reduce the estimated asymptotic standard errors of selected subsets of parameters for two ordinary differential equation models representative of ODE systems used in industrial applications: the log-scaled HIV model (3.2) and the model of [49] for the Calvin cycle. These examples suggest the strengths of each design method in selecting appropriate observable variables and sampling times.

Collecting patient data for the treatment of HIV is limited by the types and expenses of assays available. In efforts with experimental data [2], parameter estimation for the HIV model (3.2) has only been performed with one or two observed quantities. Our tests for determining the optimal sets of 1, 2, and 3 observables when data is collected at uniformly distributed times show that the SE-optimal design criterion performs very similarly to the D-optimal criterion; moreover, a small number of observables may be compensated for by allowing more sampling times. While only measuring one observable may not quite be adequate for accurate parameter estimates, taking either 201 or 401 measurements of two observables yielded similar standard errors when compared to an experiment in which three observables are sampled 51-101 times.

For the HIV model, we compared the performance of all three design criteria under time point distribution constraints C2 and C3. When determining the optimal distribution of 35 time points, D-optimal yields the smallest ASE's for the most parameters, followed by E-optimal, for both time point constraints. When determining the optimal distribution of 105 time points, however, D-, E-, and SE-optimal each yield the smallest ASE's for some of the parameters under C2, and both D- and E-optimal are strong performers under constraint C3. The SE-optimal standard errors, while not the lowest, are on the same order of magnitude as those of D- and E-optimal. Therefore for our selected parameter values, D- and E-optimal more reliably yield the smallest standard errors.

The Zhu model of [49] is representative of experimental environments where a great number of state variables may possibly be measured but the costs of measurements limit the number of observed variables that may be sampled. We compared the estimated asymptotic standard errors calculated using the observables and time point distributions determined with the D-,E-, and SE-optimal design methods. For the small parameter subset $\vec{\theta}_a$, a small number of observables is sufficient to obtain as small ASE's as possible - adding more observables does not improve the ASE's generated for any of the design methods. Optimizing the time point distribution under either constraint C2 or C3 improves the ASE's by up to two orders of magnitude. At both 5 and 10 observables, the D-optimal design provides the smallest ASE's for the most parameters, followed by SE-optimal. The E-optimal standard errors, however, are often on the same order of magnitude, indicating that it is still competitive.

When testing the treatment of the larger parameter subset $\vec{\theta}_c$ of the Zhu model by each of the three design methods, we found that while an excellent selection of observables and time points may still yield small standard errors (as with the SE-optimal design method with 5 observables under constraint C3), adding more information through more observables reduces the calculated ASE's in the uniform time distribution. At 5 observables, the SE-optimal cost function performs the best and E-optimal time distribution optimization fails to venture far from the uniform grid initially used; at 10 observables, E-optimal performs best under constraint C2 and all three design criteria perform the best for different parameters under constraint C3. Thus

for the Zhu model, it is difficult to predict which optimal design criterion will perform best in a particular case.

While the performance of each optimal design criteria is highly dependent upon the ODE system, parameter subset, number of observables allowed, number of time points allowed, and even constraints imposed on the time point distribution selection, the examples performed using the HIV model (3.2) and the Zhu model [49] demonstrate that the D-, E-, and SE-optimal design methods are all competitive and useful. Moreover, in the Zhu model examples, selection of optimal time points can reduce the estimated asymptotic standard errors of parameters by several orders of magnitude, thus providing data that is more useful in a parameter estimation problem without taking more samples or measuring more observables. Thus the new methodology developed and illustrated here can be an important tool in designing experimental protocols for obtaining data to be used in estimating parameters in complex dynamic models.

4.4   Concluding remarks

The calculation of sensitivity equations for 1.4 days of the compact model enables us to solve some simple optimal design problems for a C3 plant grown in greenhouse conditions similar to those described in [35]. Particularly, we may use the sensitivity equations to determine which metabolites should be measured - and when in a 24 hour period the measurements should be taken - in order to get the most information about maximum velocity parameters and, if unknown, initial metabolite concentrations assuming that the parameters used in the model are near the true values that describe the reactions occurring in plant tissue. Further model simplification would be necessary to solve an optimal design problem over the lifetime of a plant.

While it is assumed in the model that the locations of photosynthesis are fixed sizes, cells are not static structures. Cell compartments are known to change size based on water and gas availability. This, in turn, may impact the concentration of metabolites in that portion of the cell. Further investigation into the sizes of particular cellular spaces such as the lumen, the cytosolic space, and even the resource-storing vacuole will enable us to better describe how plant productivity changes when cells are different sizes or when cell compartments change in size due to environmental effects, genetic variation, or plant development.

The multiscale plant growth models listed in sections B.3 and B.4 of the appendix apply only to plant species that utilize the C3 cycle to fix carbon in cells. The models will be used to formulate a similarly structured model for C4 plants such as maize. The C4 cycle is a light-independent carbon fixing process that is similar to - but requires more reactions than - the C3 cycle. We will search for models of the C4 cycle to implement in our current mathematical framework as well as consult [4], [37], and [42] to determine how this alternative carbon fixing

mechanism impacts light reactions and overall productivity of the plant. The ability of this multiscale C4 plant growth model to approximate trends seen in crop leaf area or biomass data when optimal parameters are determined will be tested, with the intention of using this model in an experimental design framework to improve and streamline sampling in a greenhouse or field experimental setting.

REFERENCES

[1] B. M. Adams, H. T. Banks, M. Davidian, H. Kwon, H. T. Tran, S. N. Winfree, and E. S. Rosenberg, HIV dynamics: Modeling, data analysis, and optimal treatment protocols, CRSC Technical Report CRSC-TR04-05, NCSU, February 2004; *J. Comp. Appl. Math.*, **184** (2005), 10–49.

[2] B. M. Adams, H. T. Banks, M. Davidian, and E. S. Rosenberg, Model fitting and prediction with HIV treatment interruption data, CRSC Technical Report CRSC-TR05-40, NCSU, October 2005; *Bull. Math. Biol.*, **69** (2007), 563–584.

[3] B. M. Adams, *Non-parametric Parameter Estimation and Clinical Data Fitting with a Model of HIV Infection,* PhD Thesis, NC State Univ., 2005.

[4] J. S. Amthor, From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy, *New Phytologist* **188** (2010), pp. 939–959.

[5] A. Attarian. tssolve.m, Retrieved August 2011, from http://www4.ncsu.edu/ arattari/.

[6] M. Avery, H. T. Banks, K. Basu, Y. Cheng, E. Eager, S. Khasawinah, L. K. Potter, and K. L. Rehm, Experimental design and inverse problems in plant biological modeling, CRSC Technical Report CRSC-TR11-12, NCSU, October, 2011; *J. Inverse and Ill-posed Problems*, DOI 10.1515/jiip-2012-0208.

[7] H. T. Banks, *Modeling and Control in the Biomedical Sciences,* Lecture Notes in Biomathematics, Vol. 6, (S. Levin ed.) Springer-Verlag, Berlin Heidelberg New York, 1975.

[8] H. T. Banks and K. L. Bihari, Modeling and estimating uncertainty in parameter estimation, CRSC Technical Report CRSC-TR99-40, NCSU, December 1999; *Inverse Problems*, **17** (2001), 95–111.

[9] H. T. Banks, A. Cintrón-Arias, and F. Kappel, Parameter selection methods in inverse problem formulation, CRSC Technical Report CRSC-TR10-03, NCSU, revised November 2010; in *Mathematical Model Development and Validation in Physiology: Application to the Cardiovascular and Respiratory Systems*, Lecture Notes in Mathematics, Mathematical Biosciences Subseries, Springer-Verlag, 2012, to appear.

[10] H. T. Banks, M. Davidian, S. Hu, G. Kepler, and E. Rosenberg, Modeling HIV immune response and validation with clinical data, CRSC Technical Report CRSC-TR07-09, NCSU, March 2007; *J. Biological Dynamics*, **2** (2008), 357–385.

[11] H. T. Banks, M. Davidian, J. R. Samuels, and K. L. Sutton. An inverse problem statistical methodology summary, CRSC Technical Report CRSC-TR08-01, NCSU, January 2008; Chapter 11 in *Statistical Estimation Approaches in Epidemiology*, Gerardo Chowell, et. al., eds. Springer, Berlin Heidelberg New York, 2009, 249–302.

[12] H. T. Banks, S. Dediu, S. L. Ernstberger, and F. Kappel, Generalized sensitivities and optimal experimental design, CRSC-TR08-12, September, 2008, (Revised), November, 2009; *J. Inverse and Ill-posed Problems*, **18** (2010), 25–83.

[13] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhaüser, Boston, 1989.

[14] H. T. Banks and K. L. Rehm, Experimental design for vector output systems, CRSC Technical Report CRSC-TR12-11, NCSU, April, 2012; *Inverse Problems in Science and Engineering*, (2013), 1–34; DOI: 10.1080/17415977.2013.797973

[15] H. T. Banks and K. L. Rehm, Experimental design for distributed parameter vector systems, CRSC Technical Report CRSC-TR12-11, NCSU, August, 2012; *Applied Mathematics Letters,* **26** (2013), 10–14; http://dx.doi.org/10.1016/j.aml.2012.08.003.

[16] H. T. Banks, K. Holm, and F. Kappel, Comparison of optimal design methods in inverse problems, CRSC Technical Report CRSC-TR10-11, NCSU, July 2010; *Inverse Problems*, **27** (2011), 075002.

[17] H. T. Banks and H. T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton, Florida, 2009.

[18] R. H. Byrd, J. C. Gilbert, and J. Nocedal, A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming, *Mathematical Programming* **89** 1 (2000), pp. 149-185.

[19] M. Davidian and D. Giltinan. *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London, 1998.

[20] G. D. Farquhar, S. von Caemmerer, and J. A. Berry, A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 Species, *Planta* **149** (1980), pp. 78–90.

[21] M. Fink, myAD, Retrieved August 2011, from http://www.mathworks.com/matlabcentral/fileexchange/15235-automatic-differentiation-for-matlab.

[22] Govindjee, Sixty-three years sinks Kautsky: chlorophyll *a* fluorescence, *Aust. J. Plant Physiology* **22** (1995), pp. 131–160.

[23] J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie, The DSSAT cropping system model, *European J. Agronomy* **18** (2003), pp. 235–265.

[24] A. Kuntsevich and F. Kappel, SolvOpt, Retrieved July 2011, from http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/.

