ABSTRACT

Title of Dissertation:	A FRAMEWORK FOR CREDIBILITY ASSESSMENT OF SUBJECT-SPECIFIC PHYSIOLOGICAL MODELS
	Bahram Parvinian, Doctor of Philosophy, 2022
Dissertation directed by:	Professor Jin-Oh Hahn, Department of Mechanical Engineering

Physiological closed-loop controllers and decision support systems are medical devices that enable some degree of automation to meet the needs of patients in resource-limited environments such as critical care and surgical units. Traditional methods of safety and effectiveness evidence generation such as pre-clinical animal and human clinical studies are cost prohibitive and may not fully capture different performance attributes of such complex safety-critical systems primarily due to subject variability. *In silico* studies using subject-specific physiological models (SSPMs) may provide a versatile platform to generate pre-clinical and clinical safety evidence for medical devices and help reduce the size and scope of animal studies and/or clinical trials. To achieve such a goal, the credibility of the SSPMs must be established for the purpose it is intended to serve.

While in the past decades significant research has been dedicated towards development of tools and methods for development and evaluation of SSPMs, adoption of such models remains

limited, partly due to lack of trust in SSPMs for safety-critical applications. This may be due to a lack of a cohesive and disciplined credibility assessment framework for SSPMs.

In this dissertation a novel framework is proposed for credibility assessment of SSPMs. The framework combines various credibility activities in a unified manner to avoid or reduce resource intensive steps, effectively identify model or data limitations, provide direction as to how to address potential model weaknesses, and provide much needed transparency in the model evaluation process to the decision-makers. To identify various credibility activities, the framework is informed by an extensive literature review of more mature modeling spaces focusing on non-SSPMs as well as a literature review identifying gaps in the published work related to SSPMs. The utility of the proposed framework is successfully demonstrated by its application towards credibility assessment of a CO_2 ventilatory gas exchange model intended to predict physiological parameters, and a blood volume kinetic model intended to predict changes in blood volume in response to fluid resuscitation and hemorrhage. The proposed framework facilitates development of more reliable SSPMs and will result in increased adoption of such models to be used for evaluation of safety-critical medical devices such as Clinical Decision Support (CDS) and Physiological Closed-Loop Controlled (PCLC) systems.

A FRAMEWORK FOR CREDIBILITY ASSESSMENT OF SUBJECT-SPECIFIC PHYSIOLOGICAL MODELS

by

Bahram Parvinian

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Dr. Jin-Oh Hahn, Chair/Advisor, Mechanical Engineering
Dr. Prasanna Pathmanathan, Advisor, Food and Drug Administration
Dr. Balakumar Balachandran, Mechanical Engineering
Dr. Monifa Vaughn-Cooke, Mechanical Engineering
Dr. Yancy Diaz-Mercado, Mechanical Engineering

Dean's Representative:
Dr. Yang Tao, Bioengineering

Dedication

For Roya and my late grandmother

Acknowledgements

The work in this dissertation would not have initiated, continued, and concluded if it were not for the advice, encouragement, and support of many individuals in my life in the past nine years.

The idea for starting a PhD focusing on mathematical models of semi-autonomous safety-critical systems was planted in me by my former FDA mentor and supervisor Dr. Lester Schultheis. I'm grateful for his vision and for being my only mentor and supervisor in the past 15 years who genuinely cared and inspired.

I am grateful to Dr. Jin-Oh Hahn, my primary UMD advisor, for his support throughout the years, for recognizing the merit of this work, and its regulatory science application. I hope this work starts a new research direction in his lab; one that will ultimately lead to an increase in adoption of mathematical models and their utility towards evaluation of PCLC and CDS medical devices.

I would like to express my special gratitude to Dr. Prasanna Pathmanathan, my FDA advisor, who was instrumental towards the progress of this research. I never ended a meeting with him not learning a new concept, idea, or a new way of thinking about complex issues.

I am grateful to my former colleagues at the FDA, Dr. Tina Morrison, for her tireless efforts towards promoting modeling and simulation as a tool for medical device evaluation and Dr. Christopher Scully for his efforts supporting the PCLC team and public workshop.

I would like to thank Drs. Balachandran,Vaughn-Cooke, Diaz-Mercado, and Tao for agreeing to serve on my dissertation committee.

My sincere appreciation goes to Drs. George Kramer and Mark Ansermino for collecting the experimental data needed for this work.

I am grateful to Dr. Ramin Bighamian for answering all my questions particularly when I started the research phase of this work.
I am grateful for the love and affection I received from Dr. M.T.J in the past months. I hope she recovers soon from her unexpected illness.

Sincere thanks to my friends who genuinely cared about my inspirations and never ceased to check on me and my progress during this long journey.

Finally, and most importantly, I want to express my heartfelt appreciation for the unwavering love and support I received from my parents and siblings. It would not have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BP – blood pressure
BIS– bispectral index
CDS – clinical decision support
CF– cost function
CO – cardiac output
COU – context of use
CPM – computational patient model
CQE– calibration quality evaluation
$EtCO_2$ – end-tidal partial pressure of carbon dioxide
LSI – locally structurally identifiable
MAP – mean arterial pressure
MV– minute ventilation
PCLC – physiological closed-loop control
PIA– practical identifiability analysis
PSA– precalibration sensitivity analysis
PID – proportional-integral-derivative
PK/PD – pharmacokinetic-pharmacodynamic
SAS– semiautonomous system
SIA– structural identifiability analysis
SNP – sodium nitroprusside
RMSE–root mean square error
RMSNE– root mean square normalized error
SSPM– subject specific physiological model
UQ– uncertainty quantification
V&V – verification and validation

Chapter 1: Introduction[1]

1.1 Motivation

Subject-specific Physiological Models (SSPMs) for prediction of physiological parameters and variables such as blood volume, cardiac output, lung volume in response to hemorrhage have been developed in the past decades. As compared with other forms of evidence such as randomized controlled trials (RCTs) and animal studies, *in silico* studies (from computational testing) using these physiological models may provide an effective tool to generate safety evidence for the purpose of regulatory approval of medical devices intended for mechanical ventilation, fluid resuscitation, and infusion drug delivery. For medical devices such as clinical decision support (CDS) and physiological closed-loop controlled (PCLCS) systems, hereafter jointly referred to as semi-autonomous systems (SAS), the Food and Drug Administration (FDA) held a public workshop in 2015 [1] in which the potential for submitting *in silico* evidence using SSPMs was discussed. As an outcome of this meeting, the agency encouraged the use of *in silico* evidence to support preclinical and clinical safety assessment. The FDA and SAS manufacturers agreed that physiological models have the potential to assist in reducing the size and scope of clinical and/or animal studies that are needed to determine the safety and effectiveness of such medical devices [1].

Despite having a long history of development, SSPMs have mainly stayed in the realm of academic studies and not transitioned fully towards industrial adoption by medical device manufacturers and regulatory bodies such as the FDA. One reason for this minimal adoption could

[1] The material described in this chapter was published in the following article: B. Parvinian, C. Scully, H. Wiyor, A. Kumar, and S. Weininger, "Regulatory Considerations for Physiological Closed-Loop Controlled Medical Devices Used for Automated Critical Care: Food and Drug Administration Workshop Discussion Topics," *Anesth Analg*, 2017.

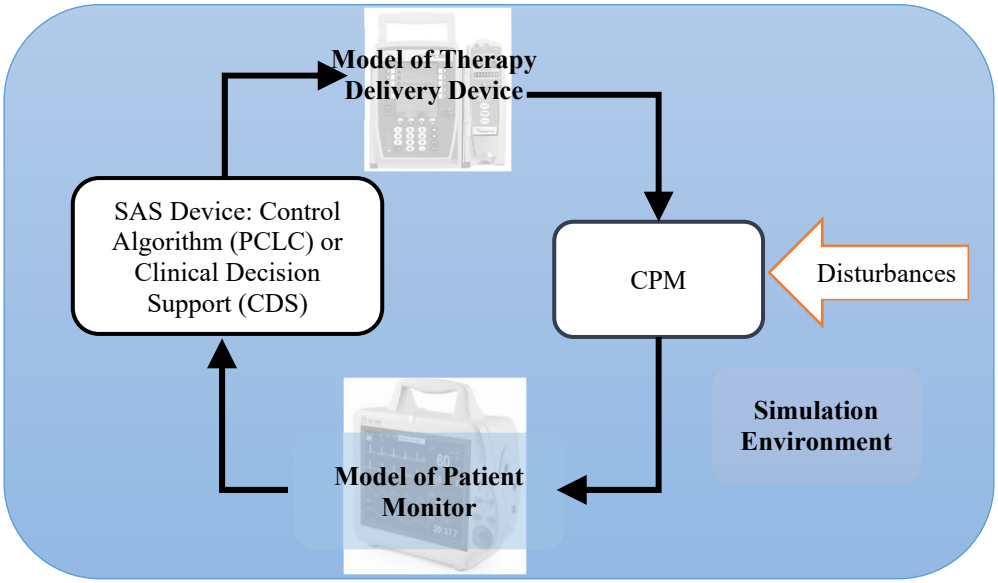
be attributed to the lack of trust in the credibility of those SSPMs. While the credibility of the existing physiological models has occasionally been evaluated as part of individual published work, such model assessment activities are typically *ad hoc* and lack a systematic and well-defined process for evaluation. As such, a unified framework by which credibility evidence and activities can be presented in a systematic and trackable fashion to the decision-makers is urgently needed. The aim of this work is to introduce such a framework in an effort to take the first step in bridging the gap between academic application of SSPMs and their utility towards clinical and regulatory applications.

1.2 Background

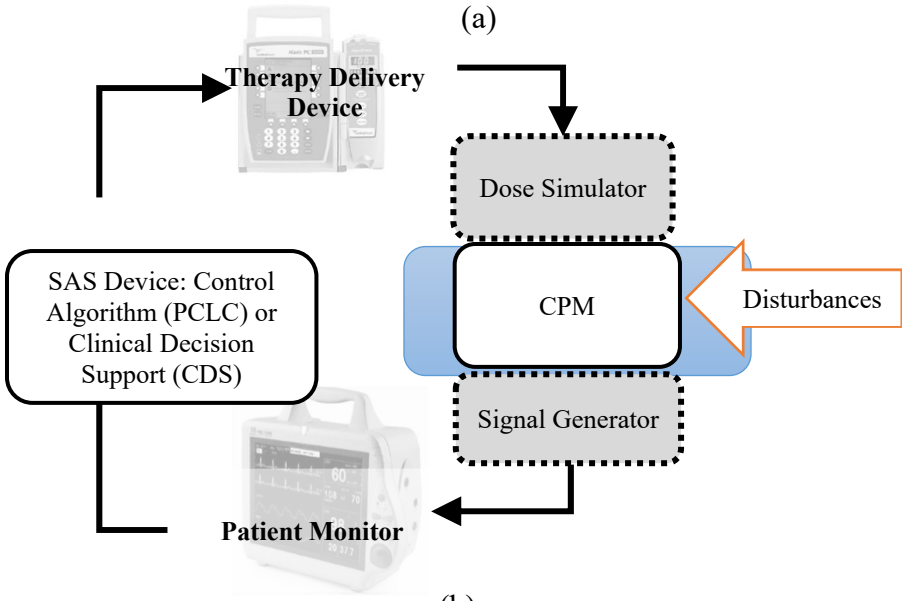
Clinical environments such as critical care and surgical units could benefit from SASs because of the high number of required clinical actions and extensive monitoring/therapeutic devices already in use[1,2]. During public health emergencies such as disaster response and global pandemics, these technologies have an even more pronounced impact on delivery and quality of care due to depletion of resources such as oxygen and shortage of clinical staff [1,2]. For critically ill patients, SASs may be used at the bedside to provide supportive therapy including automating fluid delivery [3-5] mechanical ventilation and oxygen therapy [6–8], vasopressor and/or anesthetic delivery [9-13].

Decision-recommending or automating devices need to be evaluated for their robustness to patient-to-patient variability [1]. Evidence to demonstrate the performance and safety of any medical device, in general, can include a combination of data from bench, animal, clinical, and computational testing [14]. In many cases it may not be possible or could be very cost prohibitive to establish safety of such medical devices with RCTs and animal studies. Evaluation approaches

leveraging physiological models can be used to comprehensively and rigorously stress-test automated critical care devices in a wide range of simulated scenarios using completely *in silico* approaches where computational models of the system including computational patient model (CPM) of the patient and device components are used or hardware-in-the-loop methods where a CPM is used with the actual medical devices (see Figure(1.1)).



(a)



(b)

Figure 1. 1: Computational test setups for SAS. (a) Fully computational testing uses computational models of the therapeutic delivery devices, sensors, and CPM. Initial conditions are set for the SAS algorithm settings as well as the computational models of patient, delivery device, and sensors. Simulated disturbance profiles such as timing of injuries or concomitant therapies are input to challenge the SAS. The testing may be run within a single simulation environment. (b) Hardware-in-the-loop testing uses a CPM, but one or more of the computational models of medical devices are replaced with the physical devices. This requires the use of actuator transfer mechanisms to convert therapy delivery device outputs to digital signals received by the CPM and/or signal generators to convert the output of the CPM to signals that can be recorded by the patient monitor. The highlighted blue region signifies the simulation environment in each scenario.

SSPMs, a subclass of CPMs where model parameters are adapted to each subject to provide individualized predictions, can provide the evidence needed to evaluate the degree to which the device recommendations or therapeutic decisions are robust to patient variability. For example, SSPMs can be used to evaluate the accuracy of a device intended to make decision support recommendation for fluid resuscitation or mechanical ventilation in a wide-ranging patient population where the patient responses are expected to be different. Because SSPMs are tailored to each subject, they have the potential to enable precision medicine particularly when they are utilized in decision support systems applications where the therapy decision recommendations need to be optimized based on subject-specific physiology, as opposed to a population of subjects[15],[16]. Additional applications of SSPMs beyond decision support may include the design and evaluation of PCLC systems[17–20].

SSPMs can perform their predictive purpose in various capacities. They have the potential to augment clinical trials either by reducing the size of the trial or to inform and optimize trial design [14]. They can also be leveraged in pre-clinical safety evidence generation where they can be used to build a cohort of *in silico* subjects to complement and/or replace traditional animal studies [17,

21–23]. They may further enable rapid prototyping and evaluation of new system designs potentially without having to reperform animal studies. This method of pre-clinical safety evidence generation is particularly helpful when significant inter- or intra-subject variability is expected but is difficult to capture in animal studies due to sample size requirements leading to financial or ethical limitations.

The degree to which the simulation results from an SSPM can be relied upon as evidence of SAS's performance will heavily depend on the evidence supporting the credibility of the SSPM to represent the relevant physiology. Therefore, the evidence supporting the credibility of the SSPM is a critical component of the computational testing simulations. The aim of this dissertation is to develop a novel framework for credibility assessment of SSPMs and demonstrate its merit and utility in evaluation of previously developed SSPMs.

1.3 Scope

The scope of this dissertation is limited to simulation-based *evaluation* of SASs using SSPMs that are tuned to data for **prediction of the individual subject's physiology**. A synthetic CPM that is not calibrated to data and uses sampling methods to generate a cohort of models to represent a population of subjects using non calibrated parameters is out of scope of this dissertation. Furthermore, as the objective of this work is to organize credibility activities into a unified and generally applicable framework, development of methods by which each credibility activity is enabled is beyond the scope of this work.

1.4 Contributions

The product of this dissertation is a novel framework for credibility assessment of SSPMs used for evaluation of SASs. First, the proposed framework provides a structured approach for model assessment integrated into the model development workflow, eliminating *ad hoc* steps by *a priori* specification of key model performance criteria, and with maximal efficiency towards evaluation of SSPMs. Second, it is tailored towards SSPMs and their potential weaknesses (e.g., lack of high quality and informative data for model calibration or model overparameterization). The framework is designed to identify such weaknesses as early as possible during SSPM development and evaluation, thereby saving time and resources for the model developers. Third, it enables systematic resolution of such limitations and provides direction for the modeler as to how to rectify model shortcomings. The utility and merit of the framework is demonstrated through examples of physiological models to be used for future evaluation of critical care SASs. The application of the framework to the CO_2 ventilatory gas exchange model and blood volume kinetic model are the first, as far as we are aware, end-to-end applications of a credibility assessment framework on subject-specific physiological models for critical care SASs. In addition to having regulatory science applicability, the proposed framework has utility for physiological model developers and academic researchers as it provides a cohesive and credibility-focused process for assessment of SSPMs during model development. The framework may be applicable at various stages of medical devices lifecycle where in silico evidence can be relied upon to determine the performance of medical devices with SAS technology. This may include early product design stages where in silico evidence may be leveraged to complement pre-clinical and clinical testing. It may also include generating evidence to support marketing applications or to anticipate post market performance of medical devices with SAS technology.

1.5 Dissertation Organization

The introduction, background and motivation for this dissertation is discussed in Chapter 1 . Chapter 2 begins with a literature review on credibility activities conducted on non-SSPMs in other mathematical modeling domains. A comprehensive literature review of SSPMs is then provided in the remainder of Chapter 2 to assess the degree to which such credibility activities are employed towards evaluation of SSPMs. This chapter establishes nascency of such credibility activities as applied to SSPMs and serves as the foundation upon which the proposed framework is built. In Chapter 3^3, we present our credibility framework for assessing the credibility of SSPMs, comprised of a range of credibility activities in a carefully chosen order. Detailed discussions are presented to justify the order of activities. Furthermore, different methods by which each of the activities can be accomplished are presented and discussed. Chapter 4 provides the application of the framework to a physiological model of CO_2 exchange dynamics intended to predict ventilatory parameters such as lung volume and cardiac output in pediatric surgical patients. Chapter 5 provides the application of the framework to a physiological model of blood volume kinetics intended to predict blood volume response to fluid delivery and hemorrhage. We conclude with Chapter 6 which highlights important future research directions for credibility assessment of SSPMs.

Chapter 2: Literature Review

2.1 Introduction[2]

The literature review informing the proposed framework consisted of two stages. First, we conducted a review of the established credibility activities on non-SSPM models and in other industry domains with more mature mathematical modeling practice; see Section 2.2. Such fields included mathematical models of disease epidemiology, systems biology, and environmental systems. The search was focused on the types of mathematical models that are parameterized and adapted to the modelled system through the process of calibration analogous to SSPMs. The distilled information on the types of credibility activities, obtained from the first stage, is used to inform the second stage, where we performed a comprehensive literature search on application of such credibility activities in SSPMs, focusing on credibility activities, their level of rigor, and their method of presentation in published literature; see Section 2.3. The results of this literature review directly impacts the proposed framework in Chapter 3, where we organize the identified credibility activities in a particular order such that overall credibility of the SSPM can be systematically assessed with maximal efficiency.

2.2 Characterization of Credibility Activities in Other Domains

We identified credibility activities used for evaluation of parametrized mathematical models in domains such as epidemiological models [24]–[27], systems biological models [28]–

[2] B. Parvinian *et al.*, "Credibility evidence for computational patient models used in the development of physiological closed-loop controlled devices for critical care medicine," *Front. Physiol.*, 2019

[33], environmental models [34]–[36] and animal models[37], [38]. This search identified 11 credibility activities. Table (2.1) provides a list of the identified activities.

Table 2. 1:Identification of different types of credibility activities. The activities in this table are not arranged in any particular order.

Activity	Description	Example Reference #
Characterization of context of use (COU)	A statement that defines the specific role and scope of the mathematical model used to address a question of interest.	36,37
Sensitivity analysis	The process of determining how a change in a model input (e.g., parameters or initial conditions) affects model outputs.	30,33
Model calibration	The model building process of tuning or optimizing parameters in a mathematical model to minimize the difference between model outputs and real-world data.	31
Face validity	Activity that determines whether the model output is consistent with expectations of domain experts.	30
Identifiability analysis	The process of determining the uniqueness and reliability of parameter estimates from the structure of a mathematical model and experimental data.	28–30
Parameter physical relevance	Activity that determines whether the model parameters make physical sense and are consistent with their physical value expectation.	38
Assessment of fit quality	Quantification of agreement between model output and the experimental data used for model calibration	28
Model formulation	Information on how the equations of a mathematical model are derived, including basic simplifying assumptions.	28–30
Model validation	The process of determining the degree to which a model or a simulation is an accurate representation of the real world.	27
Model parsimony determination	The process of determining the complexity (including the order) of a mathematical model and determining if a simpler mathematical model can be selected (e.g., by fixing a subset of parameters) for the same level of agreement with data.	33, 35
Uncertainty quantification	The process of determining the uncertainty in model inputs (e.g., parameters or initial conditions) and computing the resultant uncertainty in model outputs.	25, 33

Credibility activities above range from those that affect the structure of the mathematical model such as model formulation and parametrization to those that include quality of data to be

used for parameter estimation and calibration of the mathematical model. Comparison of the model output to experimental data not involved in the parameter estimation (i.e., independent data) under various input conditions relevant to the context of use of the mathematical model may serve as the major test of model predictive capability [2]. This comparison is ideally accompanied by some form of uncertainty quantification that captures uncertainty in model inputs then propagates that uncertainty through the model to obtain uncertainty in model output. It is well known [39], [40] that the validation and credibility evaluation process is highly iterative with potential modifications to the order of activities depending on the result of each step.