[25] A. Laisk, H. Eichelmann, V. Oja, A. Eatherall, and D. A. Walker, A mathematical model of the carbon metabolism in photosynthesis. Difficulties in explaining oscillations by fructose 2,6-biphosphate regulation, *Proc. R. Soc. Lond.* **237** (1989), pp. 389–415.

[26] A. Laisk, H. Eichelmann, and V. Oja, Leaf C_3 photosynthesis in silico: Integration carbon/nitrogen metabolism, Chapter 13 in *Photosynthesis in Silico: Understanding Complexity from Molecules to Ecosystems*, A. Laisk, L. Nedbal, and Govindjee, eds. (Advances

in Photosynthesis and Respiration **29**, Govindjee, ed.), Springer, Dordrecht, The Netherlands, 2009, pp. 295–322.

[27] A. Laisk and D. A. Walker, Control of phosphate turnover as a rate-limiting factor and possible cause of oscillations in photosynthesis: a mathematical model, *Proc. R. Soc. Lond.* **227** (1986), pp. 281–302.

[28] D. Lazar, Modelling of light-induced chlorophyll a fluorescence rise (O-J-I-P transient) and changes in the 820 nm-transmittance signal of photosynthesis, *Photosynthetica* **47** (2009), pp. 483–498.

[29] L. Menten and M. I. Michaelis, Die Kinetik der Interinwirkung, *Biochem Z* **49** (1913), pp. 333–369.

[30] W. G. Nolan and R. M. Smillie, Temperature-induced changes in Hill activity of chloroplasts isolated from chilling-sensitive and chilling-resistant plants, *Plant Physiology* **59** (1977), pp. 1141–1145.

[31] G. Pettersson and U. Ryde-Pettersson, A mathematical model of the Calvin photosynthesis cycle, *European J. Biochem.* **175** (1988), pp. 661–672.

[32] M. G. Poolman, D. A. Fell, and S. Thomas, Modelling photosynthesis and its control, *J. Experimental Botany* **51** (2000), pp. 319–328.

[33] J. R. Potter and J. W. Jones, Leaf area partitioning as an important factor in growth, *Plant Physiology* **59** (1977), pp. 10–14.

[34] Y. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.*, **1** (1956), 157-214.

[35] K. A. Pyke, J. L. Marrison, and R. M. Leech, Temporal and spatial development of the cells of the expanding first leaf of Arabidopsis thaliana (L.) Heynh., *J. Experimental Botany* **42** (1991), pp. 1407-1416.

[36] A. Rubin and G. Riznichenko, Modeling of the primary processes in a photosynthetic membrane, Chapter 7 in *Photosynthesis in Silico: Understanding Complexity from Molecules to Ecosystems*, A. Laisk, L. Nedbal, and Govindjee, eds. (Advances in Photosynthesis and Respiration **29**, Govindjee, ed.), Springer, Dordrecht, The Netherlands, 2009, pp. 152–176.

[37] D. Schomburg, BRENDA, retrieved January 2013 from http://www.brenda-enzymes.org.

[38] L. F. Shampine and M. W. Reichelt, The MATLAB ODE Suite, *SIAM Journal on Scientific Computing* **18** (1997) pp. 1–22.

[39] T. D. Setiyono, A. Weiss, J. E. Specht, K. G. Cassman, and A. Dobermann, Leaf area index simulation in soybean grown under near-optimal conditions, *Field Crops Res.* **108** (2008), pp. 82–92.

[40] T. R. Sinclair, Water and nitrogen limitations in soybean grain production: i. model development, *Field Crops Res.* **15** (1986), pp. 125–141.

[41] K. Song, Chloroplast Structure, retrieved July 2013, from http://en.wikipedia.org/wiki/File:Chloroplast_structure.svg.

[42] SRI International, MetaCyc Encyclopedia of Metabolic Pathways, retrieved January 2013, from http://metacyc.org.

[43] P. Steduto, T. C. Hsiao, D. Raes, and E. Fereres, AquaCrop – The FAO crop model to simulate yield response to water: I. Concepts and underlying principles, *Agronomy J.* **101** (2009), pp. 426–437.

[44] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Wiley, New York, 1989.

[45] K. Thomaseth and C. Cobelli, Generalized sensitivity functions in physiological system identification, *Ann. Biomed. Eng.*, **27** (1999), 607–616.

[46] M. R. Villarreal, Plant Cell Structure, retrieved July 2013, from http://en.wikipedia.org/wiki/File:Plant_cell_structure_svg.svg.

[47] S. von Caemmerer, G D. Farquhar, and J. A. Berry, Biochemical model of C_3 Photosynthesis, Chapter 9 in *Photosynthesis in Silico: Understanding Complexity from Molecules to Ecosystems*, A. Laisk, L. Nedbal, and Govindjee, eds. (Advances in Photosynthesis and Respiration **29**, Govindjee, ed.), Dordrecht, The Netherlands, Springer, 2009, pp. 209–230.

[48] R. Wheeler, Leaf Tissue Structure, retrieved July 2013, from http://en.wikipedia.org/wiki/File:Leaf_Tissue_Structure.svg.

[49] X.-G. Zhu, E. de Sturler, and S. P. Long, Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm, *Plant Physiology* **145** (2007), pp. 513–526.

[50] X.-G. Zhu, Govindjee, N. R. Baker, E. de Sturler, D. R. Ort, and S. P. Long, Chlorophyll *a* fluorescence induction kinetics in leaves predicted from a model describing each discrete step of excitation energy and electron transfer associated with photosystem II, *Planta* **223** (2005), pp. 114–133.

APPENDICES

Appendix A

State variables and parameters

A.1 State variables, descriptions, and initial values

Symbol	Eqn, M1	Eqn, M2	IC	Location	Description
RuBP	x_1	x_1	2	Stroma	Ribulose 1,5-biphosphate
PGA	x_2	x_2	2.4	Stroma	3-Phosphoglycerate
DPGA	x_3	x_3	0.0011	Stroma	1,3-biphosphoglycerate
T3P	x_4	x_4	0.5	Stroma	Triose phosphates (DHAP and GAP)
DHAP	y_{10}	y_{10}	-	Stroma	Dihydroxyacetone-phosphate
GAP	y_{11}	y_{11}	-	Stroma	Glyceraldehyde 3-phosphate
FBP	x_5	x_5	0.67	Stroma	Fructose 1,6-bisphosphate
E4P	x_6	x_6	0.05	Stroma	Erythrose 4-phosphate
S7P	x_7	x_7	2	Stroma	Sedoheptulose 7-phosphate
ATP	x_8	x_8	0.8019	Stroma	Adenosine triphosphate
SBP	x_9	x_9	0.3	Stroma	Sedoheptulose 1,7-bisphosphate
NADPH	x_{10}	x_{10}	0.0485	Stroma	Reduced NADP
ADP	x_{11}	x_{11}	0.82	Stroma	Adenosine diphosphate
NADP	x_{12}	x_{12}	0.1752	Stroma	NAD phosphate
HexP	x_{13}	x_{13}	2.2	Stroma	Hexose phosphates F6P, G6P, G1P
F6P	y_{12}	y_{12}	-	Stroma	Fructose 6-phosphate
G6P	y_{13}	y_{13}	-	Stroma	Glucose 6-phosphate
G1P	y_{14}	y_{14}	-	Stroma	Glucose 1-phosphate
PenP	x_{14}	x_{14}	0.2742	Stroma	Pentose phosphates Ri5P, Ru5P, Xu5P
Ru5P	y_{15}	y_{15}	-	Stroma	Ribulose 5-phosphate
Ri5P	y_{16}	y_{16}	-	Stroma	Ribose 5-phosphate
Xu5P	y_{17}	y_{17}	-	Stroma	Xylulose 5-phosphate
NADHc	x_{16}	x_{16}	0.3244	Cytosol	Reduced NAD
NADc	x_{17}	x_{17}	0.1149	Cytosol	Nicotinamide adenine dinucleotide
ADPc	x_{18}	x_{18}	0.0122	Cytosol	Adenosine diphosphate
ATPc	x_{19}	x_{19}	0.0412	Cytosol	Adenosine triphosphate
GLUc	x_{20}	x_{20}	18.724	Cytosol	Glutamate
KGc	x_{21}	-	0.4824	Cytosol	α-Ketoglutarate
GCEA	x_{23}	x_{23}	0.1812	Stroma	Glycerate
GCA	x_{24}	x_{24}	0.36	Stroma	Glycollate

PGCA	x_{25}	x_{25}	0.0029	Stroma	Phosphoglycollate
GCAc	x_{26}	x_{26}	0.36	Cytosol	Glycollate
GOAc	x_{27}	x_{27}	0.028	Cytosol	Glyoxylate
SERc	x_{28}	x_{28}	7.5	Cytosol	Serine
GLYc	x_{29}	x_{29}	1.8	Cytosol	Glycine
HPRc	x_{30}	x_{30}	0.0035	Cytosol	Hydroxypyruvate
GCEAc	x_{31}	x_{31}	0.1812	Cytosol	Glycerate
T3Pc	x_{32}	x_{32}	2.3	Cytosol	Triose phosphates
GAPc	y_4	y_4	-	Stroma	Glyceraldehyde 3-phosphate
DHAPc	y_5	y_5	-	Stroma	Dihydroxyacetone-phosphate
FBPc	x_{33}	x_{33}	2	Cytosol	Fructose 1,6-bisphosphate
HexPc	x_{34}	x_{34}	5.8	Cytosol	Hexose phosphates
G6Pc	y_1	y_1	-	Stroma	Glucose 6-phosphate
F6Pc	y_2	y_2	-	Stroma	Fructose 6-phosphate
G1Pc	y_3	y_3	-	Stroma	Glucose 1-phosphate
F26BPc	x_{35}	x_{35}	7.8e-6	Cytosol	Fructose 2,6-bisphosphate
UDPGc	x_{36}	x_{36}	0.57	Cytosol	Uridine diphosphate glucose
UTPc	x_{37}	x_{37}	0.75	Cytosol	Uridine-5'-triphosphate
SUCPc	x_{38}	x_{38}	0	Cytosol	Sucrose phosphate
SUCc	x_{39}	x_{39}	0	Cytosol	Sucrose
PGAc	x_{40}	x_{40}	0	Cytosol	3-Phosphoglycerate
$\int A$	x_{41}	x_{49}	0	Stroma	Total carbon uptake
L$^-$(HC)$^-$F	x_{42}	-	0.0014	Thylakoid membrane	State of cyt b6/f
L$^-$(HC)$^-$F$^-$	x_{43}	-	0.0012	Thylakoid membrane	State of cyt b6/f
L$^-$(HC)F	x_{44}	-	1.8270	Thylakoid membrane	State of cyt b6/f
L$^-$(HC)F$^-$	x_{45}	-	1.0747	Thylakoid membrane	State of cyt b6/f
L$^-$(HC)$^{2-}$F	x_{46}	-	0.0022	Thylakoid membrane	State of cyt b6/f
L$^-$(HC)$^{2-}$F$^-$	x_{47}	-	0.9687	Thylakoid membrane	State of cyt b6/f
L(HC)$^-$F	x_{48}	-	0.9642	Thylakoid membrane	State of cyt b6/f
L(HC)$^-$F$^-$	x_{49}	-	0.0562	Thylakoid membrane	State of cyt b6/f
L(HC)F	x_{50}	-	1.5769	Thylakoid membrane	State of cyt b6/f
L(HC)F$^-$	x_{51}	-	3.03e-4	Thylakoid membrane	State of cyt b6/f
L(HC)$^{2-}$F	x_{52}	-	7.13e-4	Thylakoid membrane	State of cyt b6/f
L(HC)$^{2-}$F$^-$	x_{53}	-	1.0949	Thylakoid membrane	State of cyt b6/f
Fd	x_{54}	x_{47}	3.5152	Stroma	Ferredoxin without electron
Fd$^-$	x_{55}	x_{48}	4.39e-4	Stroma	Ferredoxin with electron
FNRa	x_{56}	-	1.2955	Stroma	Active FNR
FNRa$^-$	x_{57}	-	0.9657	Stroma	Active singly reduced FNR
FNRa^{2-}	x_{58}	-	0.9864	Stroma	Active doubly reduced FNR
FNRi	x_{59}	-	4.2331	Stroma	Inactive FNR
P$_{680}^+$Q$_A^-$N	x_{60}	-	1.2623	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A^-$Q$_B$	x_{61}	-	1.8548	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A^-$Q$_B^-$	x_{62}	-	0.0048	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A^-$Q$_B^{2-}$	x_{63}	-	1.1239	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A$Q$_B$	x_{64}	-	0.8241	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A$Q$_B^-$	x_{65}	-	0.0014	Thylakoid membrane	State in PSII
P$_{680}^+$Q$_A$Q$_B^{2-}$	x_{66}	-	1.5088	Thylakoid membrane	State in PSII
P$_{680}$Q$_A^-$N	x_{67}	-	3.97e-4	Thylakoid membrane	State in PSII
P$_{680}$Q$_A^-$Q$_B$	x_{68}	-	0.0117	Thylakoid membrane	State in PSII
P$_{680}$Q$_A^-$Q$_B^-$	x_{69}	-	0.0016	Thylakoid membrane	State in PSII
P$_{680}$Q$_A^-$Q$_B^{2-}$	x_{70}	-	0.8327	Thylakoid membrane	State in PSII

$P_{680}Q_AN$	x_{71}	-	0.0113	Thylakoid membrane	State in PSII
$P_{680}Q_AQ_B$	x_{72}	-	0.6713	Thylakoid membrane	State in PSII
$P_{680}Q_AQ_B^-$	x_{73}	-	1.1355	Thylakoid membrane	State in PSII
$P_{680}Q_AQ_B^{2-}$	x_{74}	-	4.43e-4	Thylakoid membrane	State in PSII
P_{700}^+Fb	x_{75}	-	0.0018	Thylakoid membrane	State in PSI
$P_{700}^+Fb^-$	x_{76}	-	0.0020	Thylakoid membrane	State in PSI
$P_{700}Fb$	x_{77}	-	0.8950	Thylakoid membrane	State in PSI
$P_{700}Fb^-$	x_{78}	-	5.05e-4	Thylakoid membrane	State in PSI
PC	x_{79}	x_{44}	3.4037	Lumen	Plastocyanin
PC^+	x_{80}	x_{43}	1.85e-4	Lumen	Oxidized PC
PQ	x_{81}	x_{41}	2.8938	Thylakoid membrane	Oxidized plastoquinone
PQH_2	x_{82}	x_{42}	2.5	Thylakoid membrane	Reduced and protonated PQ
S_0	x_{83}	-	0.4721	Lumen	OEC stage 0
S_1	x_{84}	-	0.6668	Lumen	OEC stage 1
S_2	x_{85}	-	0.0022	Lumen	OEC stage 2
S_3	x_{86}	-	0.8924	Lumen	OEC stage 3
S_4	x_{87}	-	0.9062	Lumen	OEC stage 4
HIFL	x_{88}	x_{45}	1.0911	Lumen	Free H^+
HIFS	x_{89}	x_{46}	4.8302	Stroma	Free H^+
G	x_{90}	x_{50}	G_0	Exterior	Growing leaf area
S	x_{91}	x_{51}	S_0	Exterior	Senescing leaf area
L	x_{92}	x_{52}	$G_0 + S_0$	Exterior	Total leaf area
UDPc	y_6	y_6	-	Cytosol	Uridine diphosphiate
PiTc	y_7	y_7	-	Cytosol	Total phosphate
Pic	y_8	y_8	-	Cytosol	Available inorganic phosphate
PPic	y_9	y_9	-	Cytosol	Pyrophosphate
Pi	y_{18}	y_{18}	-	Stroma	Available inorganic phosphate

A.2 Parameters and values as labeled in [49] and [28]

k_A: Activation constant, k_M: Michalis Menten constant, k_I: Inhibition constant, k_E: Equilibrium constant, V:maximum velocity. C3: C3 cycle, LR: Light reactions, Suc: Sucrose synthesis, PR: Photorespiration.

Symbol	Value	Enzyme	Description
$k_{M1,1}$	0.0115	4.1.1.39	C3, PR: CO_2
$k_{M1,2}$	0.222	4.1.1.39	C3, PR: O_2
$k_{M1,3}$	0.2	4.1.1.39	C3, PR: RuBP
$k_{I1,1}$	0.84	4.1.1.39	C3: PGA
$k_{I1,2}$	0.04	4.1.1.39	C3: FBP
$k_{I1,3}$	0.075	4.1.1.39	C3: SBP
$k_{I1,4}$	0.9	4.1.1.39	C3: Pi
$k_{I1,5}$	0.07	4.1.1.39	C3: NADPH
$k_{M2,1}$	0.24	2.7.2.3	C3: PGA
$k_{M2,2}$	0.39	2.7.2.3	C3: ATP
$k_{M2,3}$	0.23	2.7.2.3	C3: ADP
$k_{M3,1}$	0.004	1.2.1.13	C3: DPGA
$k_{M3,2}$	0.1	1.2.1.13	C3: NADPH
$k_{M4,1}$	2.5	5.3.1.1	C3: DHAP
$k_{M4,2}$	0.68	5.3.1.1	C3: GAP
k_{E4}	0.05	5.3.1.1	C3
$k_{M5,1}$	0.3	4.1.2.13	C3: GAP
$k_{M5,2}$	0.4	4.1.2.13	C3: DHAP
$k_{M5,3}$	0.02	4.1.2.13	C3: FBP
k_{E5}	7.1	4.1.2.13	C3
$k_{M6,1}$	0.033	3.1.3.11	C3: FBP
$k_{I6,1}$	0.7	3.1.3.11	C3: F6P
$k_{I6,2}$	12	3.1.3.11	C3: Pi
k_{E6}	6.66e5	3.1.3.11	C3
$k_{M7,1}$	0.1	2.2.1.1	C3: Xu5P
$k_{M7,2}$	0.1	2.2.1.1	C3: E4P
$k_{M7,3}$	0.1	2.2.1.1	C3: F6P
$k_{M7,4}$	0.1	2.2.1.1	C3: GAP
k_{E7}	10	2.2.1.1	C3
$k_{M8,1}$	0.4	4.1.2.13	C3: SBP
$k_{M8,2}$	0.2	4.1.2.13	C3: DHAP
$k_{M8,3}$	1.017	4.1.2.13	C3: E4P
k_{E8}	1.017	4.1.2.13	C3
k_{M9}	0.05	3.1.3.37	C3: SBP
k_{I9}	12	3.1.3.37	C3: Pi
k_{E9}	6.66e5	3.1.3.37	C3
$k_{M10,1}$	1.5	2.2.1.1	C3: Ri5P
$k_{M10,2}$	0.1	2.2.1.1	C3: Xu5P
$k_{M10,3}$	0.072	2.2.1.1	C3: GAP
$k_{M10,4}$	0.46	2.2.1.1	C3: S7P
k_{E10}	1.1765	2.2.1.1	C3
k_{HPR}	4	3.6.3.14	LR: H^+/ATP requirement
$k_{M11,1}$	0.01447	3.6.3.14	LR: ADP
$k_{M11,2}$	0.3	3.6.3.14	LR: Pi
$k_{M11,3}$	0.3	3.6.3.14	LR: ATP