2.3 SSPM Literature Review

Our next objective was to review the potential roles of SSPMs in SASs evaluation as medical devices, and to identify the credibility evidence and activities that supported the SSPMs for that purpose. The scope of this review included SSPMs that were calibrated mathematical models of individual physiological systems or pharmacokinetic-pharmacodynamic (PK/PD) models whose parameters were tuned to individual subjects. We considered studies with SAS medical devices related to hemodynamic stability (primarily fluid resuscitation or vasoactive drug delivery), anesthetic delivery, and mechanical ventilation published between 1980 and 2019. Studies of animal or clinical testing of a SASs system without any computational testing using an SSPMs were out of scope. Moreover, while mathematical models of the closed-loop system components (e.g., patient monitors, infusion pumps) need to be considered for computational testing of SAS medical devices, these mathematical models were also out of scope. Furthermore, it is important to distinguish between validating the controller or decision-recommending algorithm and validating the SSPM; our review focused solely on the latter.

Based on the credibility activities in Table (2.1), We considered the following regarding the use of the SSPMs:

- the source of the physiological model (whether designed for a specific COU in the article or taken from previous works or combination of both);
- information on the selection of parameter values including if parameters were taken from previous studies, averaged over population, calibrated during the physiological model development process, or combination of these methods;
- whether sensitivity and/or identifiability analyses were performed;
- whether uncertainty quantification was performed;
- comparisons of physiological model performance to experimental data used in physiological model development / training / calibration processes (e.g., assess quality of fit)
- independent validation, namely the comparisons of physiological model performance to experimental data not used in any stage of physiological model development (e.g., to assess predictive performance)
- for articles which use a previously developed physiological models, justification provided to support credibility for the current COU in the article;

2.3.1 SAS Medical Devices Systems for Hemodynamic Stability

2.3.1.1 Use of SSPMs

SAS medical devices have been developed for resuscitation and vasoactive drug delivery systems with the objective of performing fluid resuscitation and/or optimizing hemodynamic stability. These devices monitor and control hemodynamic variables such as mean arterial pressure

(MAP) or cardiac output (CO). Controllers used in hemodynamic stability systems adjust the infusion rate of vasoactive drugs (e.g., sodium nitroprusside, phenylephrine) and/or time of delivery of fluids (e.g., colloids, crystalloids, or blood). A variety of controller designs have been tested with CPMs including single input-single output adaptive and model predictive controllers[41], [42], rule-based learning systems [43], proportional-integral-derivative (PID) controllers [19], [44], and multi input-multi output systems that control multiple drugs simultaneously [45]–[47]. System designs may include supervisory and decision recommending components that add a layer of safety by monitoring for known system limitations, such as noise/signal artifacts in the sensed physiological variables that could adversely impact the controller performance[48].

Multiple studies have used SSPMs to assess controller performance across a broad range of physiological responses. This type of testing can help establish patient populations for which a new control or decision-recommending algorithm may be safe for use by varying parameters within the SSPMs to represent different types of patients and expected responses. Kashihara et al. developed a physiological model to compare the robustness of various control algorithms to the sensitivity of the MAP response to norepinephrine[49]. The evidence from the simulation studies was used to initiate animal studies to further evaluate their controller designs. Bighamian et al. developed 100 different configurations of SSPMs representative of hypovolemia to assess their PID controller for a fluid resuscitation system against a range of autonomic, cardiac, and hemodynamic conditions by varying gain parameters in their physiological model [50]. Similarly, Rinehart et al. assessed the robustness of their fluid resuscitation controller to varying patient weight and cardiac contractility[51]. These types of studies enable an estimate of the potential

distributions in controller performance metrics that may be expected and can be used to identify patient conditions that may result in unsafe performance.

SSPMs are commonly used in the evaluation of closed-loop hemodynamic stability management systems to assess controller performance and robustness to changing patient conditions over time. This allows an assessment of how the controller performs during intra-patient variability simulated by changing the SSPM parameters over time. A physiological model of the MAP response to sodium nitroprusside (SNP) designed to develop and test a controller to maintain MAP by titrating delivery of SNP has been modified and further expanded by multiple groups for this purpose [52]. These modified versions of the physiological model have been used to evaluate controllers capable of tracking and responding to a patient with a changing physiological state[53], [54] and to address non-linear elements in the cardiovascular system [55], [56]. For example, Malagutti et al. considered wide variance ranges in parameters based on information in the clinical literature to enable their controller to respond to potentially unknown patient conditions [54]. Wassar et al. designed a stochastic nonlinear physiological model of MAP response to phenylephrine that considered both inter- and intra-patient variability using data from swine experiments[44]. Additional components were added to model respiratory effects and disturbances such as hemorrhage or the presence of other vasoactive drugs. The physiological model was used to design and investigate a controller under various simulated scenarios before testing the controller in swine experiments. Luspay and Grigoriadis designed a Kalman filter to track time-varying parameters for continuous scheduling of a robust linear parameter-varying controller [57].

A non-linear SSPM with online estimation of model parameters was developed to test the controller during scenarios where patients could be sensitive, nominal, or insensitive to

phenylephrine. This was done by drawing values for each calibrated parameter from pre-defined probability distributions [57].

Held and Roy used an SSPM of the canine cardiovascular system to design and assess a controller intended to adapt to patient SNP sensitivity and confirmed that the controller maintained SNP and dopamine within specified limits [45]. Rao et al. expanded this physiological model to develop a multi-input-multi-output controller. This controller was intended for hemodynamic and anesthetic stabilization [47] by modeling multiple hemodynamic measurements and a depth of anesthesia measure to control five vasoactive and anesthetic drugs. A linearized version of the physiological model was used as part of their model predictive controller and a non-linear version able to simulate multiple pathological conditions was used to evaluate the controller performance as the simulated patient's response changed over time due to the patient's pathological condition. Cote et al. modeled cerebral spinal fluid compartments to design and evaluate the stability and robustness of a closed-loop intracranial pressure regulation system against a range of relevant conditions [58].

Real-time implementations of SSPMs enable more realistic testing of the complete system under expected functional challenges that might be exhibited during normal use. Martin et al. developed a physiological model of pulsatile cardiovascular systems intended to test a PCLC system in real-time when faced with infusion rate limits, rapid physiological changes in MAP, and artifacts in the MAP signal [59]. Woodruff et al. expanded it by combining it with pharmacokinetic-pharmacodynamic (PK/PD) models of multiple vasoactive drugs [60]. Moreover, they modeled the interactions between drugs and baroreceptor components to provide a generalized physiological model of cardiovascular system for the design and evaluation of closed-

loop hemodynamic devices. Silva et al. also enabled real-time testing of embedded controller designs by implementing a physiological model on an embedded system [42] .

An additional use of SSPMs for SAS development is to be used as part of the control algorithm. For example, Urzua et al. demonstrated that using a simplified physiological model of arterial pressure control within the arterial pressure controller proposed in [61] could improve the MAP response time to the therapy [62]. To incorporate the numerous physiological mechanisms involved in arterial pressure control into their system design, Nguyen et al. used a reconstructed version of Guyton's physiological model within their SNP control system [63]. Frei et al. used a model predictive controller to control MAP during surgical stimulation by titrating anesthetic agents [64]. A separate physiological model of the MAP response to stimulation was used to evaluate the model predictive controller.

2.3.1.2 Credibility Evidence

The development of cardiovascular SSPMs has a long robust history. Therefore, many SSPMs used for the development of SASs for hemodynamic stability management are modified versions of previously reported physiological models of cardiovascular system. When existing SSPMs were modified, the impact of the modification on the model performance was rarely assessed or validated for the new COU. The physiological model of MAP response to SNP initially developed by Slate [52] was used and modified by numerous groups to evaluate SAS designs during periods of disturbances and time-variance in patient responses as well as during the handling of parametric uncertainty related to patient responses. Multiple papers [55], [56] modified the Slate's physiological model to assess their controller performance against drug sensitivity by modifying parameters to establish the safety of the controller given unknown patient conditions.

However, additional credibility evidence was rarely provided to support that the model output remained physiologically relevant across the range of modified parameters.

Qualitative assessment of steady-state and dynamic responses was commonly presented to support the use of a physiological model. Rinehart et al. described the simplicity and previous validation to support using a physiological model of Frank-Starling baroreceptor system to assess their closed-loop fluid resuscitation system [51]. The modeled cardiac output response to blood loss was presented to demonstrate that their modifications produced a physiologically plausible response [51]. While acknowledging the limitations of the physiological model used, the authors discussed the suitability for the questions being addressed in the study. Bighamian et al. used a previously developed physiological model to simulate 100 different hypovolemic patient conditions [19]. An assessment of the distributions of hemodynamic variables was made to consider if the hemodynamic responses generated by the different parameter configurations were reasonable. The authors noted that the physiological model did not include key physiological elements such as urine excretion that may impact the fluid volume and thus the results they obtained. Urzua et al. provided a detailed description of the physiological basis to support the physiological model they used within their controller [62]. The qualitative behavior of the physiological model in the presence of disturbances was contrasted against experimental data to demonstrate the physiological relevance and identify behaviors that may not be physiologically plausible.

Woodruff et al. provided a comprehensive description of the development and rationale behind the various components of their cardiovascular simulator that allowed inputs for multiple vasoactive drugs and presented qualitative and quantitative assessments of model performance [60]. A sub-model approach to validation was used where additional components were

subsequently added to the existing version of the physiological model and then model predictions were qualitatively and quantitatively compared to published animal and clinical data. An assessment of the realism of the physiological model was also performed by having anesthesiologists with varying levels of experience control SNP using a simulator [60].

A series of articles built on the physiological model of the canine cardiovascular system developed and evaluated in Yu et al. [65]. Held et al. provided a qualitative assessment of the dynamic and steady-state behavior of the physiological model under specific conditions with a discussion of how this relates to expected physiological changes following their modifications to improve the run-time and agreement with experimental results [45]. Rao et al [47]added a PK/PD model for propofol to use the model for testing simultaneous hemodynamic and anesthetic control systems [47]. The authors noted that they varied circulatory parameters to ensure that MAP and heart rate responses matched experimental observations, and propofol model parameters were then tuned so that the overall physiological model response matched steady-state results. The results of the calibration procedure were presented along with simulation results to assess the model-predicted hemodynamic responses due to increasing propofol infusion[47].

Kashirara et al. developed a data-driven physiological model of the MAP response to norepinephrine from experimental data tracking MAP during infusion of norepinephrine [49]. Data from three experiments were used to fit the physiological model, and a comparison of the experimental and simulated responses from a single animal was presented to support the model performance. Nguyen et al. also used data from animal experiments to tune parameters in a PK/PD model of SNP that was connected with a larger physiological model of cardiovascular system [63]. Qualitative assessment was presented that compared the blood pressure response to SNP on the calibration data used to tune the PK/PD model of SNP.

Wassar et al. used data from swine experiments to develop a stochastic physiological model of MAP response to phenylephrine that enabled varying the sensitivity of MAP to the infusion rate of phenylephrine [44]. An initial portion of experimental data was used to calibrate the parameters in the physiological model, and then those parameters were applied to the physiological model and predicted the remaining portion of the experimental data. Results were presented from at least one animal used in development of the physiological model that showed the measured and predicted responses. The authors noted that similar results occurred for the other animals. The physiological model was determined to be appropriate by the authors because it included saturation effects and allowed for inter- and intra-subject variability to be assessed.

Frei et al. evaluated the trade-off between model performance and model order when considering the appropriate physiological model to use for testing the response of their MAP controller to anesthesia and disturbances [64]. In the manuscript, they noted the challenges associated with the use of high order physiological models and presented evidence of their model performance on a single case of experimental data.

2.3.2 Closed-Loop Systems for Mechanical Ventilation

2.3.2.1 Use of SSPMs

The controllers employed in closed-loop mechanical ventilation devices adjust ventilation parameters to maintain blood gas levels or other respiratory-related variables within a physiological range of interest. Physiological modeling of the cardiorespiratory system plays a key role in design [66]–[68] and evaluation [69] of closed-loop mechanical ventilation algorithms. The majority of efforts behind the closed-loop design and evaluation of a mechanical ventilation device is to continuously adjust blood oxygenation levels (i.e., SpO_2 [70]–[73] and SaO_2[67], [74]),

carbon dioxide levels (End-tidal CO_2 fraction [75]–[77]) or respiratory variables including respiration rate and tidal volume [78]. Adjustments are made by varying the level of therapeutic settings on the device including the fraction of inspired oxygen (FiO_2) [67], [69], tidal volume [75], and positive end-expiratory pressure [78], [79]. A wide range of controller types have been applied including model predictive controllers [67], [80], PID and fuzzy controller [69], robust controllers [81], [82], and rule-based expert systems [80].

Sano et al. and Yu et al. both developed transfer functions to relate blood oxygen levels to FiO_2 by linearizing a physiological model of gas transport in order to simplify the controller design process [82]. These transfer functions were used to design controllers to adjust partial pressure of oxygen in arterial blood (PaO_2) by titrating FiO_2. Sano et al. further developed a linearized physiological model relating rate of breathing to blood CO_2 levels [82].

Many SSPMs used in the design and evaluation of PCLC mechanical ventilation devices were developed from foundational studies in modeling of cardiorespiratory system such as [83], [84] and then with improvements from [7], [85]. Fincham improved on past physiological models of respiratory systems by accounting for physiological processes such as respiratory work output and internal body feedback systems. This physiological model formed the basis for design and evaluation of closed-loop oxygenation devices in adults [73]. The performance of the control system was evaluated in simulation experiments representing different physiological conditions including low respiratory compliance and hypoventilation. The physiological model presented in [83] involved mass balance dynamics of gas transport and exchange, metabolism dynamics, chemoreceptors, and endogenous respiratory control mechanisms. The original physiological model was modified by adding parameters of the neonatal respiratory system such as lung shunts to enable design of controllers intended for neonatal oxygenation control [85]. Morozoff et al.

developed a physiological model that consisted of elements representing the cardiovascular system and the respiratory system including the effect of shunts [7], [86]. They first developed a physiological model for adult cardiorespiratory system and then used an allomeric approach and proportional scaling to derive parameters relevant to neonatal applications. They evaluated fuzzy logic and PID controllers by using this physiological model to simulate various neonatal conditions, including changing oxygen affinity, desaturation pulse duration, patient motion conditions, and combinations of these.

Data-driven physiological models of $EtCO_2$ have been presented for controller development [76], [77]. Kim et al. described lowering the order of multi-compartmental physiological models to capture CO_2 transport dynamics with and without CO_2 transport delay parameters that could be used to design and test controllers [77]. Hahn et al. developed an empirical, low-order physiological model relating $EtCO_2$ to minute ventilation, where clinical data were used to identify the parameters therein [76].

Pomrpara et al. evaluated a Policy Iteration Algorithm (PIA) controller based on a physiological model of blood oxygenation [67]. The physiological model was used to evaluate the performance of various iterations of the PIA controller in terms of settling time. The parameters of the physiological model were identified by changing FiO_2 settings and inducing acute respiratory distress syndrome in a large animal experiment. For further implementation of control system design, they approximated the non-linear physiological model via linearization. Iobbi et al. reported a physiological model of oxygen transport to evaluate a controller intended to adjust FiO_2 based on pulse oximetry feedback [71]. They assumed second-order transfer functions with delay parameters to relate blood oxygen saturation levels to inspired oxygen. The physiological model

was used to simulate hypoxic events from artificial disturbances representing variable fluctuations in patient pathophysiology.

2.3.2.2 Credibility Evidence

Yu et al. provided model formulation and descriptive information of the physiological process behind blood oxygenation, relying on their previous work to justify the structure of a first-order transfer function relating PaO_2 to FiO_2 [74]. They selected ranges for model parameters but did not provide any information about the validation of these ranges or the sources they were derived from. Similarly, Sano et al. designed a physiological model of both O_2 and CO_2 gas transport for robust controller design. Their rationale for constructing a new physiological model was that the previously developed physiological models of cardiorespiratory systems were too complex for control system design[82].

The foundational studies presented in [83], [84] generally provided great details of model formulation and structure. The physiological modeling approach and rationale for inclusion of individual components, for example dynamics and chemical control of gas levels in blood, were provided. In [83], model evaluation involved comparing the model performance at various conditions including hypoxic, hypercapnic, and exercise conditions. Direct comparison with experimental results reported in literature was performed only for resting and hypercapnic conditions. Specifically, the comparison included numerical values of selected model outputs (e.g., arterio-venous gas difference) and not of the entire range of variables that would be utilized in future studies involving design and evaluation of closed-loop mechanical ventilation devices such as [73].

Morozoff et al. described a physiological model with simplifying assumptions such as no account for the removal of oxygen as CO_2 during expiration in the mass balance of O_2 [86]. A

more detailed parametrized compartmental model of the respiratory and cardiovascular systems was described in [7]. The physiological model of the cardiovascular system was compared with another physiological model presented in [84], although the level of agreement and specific conditions were not specified. The study provided a clear list of assumptions for the physiological model of the cardiovascular system, such as uni-directionality, non-pulsatile blood flow, and perfect mixing of blood. The authors used proportional scaling and parameter values of mammals similar in weight and surface area to neonates to select model parameters specific to neonatal patients' circulatory and respiratory systems. The authors state that the physiological model was validated in stages first using Rideout data [84] for the adult cardiovascular system and then using published data for the neonatal cardiovascular system. For neonates, the physiological model was compared qualitatively with physiology textbooks.

Martinoni et al. [66] provided a simplified physiological model of the one presented by Chiari et al.[87] where they assumed constant cardiac output and constant oxygen saturation in both arterial and venous blood. The physiological model was "considered sufficiently descriptive for closed-loop control purposes"[66].

Iobbi et al. [71] used a physiological model of the oxygen dissociation curve proposed by Severinghaus [88]. They assumed that the oxygen response is a function of three factors: a baseline oxygen partial pressure, a driving partial pressure which was modeled as a sinusoidal function, and a flow-dependent term which was modeled as a second order differential equation whose parameters were selected from literature. The authors acknowledged that the parameters selected are based on average values and do not represent patient variability.

Pomprapa et al. described the input-output relationship of minute ventilation and $EtCO_2$ and compared a first-order linear physiological model and a nonlinear physiological model[81].

The authors used root-mean-square-error (RMSE) between an "estimated dataset" and a "validated dataset" to compare the results obtained from the two physiological models. They also provided qualitative comparison of linear and nonlinear physiological models. Model parameter values were calibrated by collecting experimental data from a single healthy human subject with homogenous lungs and normal body mass index. The authors suggested that the physiological model was validated using the single human subject data by mentioning comparison of the estimated physiological model "with a validated dataset", although it is unclear from the manuscript how this dataset was developed and whether the validation dataset involved measurement of model output independent of model calibration.

Hahn et al. developed an empirical, data-driven physiological model with parameters identified from data [76]. The authors demonstrated the similarity between their affine physiological model relating minute ventilation to $EtCO_2$ concentration to pharmacological modeling. They used this analogy to provide intuitive interpretation of the physiological model they developed by regarding the two states as the CO_2 concentrations in the body and the lungs. Respiratory data of 18 pediatric patients from an anonymized repository were used to calibrate the physiological model. The goodness of fit was quantitatively compared with experimental dataset using RMSE and coefficient of determination. Kim et al. identified model parameters from clinical data [77]. They demonstrated that inclusion of transport delay resulted in model prediction improvement compared with the same physiological model without transport delay. Unlike the physiological model with transport delay, the parameter estimation on the physiological model without transport delay resulted in parameter values that are not physiologically plausible.

2.3.3 SAS for Anesthetic Delivery

2.3.3.1 Use of SSPMs

SSPMs have been widely used to design and evaluate closed-loop anesthetic delivery systems. The systems control level of consciousness [11], [89]–[92], analgesia [93], [94], neuromuscular blockade [95], [96], or combinations of these goals [97]–[99]. Additionally, a series of studies designed closed-loop propofol delivery with the specific aim of inducing and maintaining pharmacological burst suppression based on processed EEG signal [100], [101]. PCLC devices intended for control of hypnosis delivered propofol and used feedback variables such as bispectral index (BIS) [102], [103], WAV_{CNS} [76], [92], or auditory evoked potential response [91] to titrate anesthetic agents including propofol and isoflurane. Some studies combined the closed-loop delivery of hypnosis with analgesia leading to closed-loop co-administration of propofol and remifentanil [93], [99]. Neuromuscular blockade agents such as rocuronium and atracurium were used and controlled using feedback variables based on muscle movement [10], [95], and in some studies, were combined with closed-loop hypnosis delivery[97], [98], [104]. Gentilini et al. used hemodynamic variables such as MAP and plasma concentration to titrate drug delivery [94].

Most CPMs used for evaluation of closed-loop anesthesia delivery systems have been derived using existing PK/PD models for the anesthetic drugs of interest. Fang et al. constructed a library of widely used virtual patient PK/PD models after investigating and collecting a number of reported PK/PD models for propofol, isoflurane, remifentanil, and atracurium administration [97]. They combined the PK/PD model database with a control algorithm database including PID control and model predictive control while facilitating four control modes (manual, automatic, switching from manual to automatic, and switching from automatic to manual). This created an

anesthesia simulation platform imitating the clinical situation. A patient response PK/PD model of hypnotic drug isoflurane was adopted in [105] to generate, by varying the PK/PD parameters, 16 representative patients selected to cover the range of observed sensitivity from 972 patients. These 16 virtual patients were then used to evaluate and compare performance across six different controllers designed for hypnosis regulation, using BIS as the controlled variable, including robustness and stability for expected surgical disturbances. Liu et al. modelled patient perceived pain using a PK model of fentanyl, a triexponential weight function, and a steady-state biochemical kinetics model. The intent was to represent the relationship between brain tissue drug level and unbound opioid receptor ratio in the central nervous system (CNS). They also used a relaxation pulse frequency model to convert pain to button pressing frequency [106]. This PK/PD model was then used to assess performance of four control algorithms in simulated patients with minimum, average, and maximum pain sensitivity. Alongside real-world data, Mendez et al. [107] used a simulation to assess a fuzzy logic controller to regulate BIS using propofol. Simulation of patient response was conducted using a PK/PD model after choosing parameters to match real data from patients undergoing similar surgery procedures and with the same drugs as those used in the control experiments. Struys et al. [108] developed a simulation methodology to compare performance between multiple controllers to control BIS using propofol. Virtual patients were generated using a model which incorporated PK/PD modeling along with effect relations modeling.