k_{E11}	6846	3.6.3.14	LR
k_{E12}	0.67	5.1.3.1	C3 12
$k_{M13,1}$	0.05	2.7.1.19	C3: Ru5P
$k_{M13,2}$	0.059	2.7.1.19	C3: ATP
$k_{I13,1}$	2	2.7.1.19	C3: PGA
$k_{I13,2}$	0.7	2.7.1.19	C3: RuBP
$k_{I13,3}$	4	2.7.1.19	C3: Pi
$k_{I13,4}$	2.5	2.7.1.19	C3: ADP
$k_{I13,5}$	0.4	2.7.1.19	C3: ADP
k_{E13}	6.846e3	2.7.1.19	C3
$k_{M16,1}$	0.014	3.6.3.14	C3: ADP
$k_{M16,2}$	0.3	3.6.3.14	C3: Pi
$k_{M16,3}$	0.3	3.6.3.14	C3: ATP
k_{E16}	5.734	3.6.3.14	C3
k_{E21}	2.3	5.3.1.9	C3
k_{E22}	0.058	5.4.2.2	C3 22
$k_{M23,1}$	0.08	multi-enzyme	C3: G1P
$k_{M23,2}$	0.08	multi-enzyme	C3: ATP
$k_{A23,1}$	0.1	multi-enzyme	C3: PGA
$k_{A23,2}$	0.02	multi-enzyme	C3: F6P
$k_{A23,3}$	0.02	multi-enzyme	C3: FBP
k_{I23}	10	multi-enzyme	C3: ADP
$k_{M31,1}$	0.077	Phosphate translocator	C3: DHAP
$k_{M31,2}$	0.63	Phosphate translocator	C3 31, Pi
$k_{M31,3}$	0.74	Phosphate translocator	C3 31, external phosphate
k_{M32}	0.25	Phosphate translocator	C3 32, PGA
k_{M33}	0.075	Phosphate translocator	C3 33, GAP
$k_{M36,1}$	1	1.6.6.1	C3: NADc est
$k_{M36,2}$	1	1.6.6.1	C3: NADPH est
k_{E36}	4.175	1.6.6.1	C3 est
k_{E38}	1.38	1.10.9.1	C3 est
$k_{M38,1}$	0.009	1.10.9.1	LR: PQH2
$k_{M38,2}$	3.38	1.10.9.1	LR: PC+, est
$k_{M38,3}$	1.4	1.10.9.1	LR: PC, est
$k_{M38,4}$	0.009	1.10.9.1	LR: PQ
k_{E40}	1	1.18.1.2	LR est
$k_{M40,1}$	0.0026	1.18.1.2	LR: Fd-, Brenda 658208
$k_{M40,2}$	0.00722	1.18.1.2	LR: NADP+, Brenda 285519
$k_{M40,3}$	0.0026	1.18.1.2	LR: Fd, Brenda 658208
$k_{M40,4}$	0.035	1.18.1.2	LR: NADPH, Brenda 658208
$k_{M41,1}$	0.01447	ATPc generation [49]	Suc
$k_{M41,2}$	0.3	ATPc generation [49]	Suc
$k_{M51,1}$	0.02	4.1.2.13	Suc: FBPc
$k_{M51,2}$	0.3	4.1.2.13	Suc: GAPc
$k_{M51,3}$	0.4	4.1.2.13	Suc: DHAPc
$k_{M51,4}$	0.014	4.1.2.13	Suc: SBPc
k_{E51}	12	4.1.2.13	Suc
k_{M52}	0.0025	3.1.3.11	Suc: FBPc
$k_{I52,1}$	0.7	3.1.3.11	Suc: F6Pc
$k_{I52,2}$	12	3.1.3.11	Suc: Pic
$k_{I52,3}$	7e-5	3.1.3.11	Suc: F26BPc

k_{E52}	6663	3.1.3.11	Suc
$k_{M55,1}$	0.14	2.7.7.9	Suc: G1Pc
$k_{M55,2}$	0.1	2.7.7.9	Suc: UTPc
$k_{M55,3}$	0.11	2.7.7.9	Suc: PPic
$k_{M55,4}$	0.12	2.7.7.9	Suc: UDPGc
k_{E55}	0.31	2.7.7.9	Suc
$k_{M56,1}$	0.8	2.4.1.14	Suc: FBPc
$k_{M56,2}$	2.4	2.4.1.14	Suc: UDPGc
$k_{I56,1}$	0.7	2.4.1.14	Suc: UDPc
$k_{I56,2}$	0.8	2.4.1.14	Suc: FPBc
$k_{I56,3}$	0.4	2.4.1.14	Suc: SUCPc
$k_{I56,4}$	11	2.4.1.14	Suc: Pic
$k_{I56,5}$	50	2.4.1.14	Suc: SUCc
k_{E56}	10	2.4.1.14	Suc
k_{M57}	0.35	3.1.3.24	Suc: SUCPc
k_{I57}	80	3.1.3.24	Suc: SUCc
k_{E57}	780	3.1.3.24	Suc
k_{M58}	0.032	3.1.3.46	Suc: F26BPc
$k_{I58,1}$	0.1	3.1.3.46	Suc: F6Pc
$k_{I58,2}$	0.5	3.1.3.46	Suc: PPic
$k_{M59,1}$	0.5	2.7.1.105	Suc: ATPc
$k_{M59,2}$	0.021	2.7.1.105	Suc: F26BPc
$k_{M59,3}$	0.5	2.7.1.105	Suc: F6Pc
$k_{I59,1}$	0.16	2.7.1.105	Suc: ADPc
$k_{I59,2}$	0.7	2.7.1.105	Suc: DHAPc
k_{E59}	590	2.7.1.105	Suc
$k_{M60,1}$	0.042	2.7.4.6	Suc: ADPc
$k_{M60,2}$	1.66	2.7.4.6	Suc: ATPc
$k_{M60,3}$	0.28	2.7.4.6	Suc: UDPc
k_{M60-4}	16	2.7.4.6	Suc: UTPc
k_{E60}	16	2.7.4.6	Suc
$k_{M62,1}$	5	Sucrose sink	Suc: SUCc
k_{M112}	0.026	3.1.3.18	PR: PGCA
$k_{I112,1}$	2.55	3.1.3.18	PR: Pi
$k_{I112,2}$	94	3.1.3.18	PR: GCA
$k_{M113,1}$	0.21	2.7.1.31	PR: ATP
$k_{M113,2}$	0.25	2.7.1.31	PR: GCEA
k_{I113}	0.36	2.7.1.31	PR: PGA
k_{E113}	300	2.7.1.31	PR
k_{M121}	0.1	1.1.3.15	PR: GCAc
$k_{M122,1}$	0.15	2.6.1.45	PR:, GOAc
$k_{M122,2}$	2.7	2.6.1.45	PR: SERc
k_{I122}	33	2.6.1.45	PR: GLYc
k_{E122}	0.24	2.6.1.45	PR
k_{M123}	0.09	1.1.1.29	PR: HPRc
k_{I123}	12	1.1.1.29	PR: HPRc
k_{E123}	2.5e5	1.1.1.29	PR
$k_{M124,1}$	0.15	2.6.1.4	PR: GOAc
$k_{M124,2}$	1.7	2.6.1.4	PR: GLUc
k_{I124}	2	2.6.1.4	PR:, GLYc
k_{E124}	607	2.6.1.4	PR

$k_{M131,1}$	6	Multi-enzyme [4]	PR: GLYc
$k_{I131,1}$	4	Multi-enzyme [4]	PR: SERc
$k_{I131,2}$	0.015	Multi-enzyme [4]	PR:, NADHc
$k_{M131,2}$	0.075	Multi-enzyme [4]	PR: NADc
$k_{M101,1}$	0.39	GCA/GCEA transport [49]	PR 101a, GCEA
$k_{I101,1}$	0.28	GCA/GCEA transport [49]	PR 101a, GCA
$k_{M101,2}$	0.2	GCA/GCEA transport [49]	PR 101b, GCA
$k_{I101,2}$	0.22	GCA/GCEA transport [49]	PR 101b, GCEA
V_1	2.9139	4.1.1.39	C3
V_2	30.1408	2.7.2.3	C3
V_3	4.0395	1.2.1.13	C3
V_5	1.2189	4.1.2.13	C3
V_6	0.7263	3.1.3.11	C3
V_7	3.1222	2.2.1.1	C3
V_8	1.2189	4.1.2.13	C3
V_9	0.3242	3.1.3.37	C3
V_{10}	3.1221	2.2.1.1	C3
V_{11}	0.8696	3.6.3.14	C3
V_{12}	0	5.1.3.1	C3
V_{13}	10.8348	2.7.1.19	C3
V_{16}	5	3.6.3.14	C3
V_{21}	0	5.3.1.9	C3
V_{22}	0	5.4.2.2	C3
V_{23}	0.2668	multi-enzyme	C3
V_{31}	1.2433	Phosphate translocator	C3
V_{32}	1.2433	Phosphate translocator	C3
V_{33}	1.2433	Phosphate translocator	C3
V_{36}	.05	1.6.6.1	C3
V_{37}	.005	1.10.3.9	LR
V_{38}	50	1.10.9.1	LR
V_{39}	1	1.97.1.12	LR
V_{40}	50	1.18.1.2	LR
V_{41}	.25	ATPc generation [49]	Suc
V_{51}	0.1074	4.1.2.13	Suc
V_{52}	0.0640	3.1.3.11	Suc
V_{55}	0.1154	2.7.7.9	Suc
V_{56}	0.0555	2.4.1.14	Suc
V_{57}	0.5550	3.1.3.24	Suc
V_{58}	0.0168	3.1.3.46	Suc
V_{59}	0.1009	2.7.1.105	Suc
V_{60}	6.1	2.7.4.6	Suc
V_{62}	2	Sucrose sink	Suc
V_{111}	0.6993	4.1.1.39	PR
V_{112}	52.4199	3.1.3.18	PR
V_{113}	5.7158	2.7.1.31	PR
V_{121}	1.4561	1.1.3.15	PR
V_{122}	3.3062	2.6.1.45	PR
V_{123}	10.0098	1.1.1.29	PR
V_{124}	2.7458	2.6.1.4	PR
V_{131}	2.4947	Multi-enzyme [4]	PR 131
$V_{101,1}$	5	GCA/GCEA transporter [49]	PR 101a