Nogueira et al. [109] studied the performance of the controller proposed in [110] to regulate BIS using propofol and remifentanil using a CPM under the presence of model parameter uncertainties, and introduced a retuning strategy for the CPM to recalibrate itself. Six simulated patients were developed by setting parameters of the CPM based on the data of real patients subjected to general anesthesia under propofol and remifentanil manually controlled by an

anesthetist. Dumont et al. [92] used PK/PD model results of a number of patients from different groups to choose controller parameters to design robust PID controllers and robust controllers based on fractional calculus to regulate hypnotic state of anesthesia using intravenous administration of propofol. [94] used a PK/PD model as a part of their controller to control MAP by controlling analgesics. Puebla et al. used a PK/PD model to design a cascade controller to control end-tidal anesthetic agent concentration in response to BIS sensor [111].

Mahfouf et al. [112] published a series of studies using a multivariable CPM of anesthesia that combined various CPMs obtained from classical PK/PD models, published literature, and the authors' own system identification experiments [104]. These experiments involved collecting drug infusion rates (inputs) and MAP (outputs), although it is unclear if this was done on measured or simulated data. They used the CPMs to synthesize a model predictive controller that would enable the next control action to be determined based on the output of the CPMs. Mendez et al. developed a fuzzy model predictive controller based on fuzzy modeling of the patient PK/PD response[107].

Westover et al. [101] used Schnider's PK/PD models of propofol [113] to construct a robust PID controller for maintenance of burst suppression. The authors used a system identification procedure to determine subject-specific PD parameters in a rodent study. PK variability was determined using published estimates of the coefficient of variation in PK parameters. The approach was similar to a series of studies in [11], [89], [90], [92]on closed-loop control of hypnosis using propofol that leveraged classical PK models of [114] and classical PD modeling using the Hill equation. The authors used such PK/PD models to build controllers robust to parametric variability and disturbances. Jin et al. [99] continued the work done by Hahn et al. [90] and designed a coordination controller that recursively adjusts the reference targets based on the

estimated dose-response relationship of a patient using a classical PK/PD model of propofol and remifentanil and the interactions between the two.

Silva et al. introduced a Wiener model consisting of linear dynamics and a static nonlinearity to characterize the response (Train-of-four-TOF) to neuromuscular blockade drugs[10], [98]. The studies focused on developing a minimal PK/PD model for the purpose of controller design. Almeida et al. [95]used Silva's PK/PD model presented in [10] to design an adaptive controller to titrate rocuronium to control muscle movement, and Nogueira et al. used it to design a nonlinear controller for controlling the depth of hypnosis by titrating propofol [109].

Kharisov designed a PK/PD model for the purpose of designing a controller for automatic delivery of isoflurane based on BIS [102], [115]. The PK/PD model was developed using clinical trial data collected in earlier studies [116], [117]. In [114], a black-box PK/PD model was developed for the purpose of controller design, whereas in [101] the same clinical data were used to estimate parameters of a gray-box linear time-invariant PK/PD model structure for controller design. Subspace identification methods were used to construct the PK/PD models for six patients pertaining to the black box PK/PD model, while for the gray-box PK/PD model, explicit parameters representing mean input and output biases and an adaptive gain parameter capturing the time-varying nature of patient tolerance to the anesthetic agents were formulated.

2.3.3.2 Credibility Evidence

Numerous studies relied on established PK/PD relationships previously reported for the drugs of interest. Gentilini et al. adopted a PK model with parameters from literature and approximated the PD model with parameters based on anesthesiologists knowledge and experience [94]. A linear relationship between MAP and the effect site concentration was assumed, and the gain for the PD relationship was determined based on expert opinion. Qualitative and descriptive

information of the resulting PK/PD model was provided. Linkens et al. leveraged PK/PD models from published literature combined with physiological models of blood pressure [104]. Parameters for the Atracurium PK/PD model were taken from literature and measurements in the operating theater. The hypnotic effect of isoflurane was derived using published clinical data [118]. Descriptive information for structure and formulation of submodels (e.g., the isoflurane to muscle relaxation interaction model) was limited. Mahfouf et al. used a CPM structure pertaining to circulatory and inhalational anesthesia in a previously reported literature and augmented it with a blood pressure delay parameter from their own previous work [112]. They also combined this CPM with a CPM of anesthetic phase change. Unknown CPM parameters were identified using 400 points from isoflurane delivery rate (input) and MAP (output) data with a fit quantitatively evaluated using RMSE.

Liu and Northrop provided detailed description of the CPM and source of calibration data [106]. CPM parameters were fit so that the CPM output followed the actual patient responses reported in previous work. Struys used the Schnider PK model and the PD model developed previously in their clinical work [94], [108]. Ten virtual patients were selected to represent different pharmacodynamic profiles defined by the effect site concentration-effect relation combined with a delay parameter. The authors relied on their previous work and published articles of Schnider to create the virtual data [94]. Similarly, Ionescu et al. [93] designed a model predictive controller using established PK/PD models from Schnider and Minto[119] along with their clinical experience to obtain model parameters.

Merigo et al. [103] used the PK model with parameters from [94] held at nominal values. They derived the PD parameters for the standard Hill equation. Elkfafi et al. provided qualitative

description of how the CPM was designed to show the change of depth of anesthesia with respect to propofol infusion rate [91].

A number of studies [89], [90], [92], [101] have reported the design and evaluation of robust controllers by considering the variability associated with patient parameters for PK/PD models reported in literature. The structure and form of the compartmental PK and effect site PD models used in these studies were reported. The range of PK model parameters were obtained from published evidence while the PD model parameter ranges were derived using clinical data in [88], [89], [91] and animal data in [100]. Credibility evidence included references to previous studies.

Unlike previous studies that adopted classical PK/PD models, Mendez et al. developed a CPM of propofol impact on BIS using the fuzzy modeling approach [107]. Information on the form of the fuzzy model and rules was provided. To develop this CPM, experimental data from a clinical study that captured input (propofol) and output (BIS) measurements was divided into two training (15 patients) and validation (10 patients) sets. The fit of the fuzzy model was reported as compared with compartmental model for one patient, but it is unclear if this case represents a calibration scenario or is from a separate independent validation data set. Mendez et al. acknowledged that, because of actual population variability and presence of disturbances during surgery that may not have been captured in their CPM, the controller performance in operating room could vary with respect to the simulated results [107].

Silva et al. provided studies of system identification of Weiner models pertaining to NMB and hypnosis [10], [98] to be used for closed-loop system design. The main emphasis of these studies was to develop parsimonious CPMs by lowering the CPM order to make it amenable to controller design. Such CPMs were subsequently adopted in [95], [109], [120] for the purpose of controller design. Original modeling studies by Silva mentioned the validity of the previous CPM

[121] and noted that the current CPM needed to be modified for the current COU. They reduced the CPM order by lumping the parameter values into a single parameter.

Kharisov et al. provided a description of their CPM and its formulation [115]. The CPM parameters were estimated by calibrating the CPM to previously collected clinical data. Portions of the patient's data were selected during isoflurane infusion to derive model parameters. The CPM was evaluated using the same clinical trial data, and the output of the CPM was reported for one patient. The CPM did not appear to follow the experimental data beyond the portions in which calibration was performed. It appears that the segmentation and selection of intervals for parameter estimation was based on the COU of the CPM which was to be used for BIS levels below 70. The predictive capability of the CPM was not evaluated on independent data that did not take part in parameter estimation. In [102], the authors provided a description of the CPM and formulation. The fit of the calibrated CPM was quantitatively evaluated on the same clinical data and the authors determined that the fit achieved was "acceptable" for their COU.

2.4 Discussion of SSPM Literature Review

Few of the studies included in this review have rigorously tackled the fundamental question of what makes an SSPM sufficiently credible for a specific evaluation of SASs systems. With regards to evidence needed to assess credibility, we found a broad spectrum in the type and rigor of credibility evidence and activities presented for SSPMs, but there was no clear link between how the SSPMs were used and the rigor and extent of the credibility evidence as a function of SSPMs context of use. The evidence ranged from mostly qualitative and descriptive information and in some cases quantitative assessment of the SSPM performance, although it is not always clear if this was performed on independent data or calibration data. The authors usually did not

report on how the decision was made about the type and rigor of the credibility evidence and activities needed to support SSPMs, and the details of such evidence were typically limited. The degree to which quantitative credibility-related requirements such as predictive accuracy was considered is unclear for various applications in hemodynamic stability management, mechanical ventilation, and anesthesia articles reviewed.

While most of the studies presented some qualitative assessment of SSPM outputs, the application of more formal assessment techniques that provide systematic and quantitative information about the SSPM performance, such as sensitivity analyses, identifiability analysis, uncertainty quantification, and comparisons to independent experimental data to assess the predictive capability of the SSPM was limited. Even when provided, the credibility activities were not cohesive part of the process and often scattered amongst various studies. Table (2.2) lists scattered credibility activities conducted for SSPMs focusing on mechanical ventilation. Similar observations were made for SSPMs focusing on fluid resuscitation and anesthesia delivery applications.

Table 2. 2: Summary of credibility activities for mechanical ventilation SSPMs

Credibility Activities									
Reference	COU	Sensitivity Analysis	Face Validity	Identifiability Analysis	Calibration Quality Evaluation	Model formulation	Model Validation	Model order reduction	Uncertainty Quantification
Yu et al. (1987)						X		X	
Sano et al. (1988)						X		X	
Fincham and Tehrani (1983)						X			
Morozoff and Saif (2008)			X			X			
Morozoff et al. (1993)			X			X			
Tehrani (2012).			X			X			

Martinoni et al. (2004)	X					X			
Iobbi et al. (2007)						X			
Pomprapa et al. (2013)						X		X	
Hahn et al. (2012)	X		X		X	X			
Kim et al. (2016)		X	X	X	X	X			X

A significant attribute affecting the credibility of an SSPM used for evaluation of SASs is the ability of the SSPM to capture inter- and intra-patient variability as well as variability in disturbance scenarios experienced by critically ill patients. Adequate characterization and quantification of uncertainty suited to an application can therefore result in much more robust and meaningful SASs evaluation methods. Recently, some of these challenges related to uncertainty in patient and system response to disturbances have formed the impetus for synthesizing PCLC controllers based on modern control synthesis methods based on robust control theory [11], [89], [92]. While this type of controller presents a powerful tool with the potential to guarantee that control specifications such as stability and robustness are met, the SSPMs used in such methods need to produce clinically relevant and widely variable disturbance scenarios. Furthermore, uncertainty due to oversimplifying assumptions embedded in SSPMs need to be considered. For robust control applications [101], overall uncertainty in SSPM and its bounds must be quantified and validated for controller evaluation. In medicine, such bounds will be a function of the patient population, patient characteristics, type of procedure, amount of drug infused, drug co-administration, and concomitant therapies.

Another method of handling inter- and intra-patient variability has been to design model predictive controllers [107], [112]. In this context, the SSPM used for design of controllers is not static (as is the case with robust control). The SSPM directly affects the outcome of the controller

as it is part of it. For example, in model predictive control of anesthesia, a forecast of the output is communicated to the controller based on an SSPM [112]. As such, the credibility evidence needed to establish validation and uncertainty quantification of such SSPM for this COU may require greater rigor due to the consequence of SSPM outputs on the control function.

2.5 Previous Work on SSPM Validation frameworks

There are a number of contributors who helped lay the foundation for validation of SSPMs. Cobelli et al. introduced a validation process for simple and complex SSPMs [122]. Simple SSPMs were regarded as those with plausibility of performing classical system identification procedures whereas complex SSPMs were those with numerous parameters that made system identification and parameter estimation not feasible. Their work outlines fundamental concepts to address credibility of SSPMs. They introduced concepts of internal consistency, algorithmic validity, and external validity such as empirical, theoretical, heuristics, and pragmatic validity, all pertaining to SSPMs. Batzel et al. provided steps for validation of SSPMs including classical and generalized sensitivity analysis, subset selection, evaluation of local characteristic of parameter estimation process, experimental design and data quality evaluations, and global analysis [40]. Summers et al. also recognized that whenever possible, outputs of an SSPM must be compared with experimental outcomes for more consequential conclusions about the validity of the SSPM. They summarized three criteria for supporting the validation of large-scale integrated SSPMs, including a qualitative assessment of changes in the SSPM output and a quantitative assessment of the steady-state and dynamic response of the SSPM, while noting that these activities should be performed considering the clinical context of use of the SSPM[123], [124]. Some recent articles proposed systematic assessment methods relevant to SSPMs and two reports in particular discuss

the importance of identifying and, when possible, quantifying uncertainty [125], [126]. While these frameworks present different methods for performing validation and assessing SSPM performance, they do not present a method for determining a context-dependent level of rigor or the extent of the evidence needed to determine that a given SSPM is sufficiently credible for a prespecified COU. Furthermore, they do not provide a unified process in which all credibility activities (e.g., identifiability analysis, UQ and model predictive capability assessment) are integrated in a systematic way with the goal of assessing the credibility of an SSPM for a specific COU. Finally, the applicability of such workflows have not been shown in other works related to SSPMs with applications towards critical care (e.g., mechanical ventilation and anesthesia delivery) SAS evaluation.

2.6 Chapter Conclusions

In this Chapter we have reviewed model credibility evidence provided for a range of SSPMs. We observed that most evidence are of qualitative or descriptive nature, and credibility assessment of such SSPMs appears to follow no particular process with *ad hoc* steps added, omitted, or modified to facilitate the physiological modeling process. Quantitative credibility activities are scattered and are not presented in a cohesive fashion throughout SSPM literature. For example, identifiability analysis as a foundational model credibility assessment activity is well established in other fields (e.g., environmental science, cancer biology, and system biology) but is underutilized in SSPM and its application is nascent for SSPMs to be used for critical care SAS evaluation. Current evidence for demonstrating that a SSPM is credible in such medical devices varies considerably, and rationales linking the evidence supporting a SSPM for a specific COU are limited in scope and presentation. Conducting the credibility activities, by following a

framework unifying each of the credibility activities in a pre-specified order and workflow that is amenable to unique attributes of SSPMs will aid in determination of credibility of SSPM to be used for evaluation of SASs. Therefore, in Chapter 3 we present a comprehensive framework which links various relevant credibility activities in a specific order for the purpose of evaluating the validity of SSPMs.

Chapter 3: A Framework for Assessing Credibility of SSPMs

3.1 Introduction

As discussed in the previous chapter, the literature review of credibility activities in other domains as well as those related to SSPMs concludes that a framework for credibility assessment of SSPMs is urgently needed with the increased emergence of SASs. In this chapter we introduce a novel framework specifically geared towards SSPMs which enables a comprehensive, step by step, and disciplined credibility assessment of SSPMs with optimal efficiency while being integrated in the model development workflow.

The majority of the existing credibility assessment processes, if they were to be organized as part of a workflow, may be depicted in Figure (3.1). In order to best inform development of an improved framework, a gap analysis informed by results presented in Chapter 2 was conducted. This analysis identified in the existing workflow used for assessment of SSPMs:

a) The model context of use (COU) is often not explicitly specified *a priori* to inform the rigor of credibility activities.
b) SSPM credibility assessment activities are often done in an *ad hoc* way with no prespecified order.

c) Other domains utilize foundational credibility activities such as evaluation of model structure to ensure for unique parameter identification. These are seldom conducted for SSPMs.

d) In the existing SSPM workflows, the emphasis is on calibration results as the primary factor determining the credibility of the SSPM. Frequently a single check is used to determine whether the calibrated parameters are credible and that is to assess the physiological relevance of the parameters without assessment of model predictive capability on independent data.

e) While some studies provide more rigorous quantitative measures post calibration for identifiability of the parameters such as evaluation of parameter confidence interval and their uniqueness, such credibility activities are often not conducted in such an order to inform the next steps towards effectively curing any unidentifiability problems.

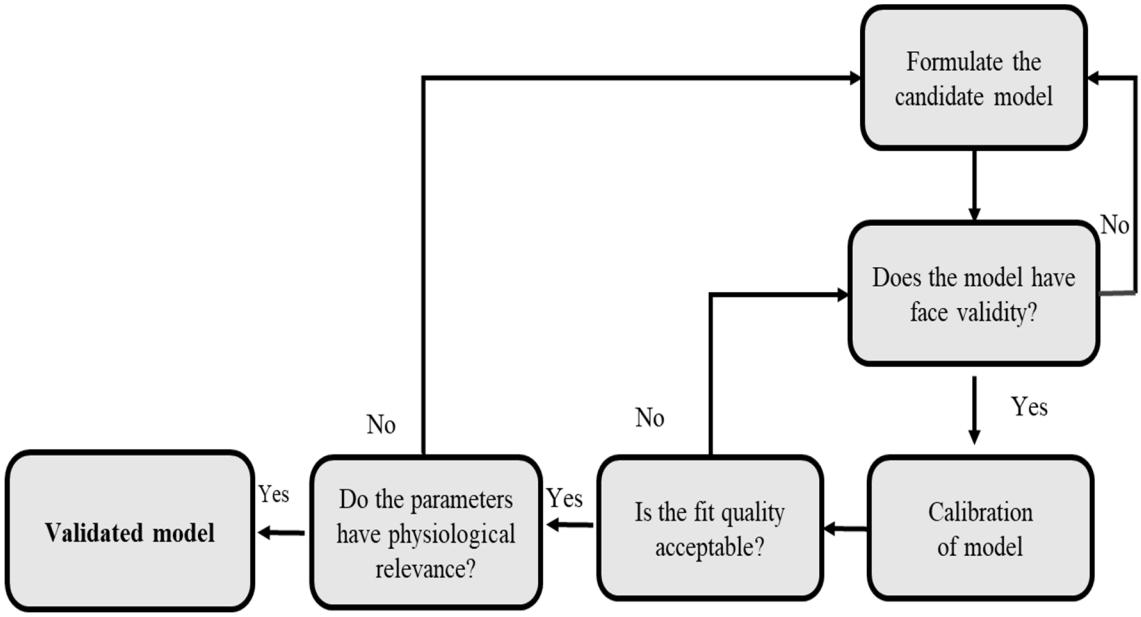


Figure 3. 1: The existing general workflow for validation of SSPMs

In this Chapter we present a novel unified credibility assessment framework for SSPMs that addresses the gaps stated. The framework has been designed to satisfy the following requirements:

1) The framework needs to be informed by a comprehensive literature review to characterize and identify the credibility activities in other domains and their current state of application in SSPMs.
2) In order to save time and resources, the framework needs to avoid *ad hoc* steps and prescribe a firm order for credibility activities that would unambiguously and expeditiously lead the evaluator either to a conclusion for the credibility of the SSPM or towards addressing the root cause of the SSPM's lack of credibility as early as possible.
3) The framework must account for unique challenges of evaluating SSPMs and focus on credibility activities addressing those challenges. For example, data used for identification of parameters in SSPMs are often noisy and sparse. Therefore, one would expect that a suitable framework for evaluation of SSPMs would likely place strong emphasis on credibility activities addressing reliability of parameter estimates considering low quality data.
4) The framework needs to be broadly applicable towards evaluation of different types of SSPMs.

The framework we present is the first ever framework that provides a systematic and maximally efficient pipeline for SSPM credibility assessment fulfilling the above requirements. Our framework is integrated into the model development process, rather than applied after a model has been fully developed. This is because an integrated model development/model assessment framework is maximally efficient in terms of avoiding wasted effort, by identifying any weaknesses with the model or data at the earliest opportunity. The framework uses all the relevant

credibility activities identified as part of review of model assessment in other domains and the review of SSPMs credibility assessment provided in the literature. The framework provides a clear and well-defined pathway for developing an SSPMs with concurrent credibility assessment, avoiding *ad hoc* steps. We have carefully devised the order of the individual credibility activities in such a way that is tailored toward evaluation of SSPMs, and if the model fails any stage, the root cause can be quicky identified and there will have been minimal wasted effort on the part of the model development team.

The remainder of the chapter is organized as follows: First, we provide in Section 3.2 a general discussion of three related concepts that are key to the framework – structural identifiability analysis, sensitivity analysis and practical identifiability analysis. Next, a high-level overview of the framework is provided in Section 3.3. Then, in Section 3.4, we provide full details for each step of the framework, including the precise goal of each step, possible methods that can be used for achieving the goal, detailed rationale for the order of the step, and options if the model fails the step.

3.2 Introduction to Model Identifiability Concepts

Since the purpose of most SSPMs is to predict observed or unobserved physiological states, outputs, or parameters, evaluating the soundness of the structure of the SSPM (e.g., symmetries in model equations) that may affect parameter estimation or model characterization error or evaluating the quality of the data to be used for parameter estimation are of paramount importance. Flaws in either SSPM structure or data used for estimation of parameters could result in poor model predictions. Figure (3.2) describes this concept which is known as the “garbage paradigm” [39]. A common flaw in the structure of SSPM arises from a model formulation that renders the resulting

estimated parameter non-unique due to symmetries in model equations [127]. Similarly poor information content in the data used for parameter estimation may result in unreliable parameter estimation [28]. Therefore, evaluation of model structure and quality of data for the purpose of parameter estimation are highly relevant activities towards credibility assessment of SSPMs.