$V_{101,2}$	6	GCA/GCEA transporter [49]	PR 101b
$k_{(HC)L}$	345.67	1.10.9.1	LR: HC $\rightarrow$ L
$k_{L(HC)}$	18.3811	1.10.9.1	LR: L $\rightarrow$ HC
k_{ct}^b	1.0498	1.10.9.1	LR: Fd $\rightarrow$ Fd$^-$
k_{ct}^f	1.0952	1.10.9.1	LR: Fd$^-$ $\rightarrow$ Fd
k_F^b	7.5421	1.10.9.1	LR: PC $\rightarrow$ PC$^+$
k_F^f	280.646	1.10.9.1	LR: PC$^+$ $\rightarrow$ PC
k_{LF}^b	10.7049	1.10.3.9	LR: PQ $\rightarrow$ PQH$_2$
k_{LF}^f	1849.71	1.10.3.9	LR: PQH$_2$ $\rightarrow$ PQ
$k_{(HC)}^f$	171.138	1.10.3.9	LR: PQ $\rightarrow$ PQH$_2$
$k_{(HC)}^b$	20.6669	1.10.3.9	LR: PQH$_2$ $\rightarrow$ PQ
k_{ct}^{b2}	95.9296	1.10.9.1	LR: Fd $\rightarrow$ Fd$^-$
k_{ct}^{f2}	18.9606	1.10.0.1	LR: Fd$^-$ $\rightarrow$ Fd
k_x^f	129.276	1.97.1.12	LR: Fd $\rightarrow$ Fd$^-$
k_x^b	7.7645	1.97.1.12	LR: Fd$^-$ $\rightarrow$ Fd
k_{Fd}^f	2.2609	1.18.1.2	LR: Fd$^-$ $\rightarrow$ Fd
k_{Fd}^b	11.673	1.18.1.2	LR: Fd $\rightarrow$ Fd$^-$
k_a	32.0263	1.18.1.2	LR: FNRi $\rightarrow$ FNRa
k_{FNR}^f	64.3104	1.18.1.2	LR: FNRa^{2-} $\rightarrow$ FNRa
$k_{OEC0,1}$	1.289e6	OEC	LR: S$_0$ $\rightarrow$ S$_1$
$k_{OEC1,2}$	13366	OEC	LR: S$_1$ $\rightarrow$ S$_2$
$k_{OEC2,3}$	24876	OEC	LR: S$_2$ $\rightarrow$ S$_3$
$k_{OEC3,4}$	6360	OEC	LR: S$_3$ $\rightarrow$ S$_4$
k_{L2}	1.283e5	1.10.3.9	LR: P$_{680}$, Q$_A$ charge separation
k_{AB1}	57883	1.10.3.9	LR: Q$_A^-$ $\rightarrow$ Q$_B$
k_{BA1}	989.369	1.10.3.9	LR: Q$_B^-$ $\rightarrow$ Q$_A$
k_B^f	16.758	1.10.3.9	LR: PQ $\rightarrow$ PQH$_2$
k_B^b	796.658	1.10.3.9	LR: PQH$_2$ $\rightarrow$ PQ
k_{AB2}	5469.1	1.10.3.9	LR: Q$_A^-$ $\rightarrow$ Q$_B^-$
k_{BA2}	6.1552	1.10.3.9	LR: Q$_B^{2-}$ $\rightarrow$ Q$_A$
k_R^f	193.648	1.97.1.12	LR: PC $\rightarrow$ PC$^+$
k_R^b	16.4036	1.97.1.12	LR: PC$^+$ $\rightarrow$ PC
k_{L1}	113.144	1.97.1.12	LR: P$_{700}$, F$_B$ charge separation
$k_{OEC4,0}$	113.122	OEC	LR: S$_4$ $\rightarrow$ S$_0$
k_{FNR}^b	220	1.18.1.2	LR: FNRa $\rightarrow$ FNRa^{2-}
V_{APS}	12	ATPc sink	Suc
PT	15		Total phosphate in stroma
UTc	1.5		Total UDP and UTP in cytosol
PTc	15		Total phosphate in cytosol
Pres	1		Atmospheric pressure in ATM
CO$_2$	0.00036		Environment: atmospheric CO$_2$
O$_2$	0.20946		Environment: atmospheric O$_2$
T_{min}	4		Minimum T required for plant development ($^\circ C$)
T_{opt}	26		Optimum T for plant development ($^\circ C$)
T_{max}	40		Maximum T allowed for plant development ($^\circ C$)
r_G/A_{opt}			Growth rate
k_G			Carrying capacity for growth
r_S			Rate of senescence
k_S			Carrying capacity for senescence

Appendix B

Model equations

B.1 C3 cycle reaction equations

$$v_1(t; \vec{\theta}) = \frac{V_1 x_1 \min(1, x_1/E_t)}{x_1 + k_{M1,3}\left(1 + \frac{x_2}{k_{I1,1}} + \frac{x_5}{k_{I1,2}} + \frac{x_9}{k_{I1,3}} + \frac{y_{18}}{k_{I1,4}} + \frac{x_{10}}{k_{I1,5}}\right)} \frac{CO_2}{CO_2 + k_{M1,1}\left(1 + \frac{O_2}{k_{M1,2}}\right)}$$

$$v_2(t; \vec{\theta}) = \frac{V_2 x_2 x_8 - x_3 x11/k_{E2}}{(x_2 + k_{M2,1})\left(x_8 + k_{M2,2}\left(1 + \frac{x_{11}}{k_{M2,3}}\right)\right)}$$

$$v_3(t; \vec{\theta}) = \frac{V_3 x_3 x_{10}}{(x_3 + k_{M3,1})(x_{10} + k_{M3,2})}$$

$$v_4(t; \vec{\theta}) = \frac{V_5\left(y_{11}y_{10} - \frac{x_5}{k_{E5}}\right)}{(k_{M5,1}k_{M5,2})\left(1 + \frac{y_{11}}{k_{M5,1}} + \frac{y_{10}}{k_{M5,2}} + \frac{x_5}{k_{M5,3}} + \frac{y_{11}y_{10}}{k_{M5,1}k_{M5,2}}\right)}$$

$$v_5(t; \vec{\theta}) = \frac{V_6\left(x_5 - y_{12}\frac{y_{18}}{k_{E6}}\right)}{x_5 + k_{M6,1}\left(1 + \frac{y_{12}}{k_{I6,1}} + \frac{y_{18}}{k_{I6,2}}\right)}$$

$$v_6(t; \vec{\theta}) = \frac{V_7\left(y_{12}y_{11} - \frac{y_{17}x_6}{k_{E7}}\right)}{\left(y_{12} + k_{M7,3}\left(1 + \frac{y_{17}}{k_{M7,1}} + \frac{x_6}{k_{M7,2}}\right)\right)(y_{11} + k_{M7,4})}$$

$$v_7(t; \vec{\theta}) = \frac{V_8\left(y_{10}x_6 - \frac{x_9}{k_{E8}}\right)}{(x_6 + k_{M8,3})(y_{10} + k_{M8,2})}$$

$$v_8(t; \vec{\theta}) = \frac{V_9\left(x_9 - \frac{y_{18}x_7}{k_{E9}}\right)}{x_9 + k_{M9}\left(1 + \frac{y_{18}}{k_{I9}}\right)}$$

$$v_9(t; \vec{\theta}) = \frac{V_{10}\left(y_{11}x_7 - \frac{y_{16}y_{17}}{k_{E10}}\right)}{\left(y_{11} + k_{M10,3}\left(1 + \frac{y_{17}}{k_{M10,2}} + \frac{y_{16}}{k_{M10,1}}\right)\right)(x_7 + k_{M10-4})}$$

$$v_{10}(t; \vec{\theta}) = \frac{V_{13}\left(x_8 y_{15} - \frac{x_{11}x_1}{k_{E13}}\right)}{\left(x_8\left(1 + \frac{x_{11}}{k_{I13,4}}\right) + k_{M13,2}\left(1 + \frac{x_{11}}{k_{I13,5}}\right)\right)\left(y_{15} + k_{M13,1}\left(1 + \frac{x_2}{k_{I13,1}} + \frac{x_1}{k_{I13,2}} + \frac{y_{18}}{k_{I13,3}}\right)\right)}$$

$$v_{11}(t;\vec{\theta}) = \frac{V_{11}x_{11}y_{18} + \frac{x_8}{k_{E11}\left(\frac{HIFL}{HIFS}\right)^{k_{HPR}}}}{k_{M11,1}k_{M11,2}\left(1 + \frac{x_{11}}{k_{M11,1}} + \frac{y_{18}}{k_{M11,2}} + \frac{x_8}{k_{M11,3}} + \frac{x_{11}y_{18}}{k_{M11,1}k_{M11,2}}\right)}$$

$$v_{12}(t;\vec{\theta}) = \frac{V_{23}y_{14}x_8\frac{x_2}{3}}{\left(y_{14} + k_{M23,1}\right)\left(\left(1 + \frac{x_{11}}{k_{I23}}\right)\left(x_8 + k_{M23,2}\right) + \frac{k_{M23,2}y_{18}}{k_{A23,1}x_2 + k_{A23,2}y_{12} + k_{A23,3}x_5}\right)}$$

$$v_{13}(t;\vec{\theta}) = \frac{V_{31}y_{10}\frac{x_2}{3}}{\left(1 + \left(1 + \frac{k_{M31,3}}{y_8}\right)\left(\frac{y_{18}}{k_{M31,2}} + \frac{x_2}{k_{M32}} + \frac{y_{11}}{k_{M33}} + \frac{y_{10}}{k_{M31,1}}\right)\right)k_{M31,1}}$$

$$v_{14}(t;\vec{\theta}) = \frac{V_{32}\frac{x_2^2}{3}}{\left(1 + \left(1 + \frac{k_{M31,3}}{y_8}\right)\left(\frac{y_{18}}{k_{M31,2}} + \frac{x_2}{k_{M32}} + \frac{y_{11}}{k_{M33}} + \frac{y_{10}}{k_{M31,1}}\right)\right)k_{M32}}$$

$$v_{15}(t;\vec{\theta}) = \frac{V_{33}y_{11}\frac{x_2}{3}}{\left(1 + \left(1 + \frac{k_{M31,3}}{y_8}\right)\left(\frac{y_{18}}{k_{M31,2}} + \frac{x_2}{k_{M32}} + \frac{y_{11}}{k_{M33}} + \frac{y_{10}}{k_{M31,1}}\right)\right)k_{M33}}$$

$$v_{16}(t;\vec{\theta}) = \frac{V_{51}\left(y_4y_5 - \frac{x_{33}}{k_{E51}}\right)}{k_{M51,2}k_{M51,3}\left(1 + \frac{y_4}{k_{M51,2}} + \frac{y_5}{k_{M51,2}} + \frac{x_{33}}{k_{M51,1}} + \frac{y_4y_5}{k_{M51,2}k_{M51,3}}\right)}$$

$$v_{17}(t;\vec{\theta}) = \frac{V_{52}\left(x_{33} - \frac{y_2y_8}{k_{E52}}\right)}{k_{M52}\left(1 + \frac{x_{35}}{k_{I52,3}}\right)\left(1 + \frac{x_{33}}{k_{M52}\left(1 + \frac{x_{35}}{k_{I52,3}}\right)} + \frac{y_8}{k_{I52,2}} + \frac{y_2}{k_{I52,1}} + \frac{y_8y_2}{k_{I52,1}k_{I52,2}}\right)}$$

$$v_{18}(t;\vec{\theta}) = \frac{V_{55}\left(x_{37}y_3 - \frac{x_{36}y_9}{k_{E55}}\right)}{k_{M55,1}k_{M55,2}\left(1 + \frac{x_{37}}{k_{M55,1}} + \frac{y_3}{k_{M55,2}} + \frac{x_{36}}{k_{M55,3}} + \frac{y_9}{k_{M55,4}} + \frac{x_{37}y_3}{k_{M55,1}k_{M55,2}} + \frac{x_{36}y_9}{k_{M55,3}k_{M55,4}}\right)}$$