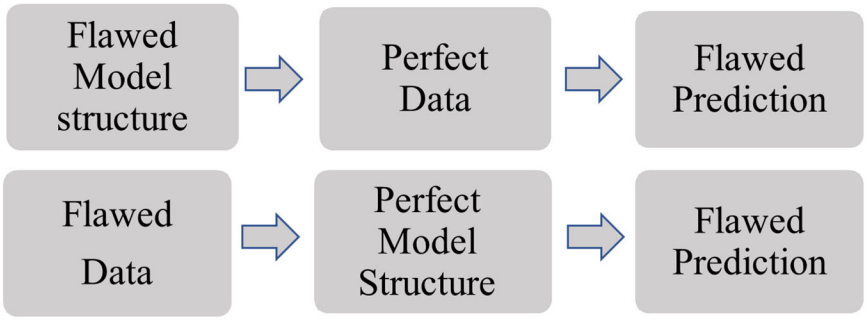


Figure 3. 2: Two extreme scenarios that may arise during development and system identification of SSPMs [39].

The credibility activity central to assessing uniqueness and reliability of SSPM parameters is known as identifiability analysis [28], [128], [129]. It is a control-theoretic concept closely related to the observability of SSPMs [127], [130]. It answers the question of whether the calibrated SSPM parameters (i.e., parameters that result in a local or global cost function minima) are unique and reliable. Identifiability analysis can be broadly categorized as Structural (or *a priori*) and Practical (or *a posteriori*) Identifiability Analysis, hereby termed SIA and PIA respectively. SIA answers the question of whether it is possible to uniquely estimate the parameters of the SSPM under a theoretical assumption that the data for the SSPM are perfect (i.e., noise free and informative) for estimation of SSPM parameters. It is strictly a mathematical concept and independent of data quality. PIA, on another hand, is associated with the quality of the data used for calibration and answers the question of whether the parameters of the SSPM can be reliably estimated given the

SSPM's structure and the available quantity and quality of data [28], [129]. Figure (3.3) (a), (b) and (c) depicts different scenarios of parametric and model identifiability.

An intermediate identifiability-related analysis which may be conducted after SIA and before PIA known as Pre-calibration Sensitivity Analysis (PSA) may be conducted in order to evaluate the possibility of model order reduction. This in turn could facilitate SSPM calibration and inform PIA. A brief introduction to SIA, PSA, and PIA is provided herein.

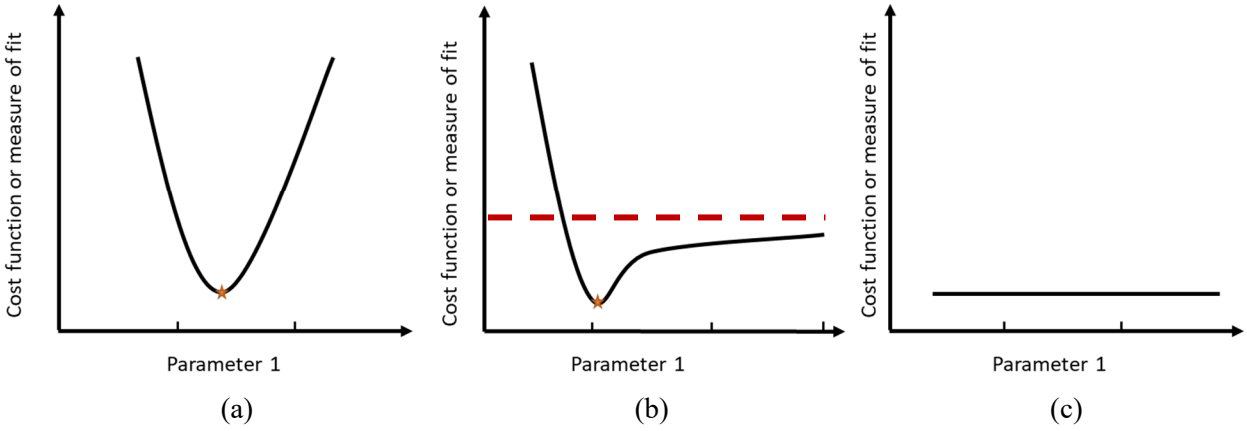


Figure 3. 3 : Example cost function curvature under different identifiability scenarios. (a) Structurally and practically identifiable. (b) Structurally identifiable but practically unidentifiable due to unbounded confidence limits (red dashed line). (c) Structurally and practically unidentifiable.

3.2.1 Structural Identifiability Analysis (SIA)

The general problem formulation for determination of global structural identifiability (GSI) or local structural identifiability (LSI) properties of a model ($\mathcal{M}$) and its set of parameters θ can be described as follows. A model is structurally identifiable if:

$$\mathcal{M}(\theta) = \mathcal{M}(\tilde{\theta}) \Rightarrow \theta = \tilde{\theta} \tag{3.1}$$

where $\tilde{\theta}$ is the set of parameters for the model $\mathcal{M}$ and θ is the set of parameters of the modeled process [129]. Eq. (3.1) essentially implies that if model structure $\mathcal{M}$ captures the process without

any error (i.e., $\mathcal{M}$ is a perfect representation of the system), then for $\mathcal{M}$ to be structurally identifiable, θ needs to be unique. In the case of GSI, the parameters are proven to be unique globally, while for LSI, there is a neighborhood within which the uniqueness of the parameters can be guaranteed. The goal of SIA is to investigate whether Eq. (3.1) holds under the theoretical assumption that the data are noise-free and with maximal information (i.e., model input and measurement times can be chosen at will) [129].

An example of a model that is trivially not structurally identifiable is $y_i = abt_i$, where y_i is the model output at time t_i and a and b are parameters to be fit. Parameters a and b cannot be uniquely determined, regardless of the quality of the data available for fitting. In practice, proving that a complex model is structural identifiable can be challenging. Methods for performing structural identifiability analysis are discussed in Section 3.4.3.

3.2.2 Pre-calibration Sensitivity Analysis (PSA)

Parameter sensitivity analysis is a credibility activity closely related to parameter identifiability. In this type of analysis, the influential parameters may be ranked and the SSPM may be eligible for order reduction (i.e., fixing of a parameter). Prior to calibration of SSPMs, nominal values of parameters can be chosen based on prior knowledge. The sensitivity matrix S in Eq. (3.2) can then be constructed by calculating the partial derivatives of the output of the SSPM with respect to individual parameters:

$$\mathrm{S}(\theta) = \begin{bmatrix} \frac{\partial \mathrm{y}(1;\theta)}{\partial \theta_1} & \cdots & \frac{\partial \mathrm{y}(1;\theta)}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{y}(\mathrm{N};\theta)}{\partial \theta_1} & \cdots & \frac{\partial \mathrm{y}(\mathrm{N};\theta)}{\partial \theta_p} \end{bmatrix} \tag{3.2}$$

In this equation y is assumed to be a time-varying output and $y(i; \theta)$ is the value of y at time t_i.

The model parameters can then be ranked using a Singular Value Decomposition (SVD) of this matrix and parameters that are relatively noninfluential may be fixed prior to calibration. Eq. (3.2) applies to both linear and nonlinear models. In the special case where the model can be formulated in linear regression form, the sensitivity matrix can be also considered as the local regressor matrix at the point θ [131].

Depending on the model's COU and type of model under evaluation, PSA may involve incorporation of some level of experimental data. For example, individual subject input measurements (model input) may be used to calculate the sensitivity matrix in Eq. (3.2) as opposed to choosing a nominal input for the model (e.g., step input). Furthermore, the PSA may involve evaluation of information content of output measurements particularly in cases where the COU of the model dictates identification of a calibration window and splitting of the data to enable independent model validation.

3.2.3 Practical Identifiability Analysis (PIA)

Post-calibration Practical Identifiability Analysis (PIA) of the parameters is a step that involves evaluating the reliability of parameter estimates. Reliability of the estimates may be defined as a post-calibration parameter confidence interval within a bounded range and reasonable size range (note that there is no strict definition of practical identifiability). Unreliable parameters may be obtained post calibration due to the limitations in the quality and quantity of the data used

in the fitting process. Reliability of the calibrated parameters can be checked by evaluating their confidence intervals based on a prespecified threshold (e.g., 95% confidence limit). Figure (3.4) depicts scenarios encountered when PIA is evaluated:

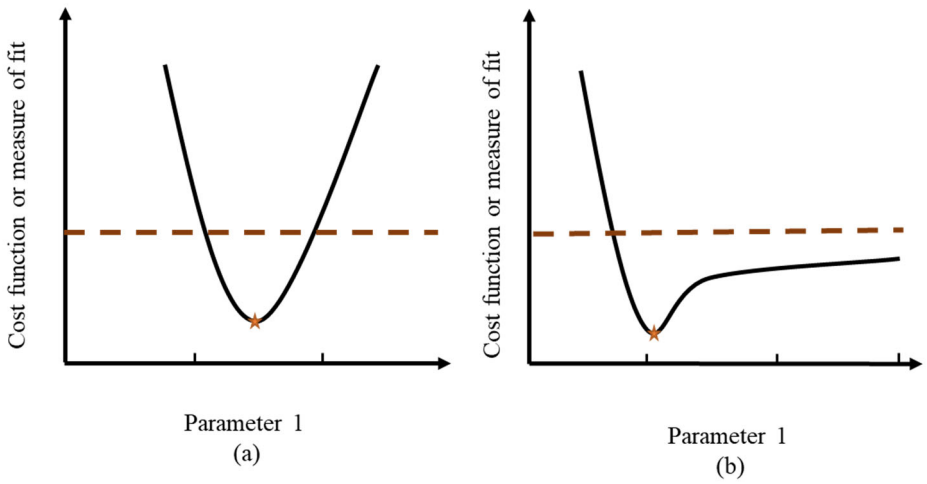


Figure 3. 4: Example cost function curvature under different practical identifiability scenarios. Red dashed lines represent the confidence limits (i.e., confidence intervals are parameter values corresponding to cost function below red dashed line). (a) A practically identifiable parameter. (b) A practically unidentifiable parameter that is unreliable due to the large (or unbounded) confidence interval.

The methods that enable SIA, PSA, and PIA are discussed in detail in section 3.4.

3.3 Overview of the SSPM Credibility Assessment Framework

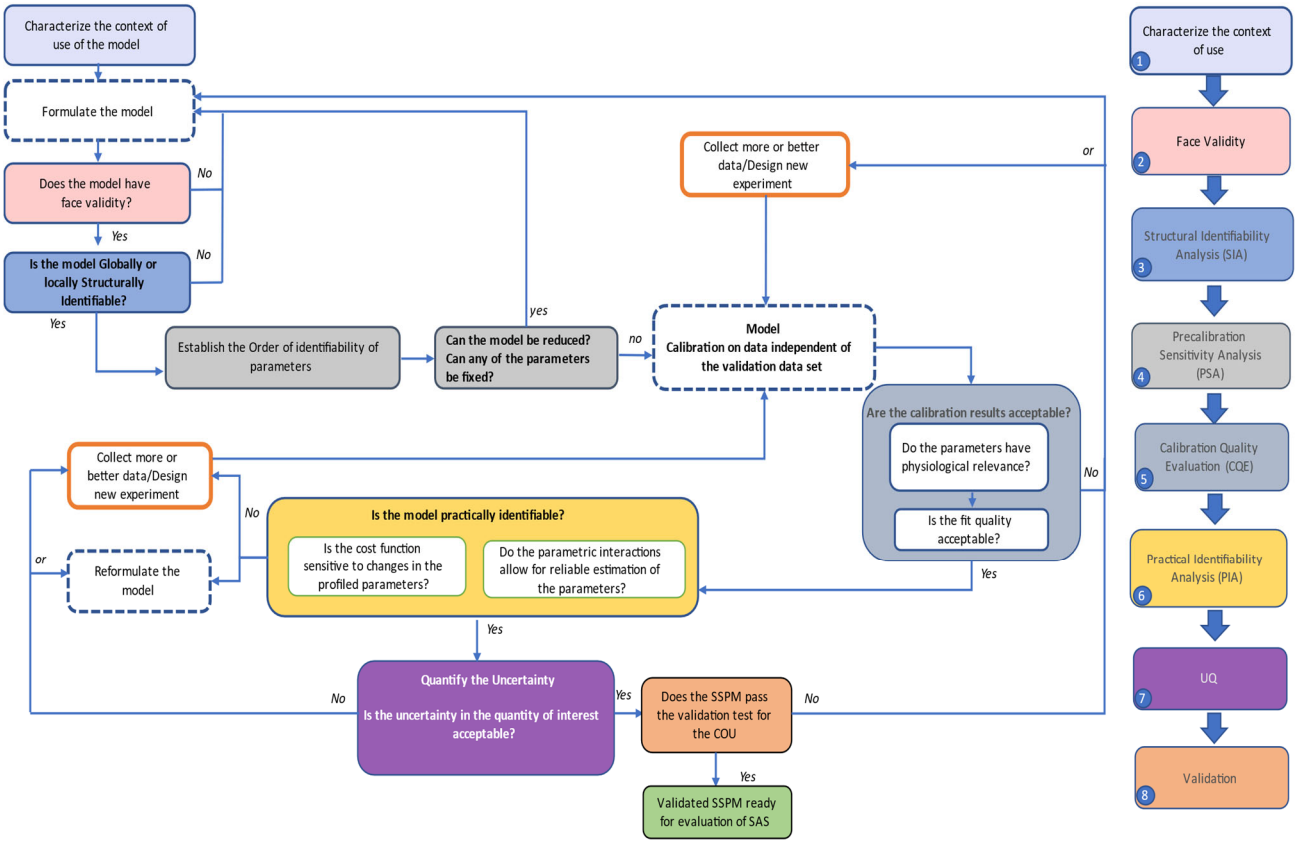


Figure 3. 5: The proposed framework for credibility assessment of SSPMs. The credibility activities are color-coded and listed on the right. The overall workflow with credibility activities included is provided in the left. Model formulation/reformulation and calibration are model building activities and thus not color-coded. See text for description of each step and rationale for ordering of credibility activities.

Figure (3.5) provides an overview of the proposed framework. The credibility assessment is integrated into the model development process, and therefore model development steps such as formulate/reformulate the model, or calibrate the model, are included in Figure (3.5). However, as these are not credibility assessment activities, they are not considered a part of the framework and not assigned a step number. In this section we only provide a high-level overview of the framework, details for each step will be provided in Section 3.4.

The first step of the process involves **characterizing the context of use (COU) of the model** (Step 1). SSPMs are typically used for prediction of model states or parameters. Alternatively, they could be used for descriptive, explanatory [39], or hypothesis testing purposes [132]. Next, the model should be formulated based on its COU (not a credibility assessment step). This involves building the model with particular attention to the physiology of interest and any mechanism that impacts that physiology, including implementing the model in software. Once the model is formulated, its **face validity** (Step 2) needs to be checked. That is, the simulated output of the model using nominal parameter values should be checked for physiological relevance. Failure to demonstrate physiological relevance in the simulation output will need to be addressed by reformulating the model (e.g., by accounting additional mechanisms in the model). Face validity is performed early because for SSPMs it is typically easier to implement and run a candidate model than performing the later analyses.

The next step involves performing **structural identifiability analysis** (Step 3), which involves evaluating the structure of the model to assess if the parameters can be estimated uniquely and reliably. As discussed in section 3.2, this activity requires a mathematical proof of Eq. (3.1) under a theoretical assumption that the data to be used for parameter estimation are fully informative and noise-free. If SIA fails, the model needs to reformulated.

Following this step and prior to calibration of model parameters to data, **pre-calibration sensitivity analysis** (Step 4) is performed whereby the model output is evaluated for sensitivity to perturbations in parameters. While this step may not be a direct and definitive test for parameter identifiability [133], it identifies model parameters that are likely to be unidentifiable as part of PIA. If the model output is insensitive to changes in parameters or their combination, the SSPM

order needs to be reduced. In the simplest case, one or more of the parameters should be fixed to nominal or population average values.

Calibration is the next step of the process, which is a model building activity and therefore detailed discussion of which is beyond the scope of this work. The calibration results should be evaluated as the next credibility assessment step, the **calibration quality evaluation** (Step 5). There are two parts to this step: (i) checking whether the calibrated parameters have physiological relevance (i.e., take physiologically reasonable values) and (ii) quantifying how well the calibrated model fits the data using a pre-specified fit quality metric (e.g., RMSE). If the calibration quality assessment does not meet the acceptance criteria set forth by the fit quality metric or the parameters do not exhibit physiological relevance, additional or higher quality data have to be gathered (e.g., higher sampling frequency or measurement of additional states). Alternatively, one can reformulate the model (e.g., reparametrize or change the model states used in calibration) to allow for acceptable calibration results with the existing data.

Upon successful completion of calibration quality evaluation, the resulting parameters need to be assessed for reliability through **post-calibration practical identifiability analysis** (Step 6). This is the final credibility activity associated with identifiability analysis and is conducted post-calibration to assess the impact of data quality on parameter reliability. PIA is conducted by evaluating the curvature of the cost function and the confidence intervals of the parameters. It should also include the influence of potential parametric interaction on reliability of the individual estimates [28], [134].

PIA is closely related to the next step of the process, **uncertainty quantification** (Step 7)**,** which is to characterize and quantify uncertainty in all model parameters and inputs, and propagating that uncertainty through the model to compute uncertainty in model outputs. In

addition, it should include quantifying the uncertainty in independent experimental data to be used for model validation. Uncertainty quantification and propagation are often ignored in previous works and have only recently been applied in models with biomedical applications [33], [135].

The final step is **model validation** (Step 8)**,** where the model predictions are compared to the experimental validation data using a prespecified validation metric. If the model output or parameters prediction are within an acceptable range using the validation metric, the model may be considered valid for the specified context of use. If not, the model has to be reformulated or additional data may need to be collected to improve the validation results. Note that in the proposed framework it is assumed that the context of use of the model is fixed and may not be changed to ameliorate the validation results.

3.4 Detailed Description of SSPM Credibility Assessment Framework

3.4.1 Step 1: Specify the COU

As the first step in credibility assessment of SSPMs, the COU and the questions the SSPM is intended to address need to be characterized. The majority of such SSPMs are developed for prediction of the patient's observed or unobserved states (e.g., changes in blood volume or interstitial fluid volume), or for prediction of particular physiological parameter values which could be used for clinical applications (e.g., lung volume) [136].

The framework prescribes that the credibility assessment of the SSPM to be initiated with explicit characterization of the COU because the COU should impact the approach followed in subsequent steps and governs the threshold for acceptance of the results (e.g., fit quality or validation metric assessment). The COU is often not stated or implicitly stated in previous work.

Furthermore, *a priori* characterization of COU will eliminate the number of *ad hoc* steps frequently added to model assessment due to facing difficulty in subsequent steps such as conducting various types of identifiability analysis. Assigning and fixing COU *a priori* instills more discipline in the credibility assessment process and enables the modeler to follow pathways to remedy the existing issues for the specific COU at hand (e.g., by designing new experiments, obtaining additional data, or by reformulating the model).

With the COU assigned, the SSPM needs to be formulated based on physiological governing equations (e.g., mass and fluid transport). This model building step often involves simplifying assumptions (e.g., assuming stationarity) which need to be described with justifications.

3.4.2 Step 2: Evaluation of Face Validity

Next, the formulated SSPM should be evaluated for face validity. Face validity in this context may be regarded as a predominantly qualitative activity evaluating 1) transparency of SSPM's input(s), parameter(s), and output(s); and 2) extent to which the SSPM's output are clinically/physiologically reasonable. This can be done by comparing the SSPM's output using nominal parameter values as input as well as internal states with expert opinion or reported values in literature. If the SSPMs output is not consistent with physiological knowledge the model needs to be reformulated. Furthermore, SSPMs output may be evaluated qualitatively for trends as a result of changing the parameters and assessed if the change in model output is consistent with the expected effect of the change in the parameter. For example, if an SSPM is formulated to have a parameter denoting the subject's effective lung volume, then an increase in this parameter should correspond to improved gas exchange. While most SSPMs reported in literature involve some form

of face validity assessment, the activity is often not conducted as part of systematic framework and not informed by COU of the model. If the SSPM passes the face validity tests, it can proceed to be evaluated by the first quantitative credibility activity.

3.4.3 Step 3: Structural Identifiability Analysis (SIA)

For parameters of the SSPM to be estimated uniquely or to be identifiable, structural identifiability is the necessary condition while practical identifiability is the sufficient condition [137], [138]. Due to SIA being independent of quality of data (i.e., it is purely an assessment of mathematical structure of the model), to save time and resources related to handling of experimental data, it should be the first of three identifiability steps that inform the credibility assessment of the model throughout the framework. This step is often neglected in existing work on SSPMs and avoiding it as the first identifiability analysis step can be costly, because unidentifiability in subsequent steps such as PIA (frequently encountered in SSPMs due to low quality or sparsity of data) may be mistakenly attributed to quality of data and lead to unnecessary data collection which can be costly and time consuming. Furthermore, in situations where data collection is to be prospective, theoretical evaluation of model structure through SIA can lead to important considerations for design of experiments (e.g., the analytical methods can reveal that initial conditions have to be specified and measured for the model in order to allow the parameters to be identifiable) [139], [140]. While SIA is often ignored in existing SSPM works, it constitutes a major credibility assessment in other types of models such as epidemiological models [24]–[27] and systems biological models [28]–[33].

SIA can be conducted using various methods in Table (3.1). These methods may be broadly divided into analytical and numerical approaches. Analytical SIA involves a mathematical proof of Eq. (3.1) to determine whether the model is globally structurally identifiable (GSI). Various methods have been developed for analytical SIA as summarized in Table (3.1). Such methods are typically tailored towards the unique characteristics of the SSPMs (e.g., linearity in parameters). For detailed and critical comparison of methods refer to [141].

Most analytical methods involve significant mathematical complexity and may be exceedingly computationally difficult or impractical to conduct for complex ordinary differential equation (ODE) models or with delay differential equation (DDE) models [129], [139], [142],[143]. In such scenarios, numerical approaches may prove effective in conducting local SIA.

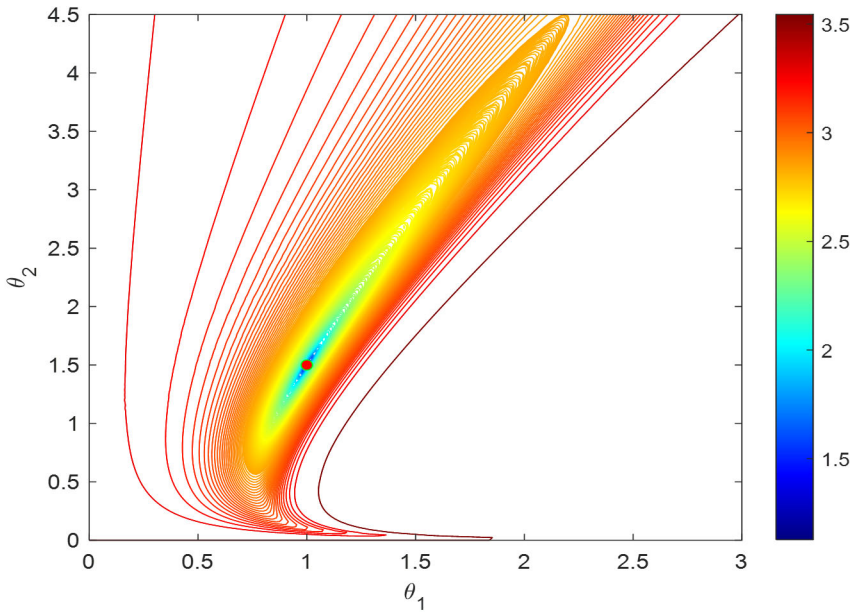
Table 3. 1: SIA methods

Method	Description	Local or global	Considerations
Analytical methods			
Transfer function approach[144]	Based on conversion of model state space to transfer function form. It's a direct test for SIA.	Global	Linear models only
Taylor series approach [129], [141]	Model output expanded as Taylor series	Global	Applicable for nonlinear systems
Similarity Transformation[129], [141]	Analytical method relying on Kalman's algebraic equivalence theorem	Global	Applicable for nonlinear systems
Differential algebra [145]	Analytical method Substitution and differentiation to derive algebraic input/output equations and to check for symmetries in model equations	Global	Applicable for nonlinear systems. May be difficult to implement for higher order models. Heavily dependent on symbolic computation. Commercially available.
Extended Observability approach [130]	Treats model parameters as constant states and evaluates their observability	Local	May be difficult to implement for higher order models. Commercial software available

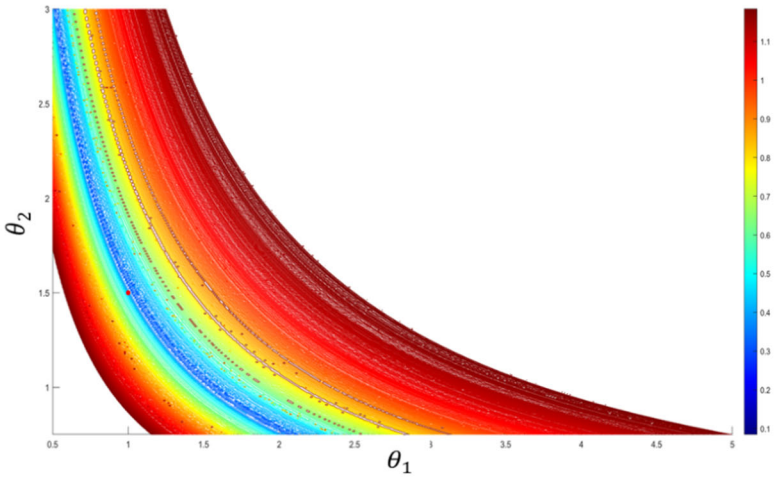
Numerical Methods			
Sensitivity based methods [129], [131]	Fisher Information Matrix Calculation /Jacobian rank calculation /confidence interval estimation	Local	Confidence Intervals estimated are asymptotic and symmetric
Profile likelihood [28], [146]	Evaluates the curvature of cost function or log-likelihood function as one parameter is profiled and others are optimized	Local	Computationally intensive due numerical reoptimization. Account for potential asymmetry in confidence intervals. Account for parametric interactions.
Bayesian method [147], [148]	Bayesian Markov chain Monte Carlo (MCMC) sampling approach.	Local	Hard to conduct for non-identifiable systems due Markov chains convergence issues. Priors need to be informative and based on biological insight.

Local SIA can be conducted by generating synthetic input resulting in synthetic model output and initializing the optimization at the same nominal assumed set of parameters. If the calibration results are identical to the true parameter values, the model may be deemed locally structurally identifiable (LSI). The drawback of this approach is that the results are local, and only valid in the neighborhood of the explored parameter space. A step-by-step process of LSI analysis can be found in [129].

Uniqueness of the model parameters may be visually inspected by contour plots of the cost function versus the two-dimensional parameter plane. Bounded contours encapsulating the identified parameter demonstrate a model that is LSI while unbounded contours signify a model that is not locally structurally identifiable (NLSI). Figure (3.6) depicts an example of such visual inspection for examples of LSI and NLSI models.



(a)



(b)

Figure 3. 6:Examples of contour plots for LSI(a) and NLSI (b) models. Color represents the cost function value. In (a) there is a single minimum, whereas the minimum is attained on a 1D curve of values in (b).

If the SSPM is deemed unidentifiable at this stage, it needs to be reformulated. This could involve reducing the number of parameters or otherwise altering the parameterization. Alternatively, new or different model outputs could be defined [129].

3.4.4 Step 4: Pre-calibration Sensitivity Analysis

If the SSPM is found to be GSI or LSI, the SSPM may subsequently be considered for calibration with data. However, prior to this step, it is necessary to quantify the influence of the parameters on the output of the SSPM or the chosen optimization cost function, a credibility activity known as Sensitivity Analysis [40], [138]. We recommend performing sensitivity analysis in advance of calibration and refer to this as Pre-calibration Sensitivity Analysis as (PSA).

By performing this preliminary sensitivity analysis, one can trace potential difficulties associated with the calibration process to the model and/or data instead of trying various optimizers thus saving time and computational cost. The previous work on development and evaluation of SSPMs do not include this step or include it in this order as a pre-calibration step and the current process often involves rushing into model calibration without any preliminary identifiability analysis which is inefficient.

Furthermore, our literature review of SSPMs reveals considerable number of SSPMs which may be over-parametrized due to attempts to capture underlying complex physiology of interest [7], [63], [84], [149]. Calibration of such SSPMs to data may be difficult, or if successful, it is possible that noise in the physiological data may be fitted (a problem known as overfitting [150]) and therefore the SSPM will have poor predictions. Ordering the parameters based on how influential they are prior to calibration of the SSPM is necessary as it allows for model order reduction by either fixing a subset of parameters or their ratios to streamline the calibration process. Another advantage of having performed PSA at this stage of SSPM evaluation process is that it

will allow for the selection of a calibration window [151] for cases where it is desirable to split the timeseries datasets into independent calibration and validation segments. Characterization of a calibration window is only necessary if the model's intended purpose is to predict one or more of the physiological state or output variables (e.g., end-tidal CO_2 concentration or interstitial fluid volume).

PSA may be conducted analytically by taking the partial derivatives of model output Eq. (3.2) with respect to the parameters. It can also be calculated numerically by applying prespecified perturbations to individual parameters and monitoring SSPM's output trajectories with respect to each parameter variation. This can be done either by visualizing the cost function as a function of parameter changes and evaluating its curvature or by applying SVD to the sensitivity matrix in Eq. (3.2)[131], [138]. It should be noted that this approach provides information on the sensitivity of the outputs of the SSPM to individual parameters and may obscure information on the influence of multiple changes in parameters on the outputs of the SSPM. Parameter collinearity can result in overall model unidentifiability [129], [133]. In such scenarios, global sensitivity analysis such as the Sobol sensitivity analysis may be leveraged to account for model output variance as a result of parametric interactions [33]. An effective method for identifying parameter collinearity is profile likelihood [28] which will be discussed in detail in section 3.4.6. Once the SSPM has passed the PSA stage, it can be calibrated to data. The calibration process itself is not a model credibility activity and therefore is outside of our framework.

3.4.5 Step 5: Calibration Quality Evaluation (CQE)

This is a necessary step following successful calibration of SSPM to data. This step is comprised of two stages. First, the physiological relevance and plausibility of the parameters must be evaluated. If a calibrated parameter demonstrates physiologically implausible values the model

formulation needs to be reevaluated. Second, the quality of calibration or an SSPM's fit to experimental data may be quantified using a variety of statistical metrics. Three commonly used statistics to evaluate the goodness of fit include F-test, R squared, and RMSE. All three metrics quantify the distance between the measurements and model predictions in one aggregate measure of fit [131], [152].

3.4.6 Step 6: Practical Identifiability Analysis (PIA)

Post-calibration identifiability analysis of the parameters in an SSPM should be conducted to evaluate the reliability of the parameter estimates. As part of PIA, parametric interactions should be also quantified to ensure that combinations of parameters with different values do not result in a nearly flat cost function curvature rendering the individual parameters and thus the overall SSPM unidentifiable. Since the SSPM has already passed the SIA and PSA tests, the likely root cause of unidentifiability at this stage would be presence of pairwise parametric interactions interaction as a result of calibration of SSPM to data.

All numerical methods used for SIA in Table (3.1) may be used for the purpose of evaluating PIA with a single difference that for SIA the data are synthetic and noise free and for PIA actual experimental inputs and outputs are applied to the calibrated SSPM. Out of these methods, the profile likelihood has been proven one of the most reliable and versatile methods to investigate practical identifiability. In this iterative method, each parameter is varied over a specified range and the value of the cost function computed after minimizing over the remaining parameters. The result may be depicted as p one-dimensional plots (p the number of parameters), each of the form: [28]

$$CF_{PL}(\theta_i) = \min_{\theta_{i\neq j}} \left\| \left(\mathrm{Y_m}(\theta, \mathrm{t}) - \mathrm{Y_e}(\mathrm{t}) \right) \right\|_2 \tag{3.3}$$

where θ_i is the profiled parameter and CF_{PL} is the cost function minimized through calibration of the rest of model parameters, $\theta_{i\neq j}$. $\mathrm{Y_m}$ is the model output or any quantity of interest that needs to be calibrated to the experimental measurements ($\mathrm{Y_e}$), and it is assumed in this equation that the cost function is the L2 norm. Profile-likelihood results provide information on interaction between parameters, as by plotting the values of the minimizing parameters in Eq. (3.4):

$$\theta_{i\neq j}(\theta_i) = \arg\min_{\theta_{i\neq j}} \left\| \left(\mathrm{Y_m}(\theta, \mathrm{t}) - \mathrm{Y_e}(\mathrm{t}) \right) \right\|_2 \tag{3.4}$$

against the varied parameter θ_i, it can be ascertained if any parameter, θ_k, changes its value to compensate for the changes in θ_i and keep the overall cost function value low.

Reliability of the parameter estimates can be evaluated using point-wise or simultaneous confidence interval estimation. This can be done in scenarios where asymptotic theory is applicable (i.e., the number of observations is much larger than the number of parameters) [131]. In cases where the number of experimental observations is comparable to the number of model parameters, a profile likelihood-based confidence interval may be leveraged [28], [128].

As such, profile likelihood may be used as a versatile method to investigate both uniqueness and reliability of the parameters in an SSPM because it can detect unidentifiability due to cost function insensitivity to individual parameters as well as unidentifiability due to parametric interactions. By quantifying the interaction between parameters, curing potential unidentifiability issue may be undertaken by reparameterization of the SSPM into identifiable parameter combinations [137], [153].

3.4.7 Step 7: Uncertainty Quantification

Uncertainty quantification (UQ) can be conducted on both the outputs of an SSPM or experimental measurements. As applied to SSPMs, UQ consists of two major steps: 1) uncertainty characterization (UC) and 2) uncertainty propagation (UP)[33]. Model UC involves the quantification of various sources of uncertainty in model inputs. This may involve both input variables and input parameters. In many applications of SSPMs, the uncertainty in model input variables such as infusion rate or minute ventilation are solely instrument dependent and is relatively small compared to the sensors measuring the variable physiological responses (e.g., blood volume) which integrate both instrumental uncertainty and physiological variability. As such, parametric uncertainty may be regarded as the predominant source of uncertainty in many applications. UQ is preceded by PIA because PIA is an enabling step for UC by providing information towards parameter confidence intervals (i.e., parametric confidence intervals are estimated as part of PIA). UP involves propagating the calculated model input uncertainty through the SSPM to gain insights into the uncertainty range in the outputs of the SSPM. For example, the calculated confidence intervals for each parameter may be propagated through the SSPM by assuming multivariate normal distribution which is a generalization of the univariate normal distribution to multiple variables.

UQ enables a probabilistic evaluation of uncertainty in the model outputs as opposed to a binary prediction and comparison with experimental data. Therefore, the experimental uncertainty in validation data should also be quantified to allow for a probabilistic comparison with model predictions in the step.

3.4.8 Step 8: Model validation

The final step may be thought of as the ultimate test for the model credibility assessment, that is to test its predictive capability using data that did not take part in the calibration procedure. Having performed UQ for both SSPM and experimental outputs in the previous step, the two can be compared in a probabilistic sense to gain insight on predictive capability of the SSPM. Various methods exist to quantify the mismatch between experimental and simulation results. In order for this comparison to be made, a validation metric based on the application of SSPM needs to be selected. A validation metric is a mathematical operator that measures the difference between a system response obtained from a simulation result and the one obtained from experimental measurements [154]. In case of probability distributions being compared, the validation metric could be a statistical measure of the degree of overlap between the two distributions [155].

Chapter 4: Applying the Framework Towards the Credibility Assessment of a Respiratory Gas Exchange Model

4.1 Introduction

In this chapter we provide the first application of the novel framework developed and described in Chapter 3 to a previously published SSPM. The specific model to be evaluated is a model of carbon dioxide (CO_2) exchange. Details of model development, formulation, and experimental protocol used for data collection can be found in [77]. A high-level description of the model is provided in Figure (4.1). This figure depicts an overview of the CO_2 exchange physiology which is governed by mass balance of CO_2. CO_2 is removed from the tissues and is transferred to the lungs by blood. The rate of this removal is dependent on the partial pressure of CO_2 in the venous blood leaving the tissue ($P_V CO_2$) and Cardiac Output (Q). CO_2 removal from the lungs is in turn determined by the product of respiratory tidal volume and breathing rate. This is known as minute ventilation (MV) or ($\dot{V}_A$) and serves as the input to the model. Circulation is completed when CO_2-deficient arterial blood with partial pressure of CO_2 denoted as ($P_A CO_2$) arrives at the tissues. CO_2 is also generated at the tissue level as part of body's maintenance of the metabolic activity ($\dot{V}_{CO2}$) which is included as part of mass balance of CO_2 in the blood entering and leaving body tissue. Two transport delays τ_1 and τ_2 may be considered for the lags existing on the venous and arterial side of the blood circulation respectively.

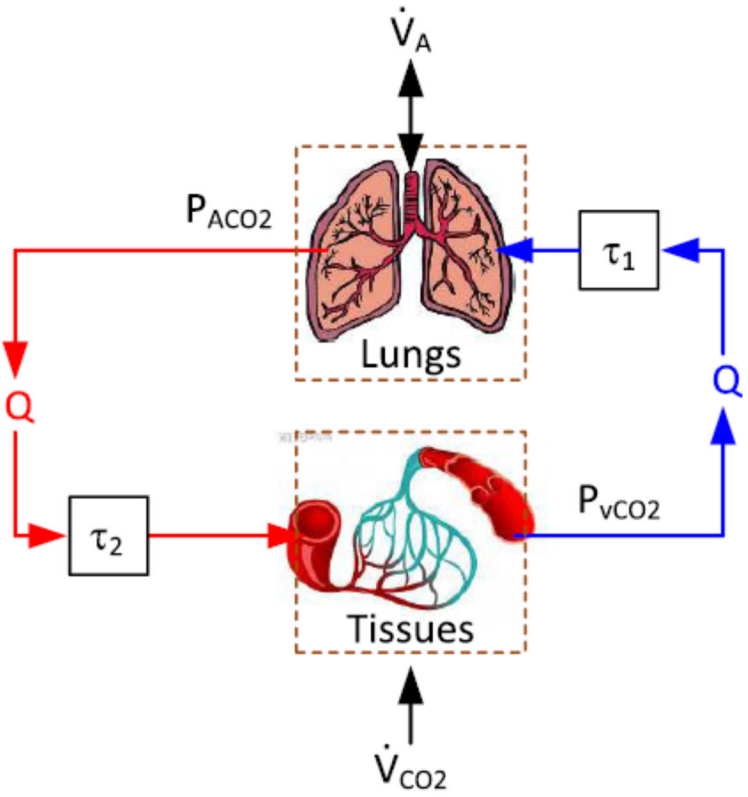


Figure 4. 1: Schematic of CO_2 exchange in blood circulation reproduced from Kim et al. 2016 [77]

4.2 Application of the Framework to CO_2 Exchange Model

4.2.1 Step 1: Specification of the CO_2 Model COU

4.2.1.1 COU

Generally, models developed for respiratory gas exchange are used for therapeutic purposes in line with intended use of SAS (e.g., informing control action or adjustment of ventilator settings such as minute ventilation through a clinical decision recommending platform). In such capacity, the model has a predictive context of use. Predictions are targeted towards the model's observed or unobserved states, or in case of a model that is equipped with physiologically relevant set of parameters, accurate and reliable prediction of such parameters for individual

patients may be extremely valuable for diagnostic and therapeutic purposes. Because the parameters of the CO_2 exchange model (below) benefit from such physiological relevance (i.e., they represent known physiological parameters) [40], [77], the COU of the model was designated to be: prediction of subject-specific model parameters representing lung volume, cardiac output, metabolic rate, and body tissue volume for pediatric patients using patient measurements of MV (model input) and end-tidal CO_2 (used to calibrate model).

4.2.1.2 Model Formulation

The two main models discussed in the referenced publication are a 4-parameter ODE model and a 6-parameter DDE model which includes the two transport delays described in the previous section. Model equations are provided for reference and additional information regarding different model formulations and assumptions behind model development may be found in the original publication [77].

The 4-parameter model, hereafter referred to as the 4p model, is described with the following state space equations:

$$\begin{aligned} \dot{x}_1 &= \theta_2(x_2 - x_1) - \theta_1 x_1 u \\ \dot{x}_2 &= \theta_3 (x_1 - x_2) + \theta_4 \\ Y &= x_1 \\ x_1(0) &= C_1, x_2(0) = C_2 \end{aligned} \tag{4.1}$$

where $x_1 = P_A CO_2$, $x_2 = P_V CO_2$, $u = \dot{V}_A$, with initial conditions $C_1 = P_A CO_2(0)$ and $C_2 = P_V CO_2(0)$.

The model's output Y is the partial pressure of CO_2 in the arterial blood circulation which the model developers assumed to be a valid surrogate for end-tidal CO_2 concentration (EtCO2).

Parameters of the model have physiological relevance and transparency. θ_1 is the model parameter corresponding to effective lung volume (V_L), θ_2 corresponds to the cardiac output (Q), θ_3 to effective tissue volume (V_B), and θ_4 corresponds to the body's CO_2 production (V_{CO2}) as the result of metabolism.

The 6-parameter model, hereafter referred to as the 6p model, has the same formulation as the 4p model with the exception that the transport delays are included for each model state representing the arterial and venous CO_2 concentrations. The 6p model is described as follows:

$$\begin{aligned} \dot{x}_1 &= \theta_2(x_2(t-\tau_1) - x_1) - \theta_1 x_1 u \\ \dot{x}_2 &= \theta_3\,(x_1(t-\tau_2) - x_2) + \theta_4 \\ Y &= x_1 \\ x_1(\mathrm{t}) &= f(t) = C_3 \quad -\tau_1 \le \mathrm{t} \le 0, \quad x_2(\mathrm{t}) = g(t) = C_4 \quad -\tau_2 \le \mathrm{t} \le 0 \end{aligned} \tag{4.2}$$

where $f(t)$ and $g(t)$ describe the history functions of the DDE, which in this work are assumed to be functions constant with respect to time.

Depending on how the history functions are treated in the parameter estimation, variations of the 6p model may be constructed, referred to as the 7p and 8p models. Table (4.1) provides a list of the variants of the CO_2 exchange model considered. The DDE models are hereafter referred to as 6p+ models.