$$v_{19}(t;\vec{\theta}) = \frac{V_{56}\left(y_2x_{36} - \frac{x_{38}y_6}{k_{E56}}\right)}{\left(y_2 + k_{M56,1}\left(1 + \frac{x_{33}}{k_{M56,2}}\right)\right)\left(x_{36} + k_{M56,2}\left(1 + \frac{y_6}{k_{M56,1}}\right)\left(1 + \frac{x_{38}}{k_{I56,3}}\right)\left(1 + \frac{y_8}{k_{I56,4}}\right)\left(1 + \frac{x_{39}}{k_{I56,5}}\right)\right)}$$

$$v_{20}(t;\vec{\theta}) = \frac{V_{57}\left(x_{38} - \frac{x_{39}y_8}{k_{E57}}\right)}{x_{38} + k_{M57}\left(1 + \frac{x_{39}}{k_{I57}}\right)}$$

$$v_{21}(t;\vec{\theta}) = \frac{V_{58}x_{35}}{k_{M58}\left(1 + \frac{x_{35}}{k_{M58}}\right)\left(1 + \frac{y_8}{k_{I58,2}}\right)\left(1 + \frac{y_2}{k_{I58,1}}\right)}$$

$$v_{22}(t;\vec{\theta}) = \frac{V_{59}\left(x_{19}y_2 - \frac{x_{18}x_{35}}{k_{E59}}\right)}{\left(y_2 + k_{M59,3}\left(1 + \frac{x_{35}}{k_{M59,2}}\right)\left(1 + \frac{y_5}{k_{I59,2}}\right)\right)\left(x_{19} + k_{M59,1}\left(1 + \frac{x_{18}}{k_{I59,1}}\right)\right)}$$

$$v_{23}(t;\vec{\theta}) = \frac{V_{60}\left(x_{19}y_6 - \frac{x_{18}x_{37}}{k_{E60}}\right)}{k_{M60,2}k_{M60,3}\left(1 + \frac{x_{19}}{k_{M60,2}} + \frac{y_6}{k_{M60,3}} + \frac{x_{19}y_6}{k_{M60,2}k_{M60,3}} + \frac{x_{18}}{k_{M60,1}} + \frac{x_{37}}{k_{M60-4}} + \frac{x_{18}x_{37}}{k_{M60,1}k_{M60-4}}\right)}$$

$$v_{24}(t;\vec{\theta}) = \frac{V_{62}x_{39}x_{19}}{x_{39} + x_{19} + k_{M62,1}}$$

$$v_{25}(t;\vec{\theta}) = \frac{V_{111}x_1 \min\left(1, x_1/E_t\right)}{x_1 + k_{M1,3}\left(1 + \frac{x_2}{k_{I1,1}} + \frac{x_5}{k_{I1,2}} + \frac{x_9}{k_{I1,3}} + \frac{y_{18}}{k_{I1,4}} + \frac{x_{10}}{k_{I1,5}}\right)} \frac{O_2}{O_2 + k_{M1,2}\left(1 + \frac{CO_2}{k_{M1,1}}\right)}$$

$$v_{26}(t;\vec{\theta}) = \frac{V_{112}x_{25}}{x_{25} + k_{M112}\left(1 + \frac{x_{24}}{k_{I112,2}}\right)\left(1 + \frac{y_{18}}{k_{I112,1}}\right)}$$

$$v_{27}(t;\vec{\theta}) = \frac{V_{113}\left(x_8 x_{23} - \frac{x_{11}x_2}{k_{E113}}\right)}{\left(x_8 + k_{M113,1}\left(1 + \frac{x_2}{k_{I113}}\right)\right)(x_{23} + k_{M113,2})}$$

$$v_{28}(t;\vec{\theta}) = \frac{V_{121}x_{26}}{x_{26} + k_{M121}}$$

$$v_{29}(t;\vec{\theta}) = \frac{V_{122}\left(x_{27}x_{28} - \frac{x_{30}x_{29}}{k_{E122}}\right)}{(x_{27} + k_{M122,1})\left(x_{28} + k_{M122,2}\left(1 + \frac{x_{29}}{k_{I122}}\right)\right)}$$

$$v_{30}(t;\vec{\theta}) = \frac{V_{123}\left(x_{30}x_{16} - \frac{x_{23}x_{17}}{k_{E123}}\right)}{x_{30} + k_{M123}\left(1 + \frac{x_{30}}{k_{I123}}\right)}$$

$$v_{31}(t;\vec{\theta}) = \frac{V_{124}\left(x_{27}x_{20} - \frac{x_{21}x_{29}}{k_{E124}}\right)}{(x_{27} + k_{M124,1})\left(x_{20} + k_{M124,2}\left(1 + \frac{x_{29}}{k_{I124}}\right)\right)}$$

$$v_{32}(t;\vec{\theta}) = \frac{V_{101,1}x_{31}}{x_{31} + k_{M101,1}\left(1 + \frac{x_{26}}{k_{I101,1}}\right)} - \frac{V_{101,1}x_{23}}{x_{23} + k_{M101,1}\left(1 + \frac{x_{24}}{k_{I101,1}}\right)}$$

$$v_{33}(t;\vec{\theta}) = \frac{V_{101,2}x_{24}}{x_{24} + k_{M101,2}\left(1 + \frac{x_{23}}{k_{I101,2}}\right)} - \frac{V_{101,2}x_{26}}{x_{26} + k_{M101,2}\left(1 + \frac{x_{31}}{k_{I101,2}}\right)}$$

$$v_{34}(t;\vec{\theta}) = \frac{x_{40}}{2\left(x_{40} + 1\right)}$$

$$v_{35}(t;\vec{\theta}) = \frac{V_{131}x_{29}}{x_{29} + k_{M131,1}\left(1 + \frac{x_{28}}{k_{I131,1}}\right)}$$

$$v_{36}(t;\vec{\theta}) = \frac{V_{36}x_{17}x_{10} - x_{16}x_{12}/4}{(x_{17} + k_{M36,1})(x_{10} + k_{M36,2})}$$

$$v_{37}(t;\vec{\theta}) = \frac{V_{37} \cdot PQ \cdot r(t) \cdot w(t)}{k_{M37} + PQ}$$

$$v_{38}(t;\vec{\theta}) = \frac{V_{38}k_{M38,1}^{-1}k_{M38,2}^{-1}\left(PQH \cdot PC^+ - \frac{PC \cdot PQ}{k_{E38}\left(\frac{H_S^+}{H_L^+}\right)^2}\right)}{1 + \frac{PQH_2}{k_{M38,1}} + \frac{PC^+}{k_{M38,2}} + \frac{PC}{k_{M38,3}} + \frac{PQ}{k_{M38,4}} + \frac{PQH \cdot PC^+}{k_{M38,1}k_{M38,2}} + \frac{PC \cdot PQ}{k_{M38,3}k_{M38,4}}}$$

$$v_{39}(t;\vec{\theta}) = \frac{V_{39} \cdot PC \cdot Fd \cdot r(t)}{(k_{M39,1} + PC)(k_{M39,2} + Fd)}$$

$$v_{40}(t;\vec{\theta}) = \frac{V_{40}k_{M40,1}^{-1}k_{M40,2}^{-1}(Fd^- \cdot NADP - Fd \cdot NADPH/k_{E40})}{1 + \frac{Fd^-}{k_{M40,1}} + \frac{NADP}{k_{M40,2}} + \frac{Fd}{k_{M40,3}} + \frac{NADPH}{k_{M40,4}} + \frac{Fd^- \cdot NADP}{k_{M40,1}k_{M40,2}} + \frac{Fd \cdot NADPH}{k_{M40,3}k_{M40,4}}}$$

$$v_{41}(t;\vec{\theta}) = \frac{V_{41}x_{18}y_8}{(x_{18} + k_{M41,1})(y_8 + k_{M41,2})}$$

$$
\begin{aligned}
y_1(t;\vec{\theta}) &= \frac{x_{34}}{1 + k_{E22} + 1/k_{E21}} \\
y_2(t;\vec{\theta}) &= \frac{x_{34}}{1 + k_{E21} + k_{E21}k_{E22}} \\
y_3(t;\vec{\theta}) &= \frac{k_{E22}x_{34}}{1 + k_{E22} + 1/k_{E21}} \\
y_4(t;\vec{\theta}) &= \frac{x_{32}}{1 + k_{E4}} \\
y_5(t;\vec{\theta}) &= \frac{k_{E4}x_{32}}{1 + k_{E4}} \\
y_6(t;\vec{\theta}) &= \text{UTc} - x_{36} - x_{37} \\
y_7(t;\vec{\theta}) &= \text{PTc} - 2(x_{33} + x_{35}) - x_{40} - x_{32} - x_{34} - x_{37} - x_{38} - x_{19} \\
y_8(t;\vec{\theta}) &= k_{E61}^2 - 4k_{E61}(\text{PTc} - 2(x_{33} + x_{35}) - x_{40} - x_{32} - x_{34} - x_{37} - x_{38} - x_{19}
\end{aligned}
$$

B.2 First-order light reaction equations from [28]

$$p_1(t; \vec{\theta}) = k_{\text{(HC)L}} x_{42} - k_{\text{L(HC)}} x_{52}$$

$$p_2(t; \vec{\theta}) = k_{\text{(HC)L}} x_{43} - k_{\text{L(HC)}} x_{53}$$

$$p_3(t; \vec{\theta}) = k_{\text{(HC)L}} x_{44} - k_{\text{L(HC)}} x_{48}$$

$$p_4(t; \vec{\theta}) = k_{\text{(HC)L}} x_{45} - k_{\text{L(HC)}} x_{49}$$

$$p_5(t; \vec{\theta}) = k_{ct}^b x_{42} x_{54} - k_{ct}^f x_{44} x_{55}$$

$$p_6(t; \vec{\theta}) = k_{ct}^b x_{46} x_{54} - k_{ct}^f x_{42} x_{55}$$

$$p_7(t; \vec{\theta}) = k_{ct}^b x_{43} x_{54} - k_{ct}^f x_{45} x_{55}$$

$$p_8(t; \vec{\theta}) = k_{ct}^b x_{47} x_{54} - k_{ct}^f x_{43} x_{55}$$

$$p_9(t; \vec{\theta}) = k_{ct}^b x_{52} x_{54} - k_{ct}^f x_{48} x_{55}$$

$$p_{10}(t; \vec{\theta}) = k_{ct}^b x_{49} x_{54} - k_{ct}^f x_{51} x_{55}$$

$$p_{11}(t; \vec{\theta}) = k_{ct}^b x_{53} x_{54} - k_{ct}^f x_{49} x_{55}$$

$$p_{12}(t; \vec{\theta}) = k_x^f x_{78} x_{54} - k_x^b x_{77} x_{55}$$

$$p_{13}(t; \vec{\theta}) = k_x^f x_{76} x_{54} - k_x^b x_{75} x_{55}$$

$$p_{14}(t; \vec{\theta}) = k_{\text{Fd}}^f x_{57} x_{54} - k_{\text{Fd}}^b x_{56} x_{55}$$

$$p_{15}(t; \vec{\theta}) = k_{\text{Fd}}^f x_{58} x_{54} - k_{\text{Fd}}^b x_{57} x_{55}$$

$$p_{16}(t; \vec{\theta}) = k_a x_{59}$$

$$p_{17}(t; \vec{\theta}) = k_{\text{FNR}}^f x_{58} x_{12} - k_{\text{FNR}}^b x_{56} x_{10}$$

$$p_{18}(t; \vec{\theta}) = k_{ct}^{b2} x_{48} x_{54} - k_{ct}^{f2} x_{50} x_{55}$$

$$p_{19}(t; \vec{\theta}) = k_{\text{F}}^b x_{42} x_{79} - k_{\text{F}}^f x_{43} x_{80}$$

$$p_{20}(t; \vec{\theta}) = k_{\text{F}}^b x_{44} x_{79} - k_{\text{F}}^f x_{45} x_{80}$$

$$p_{21}(t; \vec{\theta}) = k_{\text{F}}^b x_{46} x_{79} - k_{\text{F}}^f x_{47} x_{80}$$

$$p_{22}(t; \vec{\theta}) = k_{\text{F}}^b x_{48} x_{79} - k_{\text{F}}^f x_{49} x_{80}$$

$$p_{23}(t; \vec{\theta}) = k_{\text{F}}^b x_{50} x_{79} - k_{\text{F}}^f x_{51} x_{80}$$

$$p_{24}(t; \vec{\theta}) = k_{\text{F}}^b x_{52} x_{79} - k_{\text{F}}^f x_{53} x_{80}$$

$$p_{25}(t; \vec{\theta}) = k_{\text{LF}}^b x_{43} x_{81} - k_{\text{LF}}^f x_{48} x_{82}$$

$$p_{26}(t; \vec{\theta}) = k_{\text{LF}}^b x_{45} x_{81} - k_{\text{LF}}^f x_{50} x_{82}$$

$$p_{27}(t; \vec{\theta}) = k_{\text{LF}}^b x_{47} x_{81} - k_{\text{LF}}^f x_{52} x_{82}$$

$$p_{28}(t; \vec{\theta}) = k_{\text{(HC)}}^f x_{46} x_{81} - k_{\text{(HC)}}^b x_{44} x_{82}$$

$$p_{29}(t; \vec{\theta}) = k_{\text{(HC)}}^f x_{47} x_{81} - k_{\text{(HC)}}^b x_{45} x_{82}$$

$$p_{30}(t; \vec{\theta}) = k_{\text{(HC)}}^f x_{52} x_{81} - k_{\text{(HC)}}^b x_{50} x_{82}$$

$$p_{31}(t; \vec{\theta}) = k_{\text{(HC)}}^f x_{53} x_{81} - k_{\text{(HC)}}^b x_{51} x_{82}$$

$$p_{32}(t; \vec{\theta}) = k_{01} x_{60} x_{83}$$

$$p_{33}(t; \vec{\theta}) = k_{01} x_{61} x_{83}$$

$$p_{34}(t; \vec{\theta}) = k_{01} x_{62} x_{83}$$

$$p_{35}(t; \vec{\theta}) = k_{01} x_{63} x_{83}$$

$$p_{36}(t; \vec{\theta}) = k_{01} x_{64} x_{83}$$

$$p_{37}(t; \vec{\theta}) = k_{01} x_{65} x_{83}$$

$$p_{38}(t; \vec{\theta}) = k_{01} x_{66} x_{83}$$

$$p_{39}(t; \vec{\theta}) = k_{12} x_{60} x_{84}$$

$$p_{40}(t; \vec{\theta}) = k_{12} x_{61} x_{84}$$

$$p_{41}(t; \vec{\theta}) = k_{12} x_{62} x_{84}$$

$$p_{42}(t; \vec{\theta}) = k_{12} x_{63} x_{84}$$

$$p_{43}(t; \vec{\theta}) = k_{12} x_{64} x_{84}$$

$$p_{44}(t; \vec{\theta}) = k_{12} x_{65} x_{84}$$

$$p_{45}(t; \vec{\theta}) = k_{12} x_{66} x_{84}$$

$$p_{46}(t; \vec{\theta}) = k_{23} x_{60} x_{85}$$

$$p_{47}(t; \vec{\theta}) = k_{23} x_{61} x_{85}$$

$$p_{48}(t; \vec{\theta}) = k_{23} x_{62} x_{85}$$

$$p_{49}(t; \vec{\theta}) = k_{23} x_{63} x_{85}$$

$$p_{50}(t; \vec{\theta}) = k_{23} x_{64} x_{85}$$

$$p_{51}(t; \vec{\theta}) = k_{23} x_{65} x_{85}$$

$$p_{52}(t; \vec{\theta}) = k_{23} x_{66} x_{85}$$

$$p_{53}(t; \vec{\theta}) = k_{34} x_{60} x_{85}$$

$$p_{54}(t; \vec{\theta}) = k_{34} x_{61} x_{86}$$

$$p_{55}(t; \vec{\theta}) = k_{34} x_{62} x_{86}$$

$$p_{56}(t; \vec{\theta}) = k_{34} x_{63} x_{86}$$

$$p_{57}(t; \vec{\theta}) = k_{34} x_{64} x_{86}$$

$$p_{58}(t; \vec{\theta}) = k_{34} x_{65} x_{86}$$

$$p_{59}(t; \vec{\theta}) = k_{34} x_{66} x_{86}$$

$$p_{60}(t; \vec{\theta}) = k_{40} x_{87} w(t)$$

$$
\begin{aligned}
p_{61}(t;\vec{\theta}) &= k_{\mathrm{L}1}r(t)x_{77} \\
p_{62}(t;\vec{\theta}) &= k_{\mathrm{L}2}r(t)x_{72} \\
p_{63}(t;\vec{\theta}) &= k_{\mathrm{L}2}r(t)x_{73} \\
p_{64}(t;\vec{\theta}) &= k_{\mathrm{L}2}r(t)x_{74} \\
p_{65}(t;\vec{\theta}) &= k_{AB1}x_{61} - k_{BA1}x_{65} \\
p_{66}(t;\vec{\theta}) &= k_{AB1}x_{68} - k_{BA1}x_{73} \\
p_{67}(t;\vec{\theta}) &= k_{AB2}x_{62} - k_{BA2}x_{66} \\
p_{68}(t;\vec{\theta}) &= k_{AB2}x_{69} - k_{BA2}x_{74} \\
p_{69}(t;\vec{\theta}) &= k_B^f x_{63}x_{81} - k_B^b x_{61}x_{82} \\
p_{70}(t;\vec{\theta}) &= k_B^f x_{70}x_{81} - k_B^b x_{68}x_{82} \\
p_{71}(t;\vec{\theta}) &= k_B^f x_{66}x_{81} - k_B^b x_{64}x_{82} \\
p_{72}(t;\vec{\theta}) &= k_B^f x_{74}x_{81} - k_B^b x_{72}x_{82} \\
p_{73}(t;\vec{\theta}) &= k_R^f x_{75}x_{79} - k_R^b x_{77}x_{80} \\
p_{74}(t;\vec{\theta}) &= k_R^f x_{76}x_{79} - k_R^b x_{78}x_{80}
\end{aligned}
$$

B.3 Comprehensive model state variable differential equations

$$\dot{x}_1 = v_{10} - v_1 - v_{25}$$

$$\dot{x}_2 = 2v_1 - v_2 - v_{14} + v_{27} + v_{25}$$

$$\dot{x}_3 = v_2 - v_3$$

$$\dot{x}_4 = v_3 - 2v_4 - v_6 - v_7 - v_9 - v_{13} - v_{15}$$

$$\dot{x}_5 = v_4 - v_5$$

$$\dot{x}_6 = v_6 - v_7$$

$$\dot{x}_7 = v_8 - v_9$$

$$\dot{x}_8 = v_{11} - v_2 - v_{12} - v_{10} - v_{27}$$

$$\dot{x}_9 = v_7 - v_8$$

$$\dot{x}_{10} = -v_3 + p_{17}$$

$$\dot{x}_{11} = v_2 + v_{10} - v_{11} + v_{12} + v_{27}$$

$$\dot{x}_{12} = v_3 - p_{17}$$

$$\dot{x}_{13} = v_5 - v_6 - v_{12}$$

$$\dot{x}_{14} = v_6 + 2v_9 - v_{10}$$

$$\dot{x}_{15} = 0$$

$$\dot{x}_{16} = v_{30} - v_{35} + v_{36}$$

$$\dot{x}_{17} = -v_{30} + v_{35} - v_{36}$$

$$\dot{x}_{18} = v_{22} + v_{23} + V_{APS}v_{24}$$

$$\dot{x}_{19} = -v_{22} - v_{23} - V_{APS}v_{24}$$

$$\dot{x}_{20} = 0$$

$$\dot{x}_{21} = 0$$

$$\dot{x}_{22} = 0$$

$$\dot{x}_{23} = v_{32} - v_{27}$$

$$\dot{x}_{24} = v_{26} - v_{33}$$

$$\dot{x}_{25} = v_{25} - v_{26}$$

$$\dot{x}_{26} = v_{33} - v_{28}$$

$$\dot{x}_{27} = v_{28} - v_{29} - v_{31}$$

$$\dot{x}_{28} = v_{35} - v_{29}$$

$$\dot{x}_{29} = v_{29} + v_{31} - 2v_{35}$$

$$\dot{x}_{30} = v_{29} - v_{30}$$

$$\dot{x}_{31} = v_{30} - v_{32}$$

$$\dot{x}_{32} = v_{13} - 2v_{16} + v_{15}$$

$$\dot{x}_{33} = v_{16} - v_{17}$$

$$\dot{x}_{34} = v_{17} - v_{18} - v_{22} + v_{21} - v_{19}$$

$$\dot{x}_{35} = v_{22} - v_{21}$$

$$\dot{x}_{36} = v_{18} - v_{19}$$

$$\dot{x}_{37} = v_{23} - v_{18}$$

$$\dot{x}_{38} = v_{19} - v_{20}$$

$$\dot{x}_{39} = v_{20} - v_{24}$$

$$\dot{x}_{40} = -v_{34} + v_{14}$$

$$\dot{x}_{41} = v_1 - v_{35}$$

$$\dot{x}_{42} = -p_1 - p_5 + p_6 - p_{19}$$

$$\dot{x}_{43} = -p_{25} - p_2 - p_7 + p_8 + p_{19}$$

$$\dot{x}_{44} = -p_3 + p_5 + p_{28} - p_{20}$$

$$\dot{x}_{45} = -p_{26} - p_4 + p_7 + p_{29} + p_{20}$$

$$\dot{x}_{46} = -p_6 - p_{28} - p_{21}$$

$$\dot{x}_{47} = -p_{27} - p_8 - p_{29} + p_{21}$$

$$\dot{x}_{48} = p_{25} + p_3 - p_{18} + p_9 - p_{22}$$

$$\dot{x}_{49} = p_4 - p_{10} + p_{11} + p_{22}$$

$$\dot{x}_{50} = p_{26} + p_{18} + p_{30} - p_{23}$$

$$\dot{x}_{51} = p_{10} + p_{31} + p_{23}$$

$$\dot{x}_{52} = p_{27} + p_1 - p_9 - p_{30} - p_{24}$$

$$\dot{x}_{53} = p_2 - p_{11} - p_{31} + p_{24}$$

$$\dot{x}_{54} = -p_{18} - p_5 - p_{10} - p_7 - p_9 - p_6 - p_{11}$$
$$\quad - p_8 - p_{12} - p_{13} - p_{14} - p_{15}$$