Table 4. 1: CO_2 model variants

Model	Description	Prespecified parameters	Identified Parameters
4p	ODE with initial conditions prespecified	C_1, C_2	$\theta_1, \theta_2, \theta_3, \theta_4$
6p	DDE with delays identified but history functions prespecified as constants	C_3, C_4	$\theta_1, \theta_2, \theta_3, \theta_4, \tau_1, \tau_2$
7p	DDE with delays identified, and history functions that are constant. The first history function value is prespecified and the difference between the two history functions is identified	$f(t) = C_3$	$\theta_1, \theta_2, \theta_3, \theta_4, \tau_1, \tau_2, d$ $d = f(t) - g(t)$
8p	DDE with delays identified and history functions that are constant. The first history function value and the difference between the two history functions are identified	None	$\theta_1, \theta_2, \theta_3, \theta_4, \tau_1, \tau_2, d, C_3$

4.2.2 Step 2: Evaluation of Model's Face Validity

4.2.2.1 Model Simulation

Following formulation of the model, its face validity needs to be evaluated by simulating the model using nominal parameters. This was conducted on the 4p and 6p model as the 6p+ models involve the same type of simulation steps and only differ in the number of optimized parameters. Quality of simulation results depends on accuracy of the numerical integration process used to solve the differential equation. This is especially important for simulation of the DDE models since the DDE solver may be sensitive to properties such as step size and tolerances. We used the commercially available DDE solver dde23 as part of MATLAB which has proven utility and robustness for DDEs with constant delays [156]. The solver properties such as step size,

relative tolerance and absolute tolerance were adjusted to be sufficiently small and were fixed following this step and throughout the work.

4.2.2.2 Evaluation of Models Face Validity

As noted, parameters of the model have full physiological relevance and transparency. Therefore, nominal values for such parameters including the transport delays may be extracted from published literature and used to inform other aspects of model face validity such as evaluation of the physiological relevance of the model output Y ($EtCO_2$).

We proceeded with identifying such nominal values as listed in Table (4.2). Several simulations were then generated based on physiologically practical and relevant model inputs ($\dot{V}_A$) for the models listed in Table (4.1). The simulated model output Y ($EtCO_2$) for both 4p and 6p model was generated and evaluated for face validity by comparing the output with results in literature such as works reported in [84], [157], [158]. Figure (4.2) includes model simulations for the 4p model which is qualitatively consistent with those reported in [156] in terms of range of $EtCO_2$ reported for pediatric subjects with similar minute ventilation input. The 6p model demonstrated similar results to the 4p model. Furthermore, the influence of the changes in the parameters on model output was evaluated qualitatively and verified to make physiological sense. For example an increase in θ_1 results in a decrease in $EtCO_2$ which is expected because θ_1 corresponds to the inverse of effective lung volume V_L and a decrease in lung volume would decrease CO_2 exchange resulting in lower $EtCO_2$. The same qualitative consistency was observed in other model parameters.

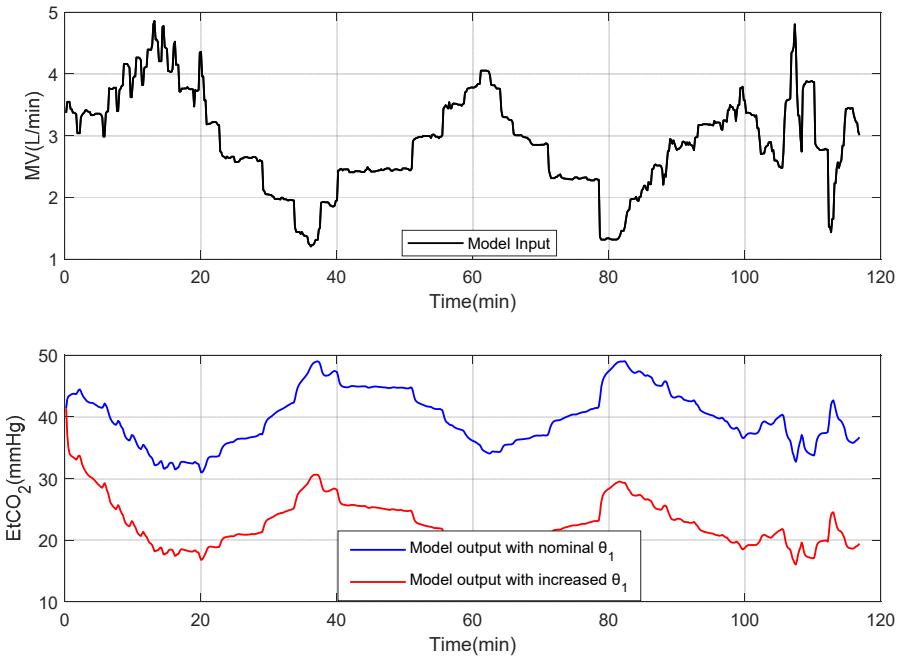


Figure 4. 2: An example scenario for evaluation of face validity of the CO_2 exchange model

Table 4. 2:Nominal values for the 4p and 6p models

Model	**Parameters**					
	θ_1	θ_2	θ_3	θ_4	$\tau_1(s)$	$\tau_2(s)$
4p	0.43	3.1	0.2	3.12	NA	NA
6p	0.66	3	1.12	5.4	5	6

4.2.3 Step 3: SIA

Performing SIA at this stage has significant advantages in terms of saving time and resources for SSPM modelers. This is realized by answering the question of whether the model structure allows for the parameters to be estimated uniquely early on in the model development process and prior to either processing retrospective data or collecting prospective data. Such a determination is crucial especially if an unidentifiability conclusion is reached as part of

subsequent steps because establishing SIA allows us to eliminate the model structure as the source of unidentifiability and focus on gathering more informative data for parameter estimation.

4.2.3.1 Methods

As discussed in Section 3.4.3 of Chapter 3, different methods exist for conducting SIA depending on the structure and properties of the model to be evaluated. The models listed in Table (4.1) have different structures and properties. The 4p CO_2 model is a nonlinear ODE model which does not involve transport delay parameters and is specified by initial conditions. The rest of the models included in Table (4.1), in addition to being nonlinear are DDE and thus require specification of a history function.

For the 4p model we leveraged the analytic method provided by an open-source software known as COMBOS [159] which uses differential algebra to systematically provide an exhaustive ordering and reordering of Grobner bases [159]. The result was that the 4p model is indeed GSI.

While analytic methods involving GSI analysis may be applied to the 4p model, an analytical method for SIA and evaluation of such is not readily available for 6p+ models. In fact, conducting SIA analytically may not be practical to perform on such complex models even with advanced software tools available such as STRIKE GOLDD2, DAISY, or COMBOS [129], [160]. In such a circumstance, numerical approaches may be leveraged to conduct SIA and to locally evaluate structural identifiability of the DDE models . Leveraging a numerical approach, first nominal parameter values are used to generate simulated synthetic data to be used for calibration of the model parameters in the subsequent step by minimizing Cost Function (CF) in Eq. (4.3) where Y_m and $Y_{synthetic}$ are model output and the synthetic data generated by the model, respectively. If the resulting parameters are the same as the original set of parameters, the model

may be LSI at the nominal parameter values[129]. For 6p+ models we employed a combination of sensitivity-based methods and the profile likelihood method to develop three numerical *structural identifiability tests* for SIA discussed below.

$$\mathrm{CF}(\theta) = \left\| \left(\mathrm{Y_m}(\theta, \mathrm{t}) - \mathrm{Y_{synthetic}}(\mathrm{t}) \right) \right\|_2 \tag{4.3}$$

4.2.3.2 Description of numerical structural identifiability test for SIA of 6p+ models

Structural Identifiability Test Development

As the first step for development of the test, True Positive (TP) and True Negative (TN) scenarios were defined and determined using two models with known identifiability properties to serve as reference models. For the first reference model, a PK/PD model described in Eq. (4.4) with known structural nontrivial structural unidentifiability was selected. This model is only identifiable when θ_2 is fixed [138].

$$\begin{aligned} \dot{x}_1 &= -(\theta_1+\theta_4)x_1 + \theta_3 x_2 + u, \\ \dot{x}_2 &= \theta_4 x_1 - (\theta_2+\theta_3)\, x_2 \\ Y &= x_1,\ x_1(0) = 0, x_2(0) = 0 \end{aligned} \tag{4.4}$$

The 4p model which was determined analytically to be GSI was leveraged as the second reference model. Because this model as formulated in Eq. (4.1) is GSI, structural unidentifiability was introduced trivially by introducing an additional parameter θ_5 as follows (Eq. 4.5).

$$\dot{x}_1 = \theta_2(x_2 - x_1) - \theta_1\theta_5 x_1 u$$
$$\dot{x}_2 = \theta_3\,(x_1 - x_2) + \theta_4$$
$$y = x_1,\; x_1(0) = 0, x_2(0) = 0 \tag{4.5}$$

This model is referred to as 4p+1 in the results section. Overall, we have four reference model variants, two that are LSI (and GSI), two that are not locally structurally identifiable (NLSI). Next, we defined three numerical LSI tests. The tests are summarized below and in Table (4.3).

Table 4. 3: Numerical structural identifiability tests.

Test	Description	Positive outcome	Negative outcome
1) Profile Likelihood	Evaluating the curvature of the cost function (CF)	Model is LSI if all parameters when profiled produce curvature in CF	Model is NLSI if any of the parameters profiled produce flat CF
2) Rank of the Jacobian	Evaluating rank of the Jacobian of the model output for 1000 points in parameter space	Model is LSI if the Jacobian matrix has full rank for percentage of points above sensitivity threshold. Sensitivity threshold: 95%	Model is NLSI if the Jacobian matrix is rank deficient for percentage of points above specificity threshold. Specificity threshold: 95%
3) Confidence Interval (CI)	Evaluating the boundedness of the CIs for N=1000 points in the parameter space when 1% noise is introduced in the data	Model is LSI if all P parameters have bounded CI for percentage of points above sensitivity threshold. Boundedness Thresholds: 1×10^2 Sensitivity threshold: 95%	Model is NLSI if any of the P parameters have unbounded CI for percentage of points above specificity threshold. Boundedness Thresholds: 1×10^2 Specificity threshold: 95%

1) Profile likelihood of the individual parameters was evaluated in the neighborhood of physiologically relevant nominal parameters values. Specifically, nominal parameter values are chosen, synthetic data are constructed by solving the model at these parameters, and that synthetic data is used as the experimental data for calculating cost functions. Profile likelihood was then performed for each parameter. The test threshold distinguishing between identifiable and non-identifiable model was based on the relative curvature of the cost function. A flat curvature represents (Not Locally Structurally Identifiable) NLSI. Nominal values were selected based on the reported values in literature included in Step 2.
2) Because profile likelihood test is computationally costly, and we are interested in structural identifiability properties of the model beyond the chosen nominal values and in the neighborhood of the selected parameters, a sensitivity-based test known as the "rank test" was employed. Here, a point in parameter space was sampled and then we calculated the rank of the sensitivity matrix or the Jacobian of the model output in Eq. (3.2). A full-rank matrix indicates the model is LSI at the sampled point. The process was repeated using N=1000 randomly selected points in the parameter space.
3) As the final test and to confirm the results obtained from the previous two steps, the synthetic model output for the N=1000 points in the parameter space from Test 2 was contaminated with a small amount of gaussian noise (1% amplitude). The boundedness of the confidence intervals for each parameter estimated using this synthetic data was then evaluated to determine LSI properties. Large confidence intervals indicate the model is NLSI at the sampled point.

A threshold of 95% was selected for the confidence interval and rank tests to discern between LSI and NLSI. Because the confidence interval test relies on the length of the confidence intervals to draw conclusion on LSI vs NLSI, the boundedness of the confidence intervals was established based on a threshold of two orders of magnitude from the estimated parameter.

Results from all three tests were merged as follows. A model was assessed to be LSI if it was determined to be LSI in all three tests stated above. A model was assessed as NLSI if it was assessed as NLSI in any of the three tests. Each test was applied to each of the reference models to verify that the test can correctly determine the identifiability properties of the reference models.

Summary of Results

Table (4.4) includes the summary of LSI test results for all three tests for all reference models and the CO_2 model variants. Individual tests results are provided in the subsequent section. The numerical structural identifiability tests that we developed correctly identify the identifiability properties of the four reference model variants. The tests then show that the 6p+ models are all LSI.

Table 4. 4: Summary of LSI test results for all models

Reference Models					
Models	**Rank Test Results**	**CI test Results***	**PL test Results***	**Model known to be LSI?**	**Test Conclusion**
4p	99.1% full rank	98.1% Bounded	Unique local minimum found	Yes	LSI
4p+1	100% rank deficient	99% Unbounded	Unique local minimum not found	No	NLSI
PKPD (θ2 fixed)	98.6% full rank	98.3% Bounded	Unique local minimum found	Yes	LSI
PKPD	99.5% rank deficient	96.9% Unbounded	Unique local minimum not found	No	NLSI
Evaluated Variations of the CO_2 Model					
	Rank Test Results	**CI Test Results**	**PL Test Results**	**Model known to be LSI?**	**Test Conclusion**
6p	98.3% full rank	97.4% Bounded	Unique local minimum found	Unknown	LSI
7p	97.4% full rank	99.3% Bounded	Unique local minimum found	Unknown	LSI
8p	97% full rank	97.1% Bounded	Unique local minimum found	Unknown	LSI

*results are for all parameters profiled in the respective models

Results for 4p Reference Model

1) Profile likelihood test:

An example of the profile likelihood test result is depicted in Figure (4.3). Results are only shown for one profiled parameter as the profile likelihood test for the rest of the 4p model parameters follow the same pattern. The cost function for the 4p model demonstrates curvature,

and therefore this test correctly concludes the 4p model is LSI at the nominal parameters. The cost function for the 4p+1 model is flat and therefore the test correctly concludes the 4p+1 model is NLSI.

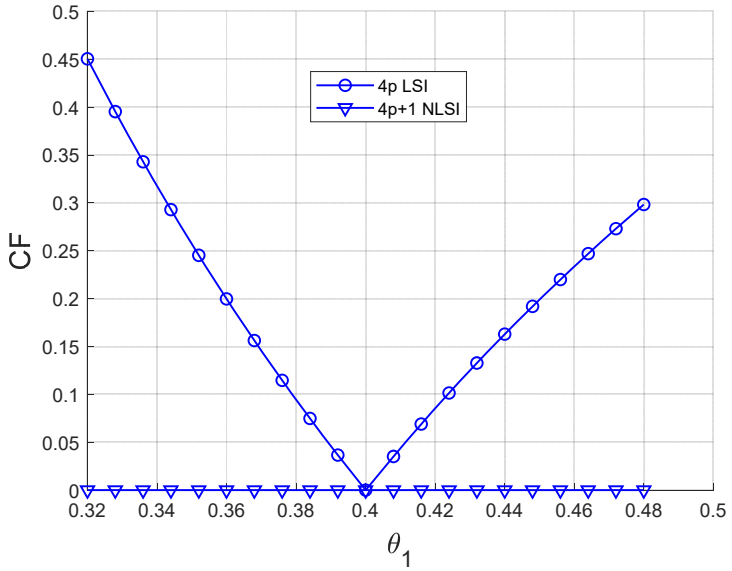


Figure 4. 3: One dimensional profile likelihood test for the 4p model.

2) Confidence Interval Test:

Confidence intervals for 4p and 4p+1 models in absence and presence of 1% noise for one set of model parameters are provided in Table (4.5). (The absence of noise results are only presented to demonstrate the need for introducing noise in this test). The 4p+1 model is correctly deemed NLSI because the span in confidence interval is several orders of magnitude larger than the nominal parameter value. The confidence interval for the 4p model with noise stays bounded and with bounds in the same order of magnitude as the nominal parameters, and therefore the 4p model is correctly deemed LSI. The results for positive outcome and negative outcome were 98.1% and 99% respectively.

Table 4. 5: Confidence intervals for 4p and 4p+1 models in absence and presence of noise, for one point in parameter space.

	4p		**4p+1**	
Parameters	Lower CI	Upper CI	Lower CI	Upper CI
θ_1	1.26	1.26	1.26	1.26
θ_2	3.58	3.58	3.58	3.58
θ_3	1.49	1.49	1.49	1.49
θ_4	2.85	2.85	2.85	2.85
θ_5	NA	NA	0.9	0.9
	with noise		**with noise**	
	Lower CI	Upper CI	Lower CI	Upper CI
θ_1	1.01	1.51	**-2.70E+08**	**2.70E+08**
θ_2	2.87	4.30	1.26	3.04
θ_3	1.19	1.79	0.64	0.69
θ_4	2.28	3.42	0.86	0.97
θ_5	NA	NA	**-1.74E+08**	**1.74E+08**

3) Rank Test

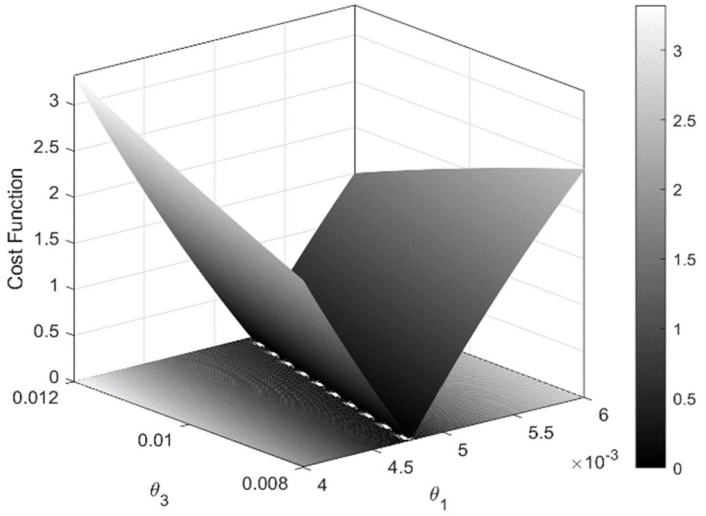
For the 4p model the Jacobian was full rank (rank=4) for 100% of points, so the test correctly deemed the 4p model to be LSI. For 4p+1 model the Jacobian was rank deficient (rank=4) for 100% of points and full rank for 99.1% of the points in parameter space, so the test correctly determined the model to be NLSI.

PK/PD Reference Model Results

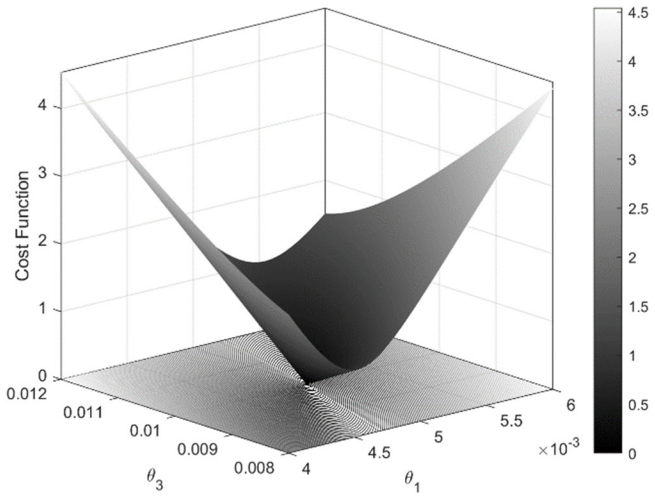
1) Profile Likelihood Test

The result of profile likelihood test for the reference PK/PD model are demonstrated in Figure (4.4) in two dimensions. When the model with its original formulation is tested using profile likelihood the model is NLSI because the valley on the two-dimensional parameter plane represents infinite number of parameter combination that result in the cost function minimum. This

could be due to symmetries in model formulation created by θ_2 [138]. When θ_2 is fixed the model becomes LSI as demonstrated by unique minimum. Profile likelihood by other parameter combinations demonstrate similar pattern.



(a)



(b)

Figure 4. 4: Two-dimensional profile likelihood result for the PK/PD reference model (a) model is NLSI due to internal symmetries created by θ_2 (b) model is LSI after fixing θ_2.

2) Confidence Interval Test:

Confidence intervals were estimated for PK/PD reference models in absence and presence of 1% noise for one set of model parameters are included in Table (4.6) for The original model formulation has nontrivial unidentifiability as demonstrated in the profile likelihood test. This is confirmed in a larger sample size (N=1000) in the neighborhood of the nominal parameters. The results were consistent in 98.3% and 96.9 % for positive and negative outcome scenarios respectively.

Table 4. 6: Confidence intervals for the original PK/PD model and PK/PD model with θ_2 fixed in absence and presence of noise

	Reparametrized (LSI)		**Original (NLSI)**	
Parameters	Lower CI	Upper CI	Lower CI	Upper CI
θ_1	0.005	0.005	0.005	0.005
θ_2	NA	NA	0.003	0.003
θ_3	0.010	0.010	0.010	0.010
θ_4	0.020	0.020	0.020	0.020
	with noise		**with noise**	
	Lower CI	Upper CI	Lower CI	Upper CI
θ_1	-0.018	0.042	**-38.10**	**38.11**
θ_2	NA	NA	**-15.19**	**15.20**
θ_3	-0.019	0.044	**-22.89**	**22.91**
θ_4	0.024	0.026	0.02	0.02

3) Rank Test:

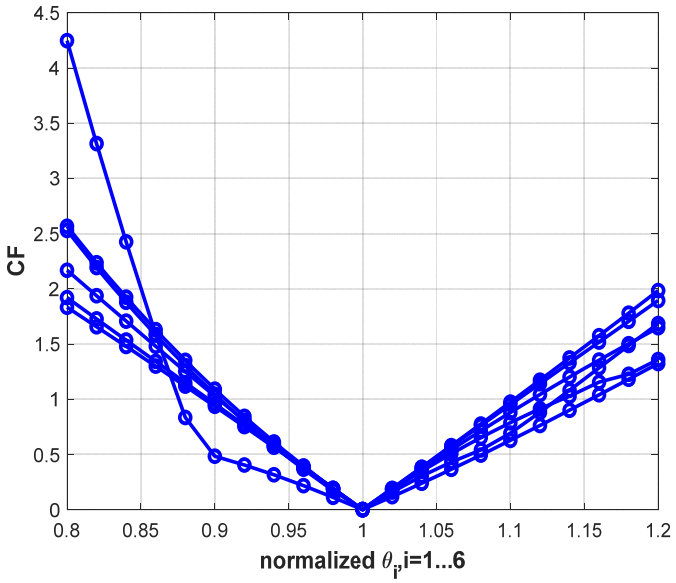
For the original PK/PD model the Jacobian was rank deficient (rank=3) for 99.5% of points, so the test correctly deemed the model to be NLSI. For the PK/PD model with θ_2, the Jacobian was full rank (rank=4) for 98.6% of points, so the test correctly determined the model to be NLSI.