$$\dot{x}_{55} = p_{18} + p_5 + p_{10} + p_7 + p_9 + p_6 + p_{11}$$
$$\quad + p_8 + p_{12} + p_{13} + p_{14} + p_{15}$$

$$\dot{x}_{56} = p_{16} + p_{14} + p_{17}$$

$$\dot{x}_{57} = -p_{14} + p_{15}$$

$$\dot{x}_{58} = -p_{15} - p_{17}$$

$$\dot{x}_{59} = -p_{16}$$

$$\dot{x}_{60} = -p_{39} - (p_{32} + p_{46} + p_{53})$$

$$\dot{x}_{61} = p_{62} - (p_{33} + p_{40} + p_{47} + p_{54}) - p_{65} + p_{69}$$

$$\dot{x}_{62} = p_{63} - (p_{34} + p_{41} + p_{48} + p_{55}) - p_{67}$$

$$\dot{x}_{63} = p_{64} - (p_{35} + p_{42} + p_{49} + p_{56}) - p_{69}$$

$$\dot{x}_{64} = -(p_{36} + p_{43} + p_{50} + p_{57}) + p_{71}$$

$$\dot{x}_{65} = -(p_{37} + p_{44} + p_{51} + p_{58}) + p_{65}$$

$$\dot{x}_{66} = -(p_{38} + p_{45} + p_{52} + p_{59}) - p_{71} + p_{67}$$

$$\dot{x}_{67} = p_{32} + p_{46} + p_{53}$$

$$\dot{x}_{68} = (p_{33} + p_{40} + p_{47} + p_{54}) - p_{66} + p_{70}$$

$$\dot{x}_{69} = (p_{34} + p_{41} + p_{48} + p_{55}) - p_{68}$$

$$\dot{x}_{70} = (p_{35} + p_{42} + p_{49} + p_{56}) - p_{70}$$

$$\dot{x}_{71} = p_{39}$$

$$\dot{x}_{72} = -p_{62} + (p_{36} + p_{43} + p_{50} + p_{57}) + p_{72}$$

$$\dot{x}_{73} = -p_{63} + (p_{37} + p_{44} + p_{51} + p_{58}) + p_{66}$$

$$\dot{x}_{74} = -p_{64} + (p_{38} + p_{45} + p_{52} + p_{59}) + p_{68} - p_{72}$$

$$\dot{x}_{75} = p_{13} - p_{73}$$

$$\dot{x}_{76} = p_{61} - p_{13} - p_{74}$$

$$\dot{x}_{77} = -p_{61} + p_{12} + p_{73}$$

$$\dot{x}_{78} = -p_{12} + p_{74}$$

$$\dot{x}_{79} = -p_{23} - p_{20} - p_{22} - p_{19} - p_{24} - p_{21} - p_{73} - p_{74}$$

$$\dot{x}_{80} = p_{23} + p_{20} + p_{22} + p_{19} + p_{24} + p_{21} + p_{73} + p_{74}$$

$$\dot{x}_{81} = -p_{71} - p_{72} - p_{69} - p_{70} - p_{26} - p_{25} - p_{27}$$
$$-p_{30} - p_{28} - p_{31} - p_{29}$$

$$\dot{x}_{82} = p_{71} + p_{72} + p_{69} + p_{70} + p_{26} + p_{25} + p_{27}$$
$$+p_{30} + p_{28} + p_{31} + p_{29}$$

$$\dot{x}_{83} = -p_{36} - p_{32} - p_{33} - p_{37} - p_{34} - p_{38} - p_{35} + p_{60}$$

$$\dot{x}_{84} = p_{36} - p_{43} + p_{32} - p_{40} + p_{33} - p_{39} + p_{37}$$
$$-p_{44} + p_{34} - p_{41} + p_{38} - p_{45} + p_{35} - p_{42}$$

$$\dot{x}_{85} = p_{43} - p_{50} + p_{40} - p_{47} + p_{39} + p_{44} - p_{51}$$
$$-p_{46} + p_{41} - p_{48} - p_{53} + p_{45} - p_{52} + p_{42} - p_{49}$$

$$\dot{x}_{86} = p_{50} - p_{57} + p_{47} - p_{54} + p_{51} - p_{58} + p_{46}$$
$$+p_{48} - p_{55} + p_{52} - p_{59} + p_{49} - p_{56}$$

$$\dot{x}_{87} = p_{56} + p_{53} + p_{57} + p_{54} + p_{58} + p_{55} + p_{59} - p_{60}$$

$$\dot{x}_{88} = 0.001 \left(16 p_{60} - k_{HPR} v_{40} \right)$$

$$\dot{x}_{89} = 0.001 \left(k_{HPR} v_{40} - 4 p_{17} - v3 \right)$$

$$\dot{x}_{90} = r_G / A_{opt} x_{90} \left(1 - x_{90} / k_G \right)$$

$$\dot{x}_{91} = r_S x_{91} \left(1 - x_{91} / k_G \right)$$

$$\dot{x}_{92} = \dot{x}_{90} \dot{x}_{41} - \dot{x}_{91}$$

B.4 Compact model state variable differential equations

$$\dot{x}_1 = v_{10} - v_1 - v_{25}$$

$$\dot{x}_2 = 2v_1 - v_2 - v_{14} + v_{27} + v_{25}$$

$$\dot{x}_3 = v_2 - v_3$$

$$\dot{x}_4 = v_3 - 2v_4 - v_6 - v_7 - v_9 - v_{13} - v_{15}$$

$$\dot{x}_5 = v_4 - v_5$$

$$\dot{x}_6 = v_6 - v_7$$

$$\dot{x}_7 = v_8 - v_9$$

$$\dot{x}_8 = v_{11} - v_2 - v_{12} - v_{10} - v_{27}$$

$$\dot{x}_9 = v_7 - v_8$$

$$\dot{x}_{10} = v_{40} - v_3 - v_{36}$$

$$\dot{x}_{11} = v_2 + v_{10} - v_{11} + v_{12} + v_{27}$$

$$\dot{x}_{12} = v_3 + v_{36} - v_{40}$$

$$\dot{x}_{13} = v_5 - v_6 - v_{12}$$

$$\dot{x}_{14} = v_6 + 2v_9 - v_{10}$$

$$\dot{x}_{15} = 0$$

$$\dot{x}_{16} = v_{30} - v_{35} + v_{36}$$

$$\dot{x}_{17} = -v_{30} + v_{35} - v_{36}$$

$$\dot{x}_{18} = v_{22} + v_{23} - v_{41}$$

$$\dot{x}_{19} = -v_{22} - v_{23} + v_{41}$$

$$\dot{x}_{20} = 0$$

$$\dot{x}_{21} = 0$$

$$\dot{x}_{22} = 0$$

$$\dot{x}_{23} = v_{32} - v_{27}$$

$$\dot{x}_{24} = v_{26} - v_{33}$$

$$\dot{x}_{25} = v_{25} - v_{26}$$

$$\dot{x}_{26} = v_{33} - v_{28}$$

$$\dot{x}_{27} = v_{28} - v_{29} - v_{31}$$

$$\dot{x}_{28} = v_{35} - v_{29}$$

$$\dot{x}_{29} = v_{29} + v_{31} - 2v_{35}$$

$$\dot{x}_{30} = v_{29} - v_{30}$$

$$\dot{x}_{31} = v_{30} - v_{32}$$

$$\dot{x}_{32} = v_{13} - 2v_{16} + v_{15}$$

$$\dot{x}_{33} = v_{16} - v_{17}$$

$$\dot{x}_{34} = v_{17} - v_{18} - v_{22} + v_{21} - v_{19}$$

$$\dot{x}_{35} = v_{22} - v_{21}$$

$$\dot{x}_{36} = v_{18} - v_{19}$$

$$\dot{x}_{37} = v_{23} - v_{18}$$

$$\dot{x}_{38} = v_{19} - v_{20}$$

$$\dot{x}_{39} = v_{20} - v_{24}$$

$$\dot{x}_{40} = -v_{34} + v_{14}$$

$$\dot{x}_{41} = -2v_{37} + v_{38}$$

$$\dot{x}_{42} = 2v_{37} - v_{38}$$

$$\dot{x}_{43} = -2v_{38} + v_{39}$$

$$\dot{x}_{44} = 2v_{38} - v_{39}$$

$$\dot{x}_{45} = 0.01(6v38 - .75k_{HPR}v_{11})$$

$$\dot{x}_{46} = 0.01(-3v_{38} - v_{40} + k_{HPR}v_{11} - v3)$$

$$\dot{x}_{47} = -v_{39} + 2v_{40}$$

$$\dot{x}_{48} = v_{39} - 2v_{40}$$

$$\dot{x}_{49} = 0.001(v_1 - v35)$$

$$\dot{x}_{50} = r_G/A_{opt}x_{50}\left(1 - x_{50}/k_G\right)$$

$$\dot{x}_{51} = r_Sx_{51}\left(1 - x_{51}/k_S\right)$$

$$\dot{x}_{52} = \dot{x}_{50}\dot{x}_{49} - \dot{x}_{51}$$