Conclusion of the application of SI test to the reference model

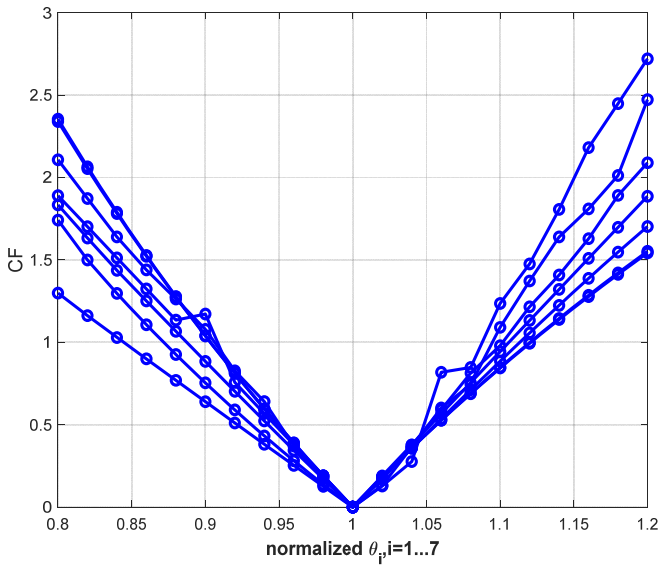
All three sets of results from the profile likelihood, confidence interval and rank test are consistent and meet the prespecified criteria for the test. The combination of these tests can accurately and effectively discern between NLSI and LSI models. The tests can be applied towards the 6p+ models.

Results for 6-8 Parameter Models

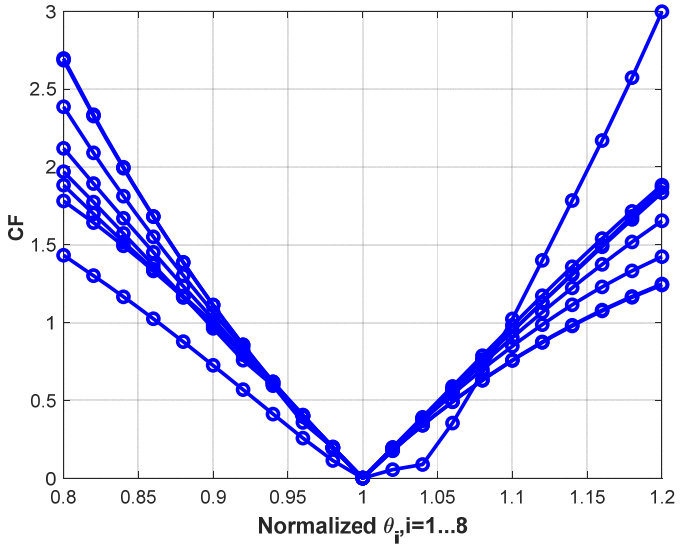
The profile likelihood results for 6p+ models are provided in Figure (4.5). All model parameters for the three sets of models when profiled result in noticeable curvature in the cost function. Therefore, all three models are LSI based on this test.



(a)



(b)



(c)

Figure 4. 5: Profile likelihood results for all parameters of the 6p (a), 7p(b), and 8p(c) models

4.2.3.3 Conclusion of Step 3

Based on the identifiability test developed and the prespecified test acceptance criteria that were verified as part of checking the identifiability properties of the two reference models, all models under evaluation listed in Table (4.1) passed each one of the profile likelihood, rank, and confidence interval tests. The models are therefore LSI and the model parameters may be estimated uniquely.

4.2.4 Step 4: PSA

Since the models have been determined to be LSI, we can proceed to the next step of the framework which is PSA. However, at this juncture we decided to preliminarily only apply the framework to the 4p and 6p models and include other model variants only if the results obtained from these models unambiguously allow progress towards next steps of the framework.

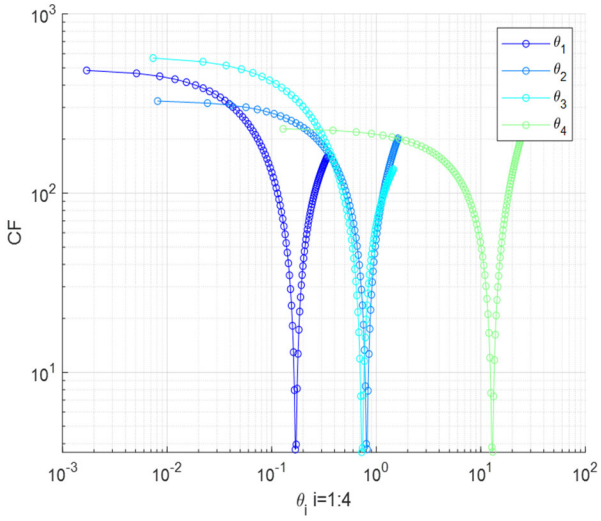
4.2.4.1 Methods

Individual parametric sensitivity for 4p and 6p models were first evaluated by evaluating model cost function in a range of parameters centered around each individual parameter's nominal values. In this method of sensitivity analysis, the influence of variation in each parameter on model output or cost function is evaluated and the influence of parametric interactions is not accounted for [40].

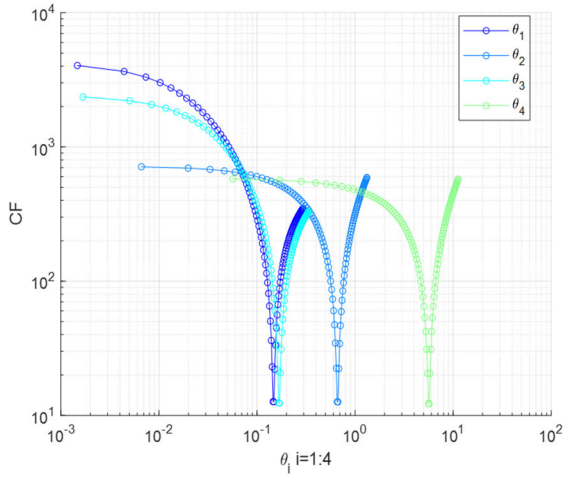
4.2.4.2 Results

The results for application of PSA to the 4p and 6p models are provided in Figure (4.6) a through c. All four parameters in the 4p model demonstrate cost function sensitivity when varied

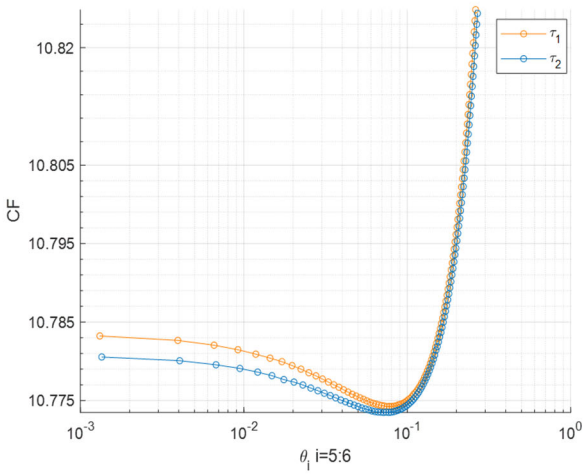
around their nominal values. The 6p model however, is found to have two parameters (the transport delays) that have minimal impact on model output and thus should be fixed prior to model calibration.



(a)



(b)



(c)

Figure 4. 6: One-at-a-time cost function sensitivity to parameters of 4p model (a) and 6p model (b) and (c). Nominal parameter values were used to generate the plots.

4.2.4.3 Conclusion of Step 4

All parameters of the 4p model demonstrate cost function sensitivity. The lags in the 6p model however do not influence the cost function or model output and therefore need to be fixed.

4.2.4.4 Calibration

The process of calibration is a model building step and as such not in the scope of the proposed framework. However, it needs to be carried out in order for the SSPM to proceed to the next step. Therefore, brief details about calibration methods and calibration data are provided in the subsequent sections:

4.2.4.5 Methods

Calibration is a model building step and thus not part of model evaluation activities listed as part of the proposed framework. It involves application of calibration data to the model.

4.2.4.6 Calibration Data

Model output measurements ($PetCO_2$) and input measurements (MV) in Figure (4.7) from 24 anonymous pediatric subjects receiving pressure-controlled mechanical ventilation during a surgical procedure were applied for model calibration. The data collection protocol was approved, as part of a larger investigation, by the Children's and Women's Health Centre of British Columbia. A standard respiratory module (M-CAiOVX, Datex-Ohmeda, Finland) was utilized for data collection. For additional information on data collection and experiment protocol see [77].

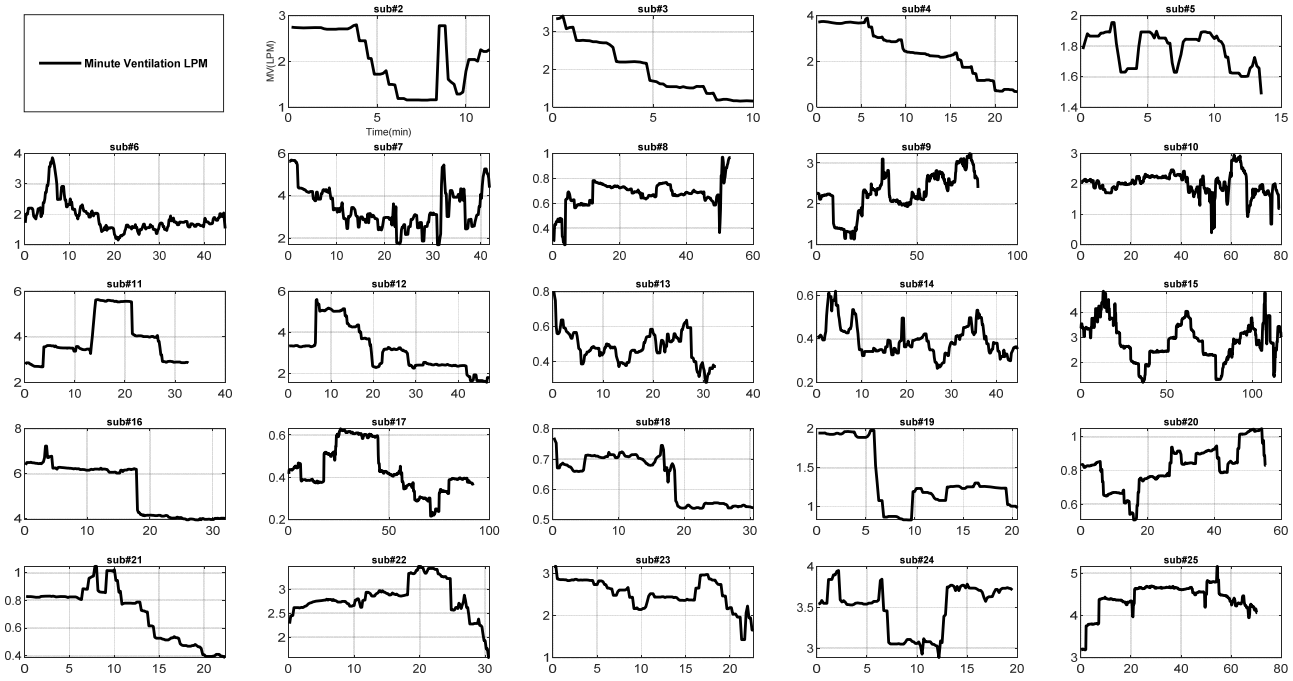


Figure 4. 7: Subject specific MV measurements used as model input for parameter calibration

4.2.4.7 Model Calibration

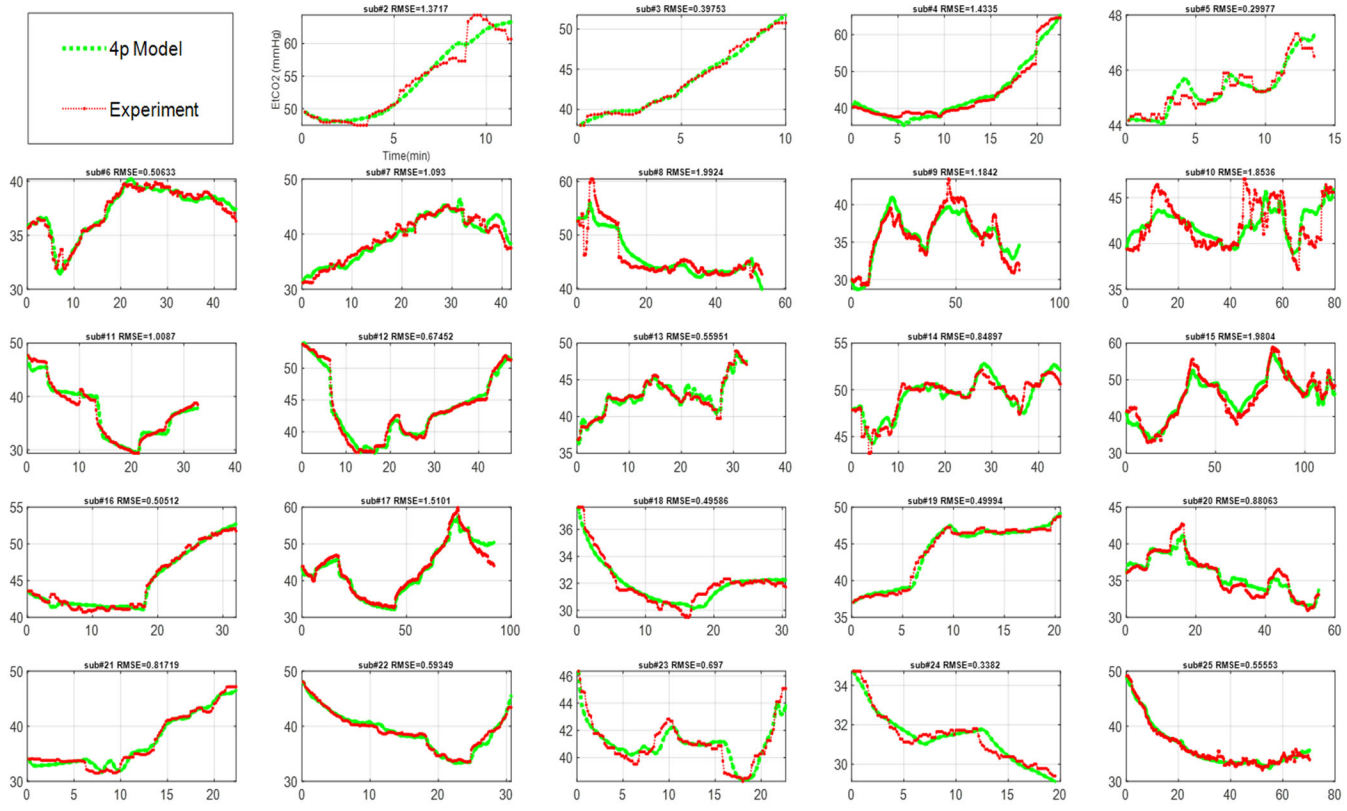
Two classes of optimizers were utilized for model calibration. First a gradient based optimizer employing nonlinear regression based on Levenberg-Marquardt method with a constant error model was utilized. In addition, a second class of optimizer based on non-gradient method employing Nelder-mead simplex direct search method was applied. Both optimizations utilized the same quadratic cost function in Eq. (4.6). In this equation $Y_m(\theta, t)$ is the model output and $Y_e(t)$ is $EtCO_2$ measurements. We selected a prespecified average RMSE Eq. (4.7) of 1.5 mmHg as the threshold for determination of calibration failure. The delay parameters were fixed to 5s and 6s (0.08 and 0.1 min respectively) based on average values reported in Table (4.2). Constraints for other model parameters were implemented based on conservative estimation of physiological bounds.

$$\theta^* = \{\theta_1, \theta_2, \theta_3, \theta_4\} = \arg\min_{\theta} \left\| \left(\mathrm{Y_m}(\theta, \mathrm{t}) - \mathrm{Y_e}(\mathrm{t}) \right) \right\|_2 \tag{4.6}$$

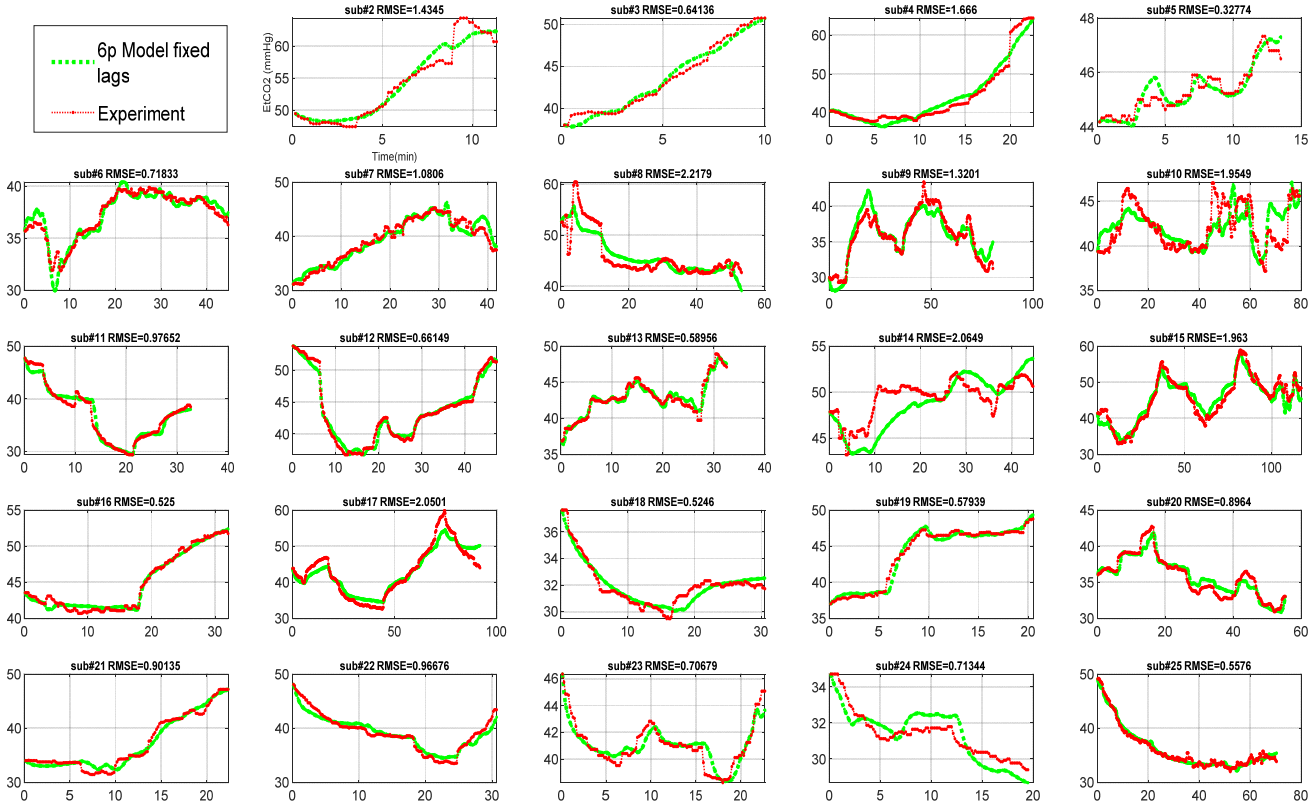
$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{Y}_m(\theta^*, t_i) - \mathrm{Y}_e(t_i))^2}{N}} \tag{4.7}$$

4.2.4.8 Calibration Results

The fit quality results for both classes of the optimizers were similar. The calibration plots are only provided for the gradient-based optimizer and the overall summary results for both optimizers are included in Tables (4.7) and (4.8). Figures (4.8) (a) and (b) depict the subject-specific fit quality for the 4p and 6p models respectively.



(a)



(b)

Figure 4. 8. Subject-specific calibration results for the 4p (a) and 6p model with fixed lags (b).

4.2.5 Step 5: CQE

4.2.5.1 Evaluation of physiological relevance of the parameters

The mean and median of all identified parameters listed in Table (4.7) are comparable to the previously reported nominal values listed in Table (4.2) for 4p and 6p models. Furthermore, Subject-specific calibrated parameters were consistent with physiologically relevant values.

4.2.5.2 Evaluation of Fit Quality

The results for both optimization methods passed the prespecified average RMSE acceptance criterion of 1.5 mmHg for both 4p and 6p models (Tables (4.7) and (4.8)). The mean RMSE for both optimization methods for 4p and 6p models were comparable despite the mean of the actual parameter values being significantly different.

Table 4. 7: Calibration results for the 4p model

4p	gradient based optimization					non gradient based optimization				
	θ_1	θ_2	θ_3	θ_4	RMSE(mmHg)	θ_1	θ_2	θ_3	θ_4	RMSE(mmHg)
mean	0.48	2.31	0.29	4.48	1.03	0.75	1.76	0.23	4.04	0.92
median	0.33	1.41	0.24	3.37	0.77	0.18	0.99	0.21	2.87	0.76
std	0.34	2.31	0.24	3.75	0.56	1.35	1.95	0.20	3.66	0.51
CV	0.71	1.00	0.81	0.84	0.54	1.80	1.11	0.85	0.91	0.55

Table 4. 8: Calibration results for the 6p model

6p with fixed lags	gradient based optimization							non gradient based optimization						
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	RMSE(mmHg)	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	RMSE(mmHg)
mean	0.29	0.87	0.14	2.75	0.2	0.2	1.11	0.96	2.48	0.20	3.89	0.2	0.2	0.90
median	0.18	0.90	0.10	2.41	0.2	0.2	0.90	0.21	0.77	0.18	3.71	0.2	0.2	0.74
std	0.32	0.50	0.11	1.96	NA	NA	0.58	1.59	4.14	0.19	4.35	NA	NA	0.50
CV	1.10	0.58	0.83	0.71	NA	NA	0.53	1.65	1.67	0.99	1.12	NA	NA	0.55

However, evaluation of individual fitting results reveals potential practical unidentifiability and unreliability of both 4p and 6p models. Tables (4.9) and (4.10) includes examples of values of model parameters for both 4p and 6p models which are drastically different despite the models having the same fitting quality as measured by RMSE. This lack of reliability in parameter estimates is particularly problematic because of the COU characterized in Step 1. Having a COU focusing on prediction of physiological parameters requires the parameters to be estimated reliably for the model to fulfill its requirements in the later stages of the framework such as actual validation with independent data.

These results point to unreliability of the estimates and warrant further investigation as part the next step of the framework.

Table 4. 9: Selected subject-specific calibration results for the 4p model

	Non-gradient based					gradient based				
subject #	**θ_1**	**θ_2**	**θ_3**	**θ_4**	**RMSE**	**θ_1**	**θ_2**	**θ_3**	**θ_4**	**RMSE**
5	0.10	0.80	0.08	1.53	0.29	0.19	0.93	0.15	2.73	0.30
6	0.06	0.86	0.07	0.88	0.70	0.08	0.40	0.01	0.17	0.51
7	0.08	2.09	0.31	3.32	1.19	0.08	1.05	0.32	3.44	1.09
8	0.40	0.75	0.62	6.13	1.93	0.31	0.41	0.36	7.21	1.99
9	0.10	0.75	0.07	1.92	1.39	0.11	0.33	0.06	1.83	1.18
10	0.06	1.04	0.25	3.41	1.96	0.08	0.52	0.22	3.01	1.85
11	2.10	6.51	0.13	2.45	1.00	0.49	3.14	0.12	2.22	1.01

Table 4. 10 :Selected subject-specific calibration results for the 6p model

	Non gradient based							gradient based						
subject #	**θ_1**	**θ_2**	**θ_3**	**θ_4**	**$\tau_{1(s)}$**	**$\tau_{2(s)}$**	**RMSE**	**θ_1**	**θ_2**	**θ_3**	**θ_4**	**$\tau_{1(s)}$**	**$\tau_{2(s)}$**	**RMSE**
5	0.23	0.95	0.33	6.76	5.0	6.0	0.31	0.18	0.73	0.14	2.99	5.0	6.0	0.31
6	0.35	1.54	0.07	1.05	5.0	6.0	0.76	0.10	0.42	0.03	0.52	5.0	6.0	0.57
7	0.19	1.97	0.32	4.08	5.0	6.0	1.11	0.09	0.81	0.19	2.85	5.0	6.0	1.11
8	0.64	0.75	0.51	13.09	5.0	6.0	1.97	0.33	0.43	0.12	2.60	5.0	6.0	2.22
9	0.64	1.67	0.07	2.30	5.0	6.0	1.78	0.09	0.26	0.05	1.74	5.0	6.0	1.22
10	0.17	0.87	0.11	1.82	5.0	6.0	1.98	0.11	0.60	0.15	2.39	5.0	6.0	1.90
11	0.18	0.97	0.11	2.51	5.0	6.0	0.96	0.19	1.06	0.08	1.72	5.0	6.0	0.97
12	0.57	5.36	0.21	2.82	5.0	6.0	0.65	0.17	1.49	0.20	2.86	5.0	6.0	0.67

4.2.5.3 Conclusion of Step 5

Both models passed the physiological relevance and fit quality tests. However, the calibration results point to unreliability of the estimates and warrant further investigation as part of the next step of the framework.

4.2.6 Step 6: PIA

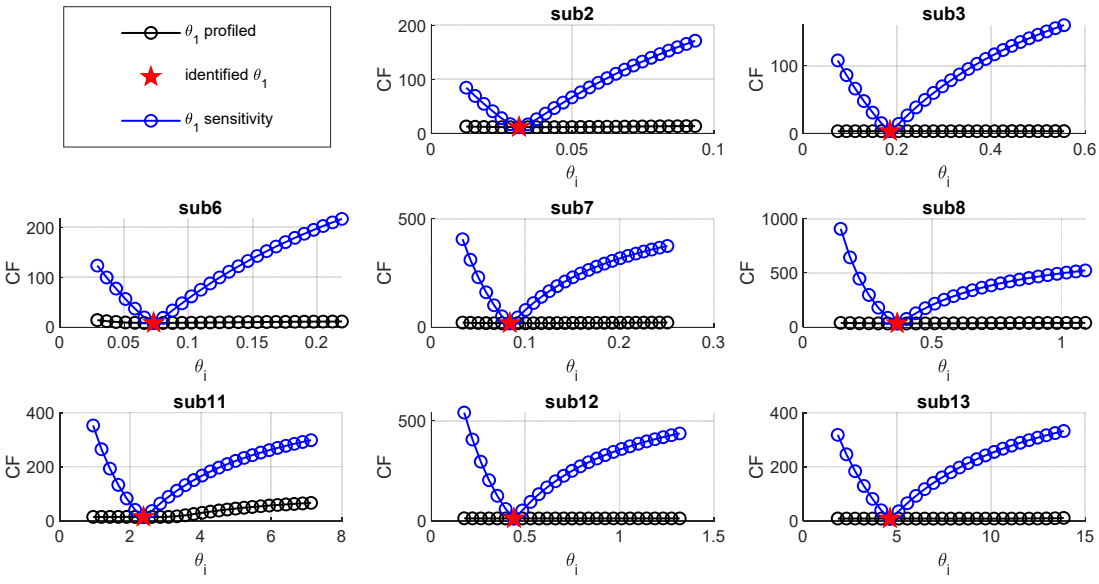
At this juncture, PIA is the most efficient step to take to assess the degree and source of weaknesses in the calibration results observed in the previous step. Having PIA conducted by profile likelihood allows for fault diagnosis and enables the modeler to pinpoint the necessary steps to take in order to ameliorate the potential problems with parameter unreliability. Without PIA the modeler may revert to change the structure of the model with no direction as to what the source of the problem is or go through an iterative process of changing optimizers hoping that parameter reliability will improve and the model can fulfill the requirements set forth by its COU (i.e., acceptable predictive capability of the identified parameters).

4.2.6.1 Methods

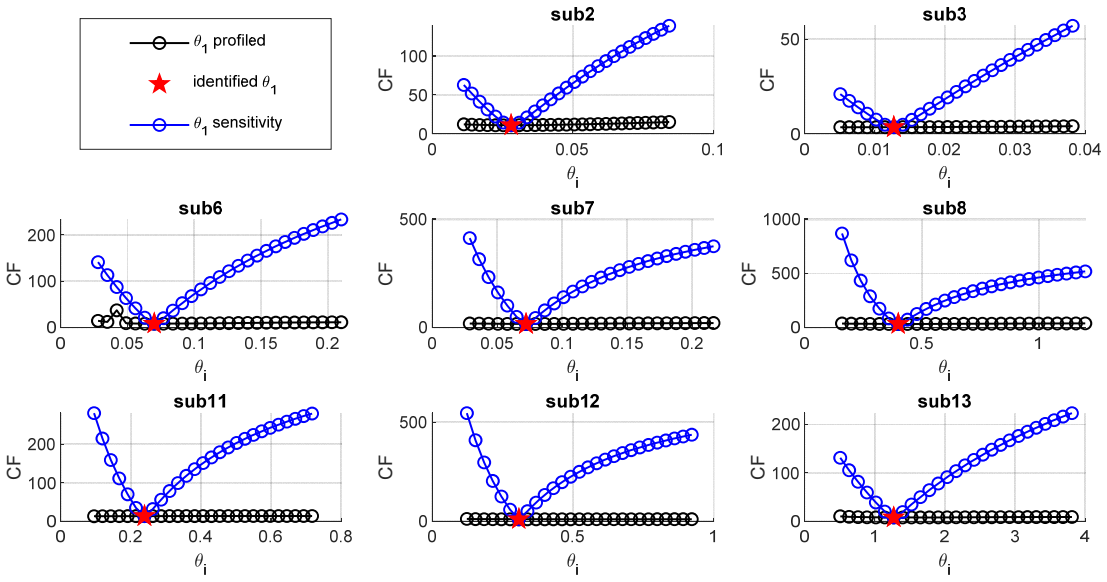
We initiate the PIA by applying the profile likelihood method to the post-calibrated model. The advantage of applying profile likelihood is that when combined with individual parameter sensitivity analysis (Step 4), it allows for identification of the root cause of practical unidentifiability by separating unidentifiability due to cost function insensitivity from unidentifiability due to parametric correlation.

4.2.6.2 Results

Subject-specific profile likelihood results for the 4p and 6p models is provided in Figure (4.9) (a) and (b) respectively for several sample subjects. The profile likelihood plot is compared with the single parameter sensitivity for the same parameter profiled. Further information may be obtained by plotting the non-profiled parameters against the profiled parameter in Figure (4.10). In this figure presence of significant parametric compensation can be observed in pairwise fashion.



(a)



(b)

Figure 4. 9:(a) Profile likelihood results for the 4p model for select subjects when θ1 is profiled. Flat curvature of the cost function for θ1 profile likelihood plot (black) compared to its sensitivity (blue) indicates presence of parametric compensation to keep the cost function relatively unchanged. Confidence intervals are large and not shown due to flatness of the cost function of the profiled parameter. (b) Profile Likelihood results for 6p model for select subjects when θ1 is profiled. Results are similar to the 4p model with the same conclusion that the 6p model is practically unidentifiable due to parametric interaction.

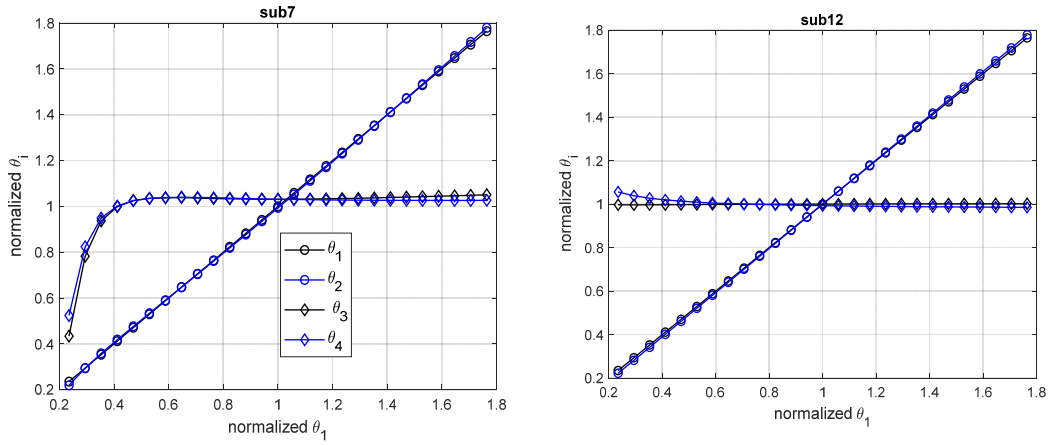


Figure 4. 10:Presence of pairwise parametric compensation for two subjects. The rest of the cohort follow the same pattern.

4.2.6.3 Conclusion of Step 6

The model cannot proceed to the next step since the 4p and 6p models are practically unidentifiable given the data available. The models need to be reparametrized and/or additional data collected to allow for reliable parameter estimates. The information obtained from PIA regarding parameter interactions effectively informs a design of a data collection protocol that would minimize parametric interaction e.g., type or timing of the input may be selected to isolate the parameters and minimize pairwise interactions during model calibration. For example, the modeler may try to amend the protocol to include measurement of cardiac output (Q) and use that as a model input. Alternatively , parameter estimation may be judiciously conducted on episodes of data where Q is relatively unchanging.

4.3 Discussion and conclusion

In this chapter we provided the results for application of the proposed novel framework to different formulations of an SSPM of CO_2 gas exchange intended for evaluation of Clinical Decision Support (CDS) systems. Even though the models could not be taken throughout the entire framework, application of a step-by-step process with firm order proved the utility of the framework in detecting model and/or data weakness and providing a path forward for SSPM model developer to rectify the detected shortcoming with maximal efficiency and without having to prematurely commit to resource intensive steps in the midst of model development.

The class of models was designated in the first step of the framework to have a COU for predicting gas exchange parameters such as lung volume in pediatric patients with diagnostic clinical utility which could be features of a CDS system. This step informed the rigor of subsequent steps such as SIA, CQE and PIA and established key requirements for the model to meet its objective set forth by COU. Application of Step 2 was successful in that it concluded that the model has face validity and has physiological and clinical consistency. Having a COU centered around predictive power of the estimated parameters dictated a comprehensive and rigorous analysis of the model's ability to yield unique parameters theoretically as part of Step 3. This step was undertaken with SIA and determined that the model variants are all LSI and that the structure of the models theoretically allows for unique parameter estimation assuming perfect and maximally informative data. This step effectively cemented the foundation for addressing potential parameter unidentifiability by eliminating weaknesses such as symmetries in model formulation as the source of unidentifiability and effectively guided the modeler towards the root cause of unidentifiability (e.g., quality or quantity of calibration data as determined in Step 6), thereby saving time and resources.

In Step 4 we determined that the 6p+ models may be reduced by fixing the lags to their respective physiological values. This optimized the calibration process by reducing computation costs and eliminating potential compromise in calibration results which could result as a result of the delay parameters interacting or compensating with other model parameters.

Step 5 was conducted successfully with models passing the prespecified acceptance criterion for RMSE and demonstrating physiological relevance of their calibrated parameters. However, the results from this step proved alarming as the parameters identified using different optimization methods produced nearly the same fit quality despite taking significantly different

values. It is exactly for this frequently encountered situation in SSPM models that Steps 1 through 4 of the proposed frameworks provide maximal utility. Following those steps allow the modeler to hypothesize the source of unidentifiability of the model to be the calibration data and not the structure of the model or steps taking as part of the calibration process. The subsequent step which is PIA confirmed this hypothesis and provided definitive evidence of the class of models being practically unidentifiable due to parametric interactions. The framework successfully and efficiently provides the next steps for the SSPM modeler which is to gather more or different type of data (e.g., different timing of the MV input or additional measurement of unobserved inputs such as $PvCO_2$ or Q) to isolate the parameters and allow for reliable estimation of the parameters. Alternatively, the modeler may elect to reparametrize the model and reapply the framework from Step 2.

In the current state of SSPM credibility assessment steps 1 to 4 are often ignored or not explicitly followed in a firm order resulting in chaotic model development and evaluation process focused on calibration quality. This frequently leads to iterative and resource intensive attempts at trying various optimizers or optimization schemes to ameliorate problems identified as part of parameter reliability assessment steps. In conclusion, the framework addresses the existing gaps in SSPM credibility assessment and provides a disciplined process to be integrated with model development process to efficiently evaluate the credibility of SSPMs.

Chapter 5: Applying the Framework Towards the credibility assessment of a Blood Volume Kinetic Model

5.1 Introduction[3]

In this chapter, we aim to evaluate the credibility of a 3-parameter (hereafter referred to as the 3p model) mathematical model of BV kinetics in response to hemorrhagic blood loss and fluid infusion to serve as the second case study for application of the proposed framework. SSPMs for prediction of physiological variables such as blood volume, cardiac output, and blood pressure in response to hemorrhage have been previously developed using sheep subjects [161], [162]. The purpose of the SSPM is to replace or complement existing animal studies such that pre-clinical assessment of SAS systems can be less cost prohibitive and more streamlined. Therefore, initiation of human trials which rely on such preclinical safety evidence can be expedited. To achieve such a goal, the credibility of the mathematical model must be established for the purpose of pre-clinical *in silico* testing. In this chapter, we apply the proposed framework to a previously developed SSPM of blood volume kinetics [159]. This SSPM is a physics-based two compartmental model of blood volume and interstitial fluid volume in Figure (5.1). The model accepts fluid infusion (u) and hemorrhage (v) as inputs. Model parameters include α_u and α_v representing the ratio between the intravascular and extravascular volumetric changes in the steady state in response to fluid infusion and fluid loss respectively. K_p represent the proportional gain determining the rate of change of BV. Additional details about model derivation and assumptions may be found in [159].

[3] B. Parvinian, R. Bighamian, C. G. Scully, J. O. Hahn, and P. Pathmanathan, "Credibility Assessment of a Subject-Specific Mathematical Model of Blood Volume Kinetics for Prediction of Physiological Response to Hemorrhagic Shock and Fluid Resuscitation," Front. Physiol., vol. 12, no. September, pp. 1–14, 2021.

Infusion, hemorrhage and blood volume recordings from 22 sheep were used to generate the SSPMs (further details below).

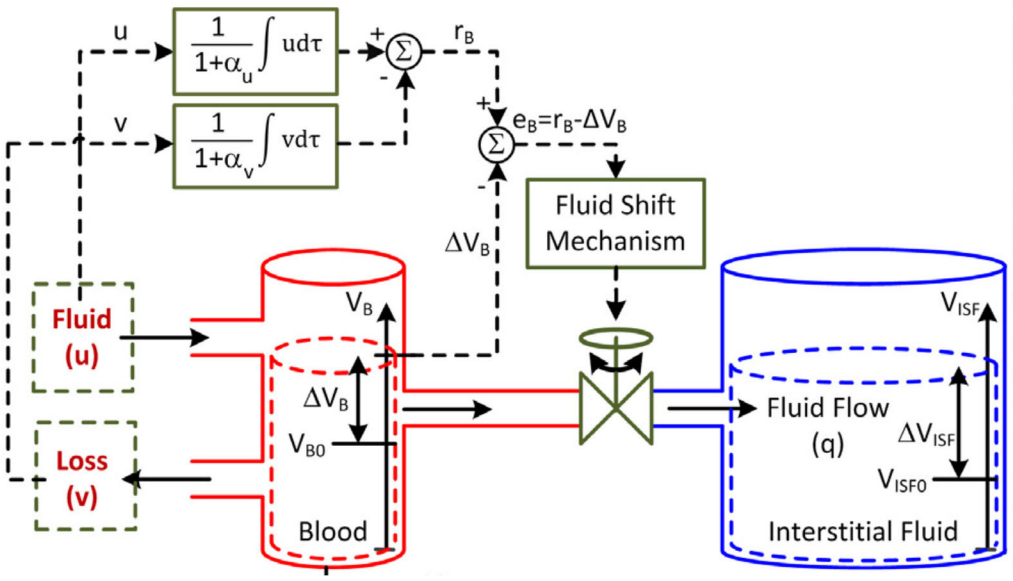


Figure 5. 1: Schematic of the two-compartment body fluid volume recreated from [162]

5.2 Application of the Framework to the BV Kinetic Model

5.2.1 Step 1: Characterization of COU and Model Formulation

5.2.1.1 COU

The COU for this SSPM is to predict changes in blood volume in response to individual as well as overlapping episodes of hemorrhage and fluid resuscitation in sheep subjects. Because of this COU, subject-specific validation is important due to variability in each subject's response.

A priori and explicit characterization of COU for this SSPM has important implications on the type and rigor of the ensuing credibility activities. For example, because the model will be used to predict the time series of the changes in blood volume, it will be essential to identify a calibration window (i.e., by splitting the subject specific time series data) such that adequate subject-specific data independent of calibration data will remain for the purpose of model validation.

5.2.1.2 Model Formulation

The model is formulated as shown in Eq. (5.1):

$$\Delta\ddot{V}_B(t) + K_p\Delta\dot{V}_B(t) = \left[\dot{U}(t) - \dot{V}(t)\right] + \frac{K_p}{1+\alpha_u}U(t) - \frac{K_p}{1+\alpha_v}V(t) \tag{5.1}$$

where ΔV_B is the change in BV, $U(t)$ and $V(t)$ are fluid infusion and hemorrhage rates, respectively, and the parameters $\alpha_u, \alpha_v,$ and K_p are the ratio of volume gain between intravascular and interstitial compartments, the ratio of volume loss between intravascular and interstitial compartments, and the rate of fluid shift between intravascular and interstitial compartments, respectively. This mathematical model was previously published in 3-parameter and 4-parameter variants [162] . In this work, we considered the 3-parameter BV kinetics model.

5.2.2 Step 2: Face Validity

We identified nominal values for the parameters of the 3p model and ran a number of simulations to verify model face validity. Figure (5.2) depicts an example scenario considered to evaluate face validity of the model. In this figure, we first generated a nominal infusion input for the model. We then assessed the simulation output (dynamic and steady state change in BV) in response to fluid infusion only. The scenario was repeated for changes in model parameters such as K_p to verify that such a change makes physiological sense. As shown in this figure, the change in BV is positive in response to fluid infusion. Moreover, the reduction in BV in the absence of infusion (or presence of minimal infusion, e.g., between 80 to 100 min) is qualitatively consistent with physiological mechanisms for fluid shift between intravascular and extravascular

compartments reported in literature [163]. Finally, we verified that an increase in the parameter representing fluid shift K_p drives blood away from the intravascular compartment resulting in lower BV in the intravascular compartment at steady state and during dynamic changes induced by fluid infusion. Changes in other model parameters in presence of both infusion and hemorrhage input produced similar face validity results.

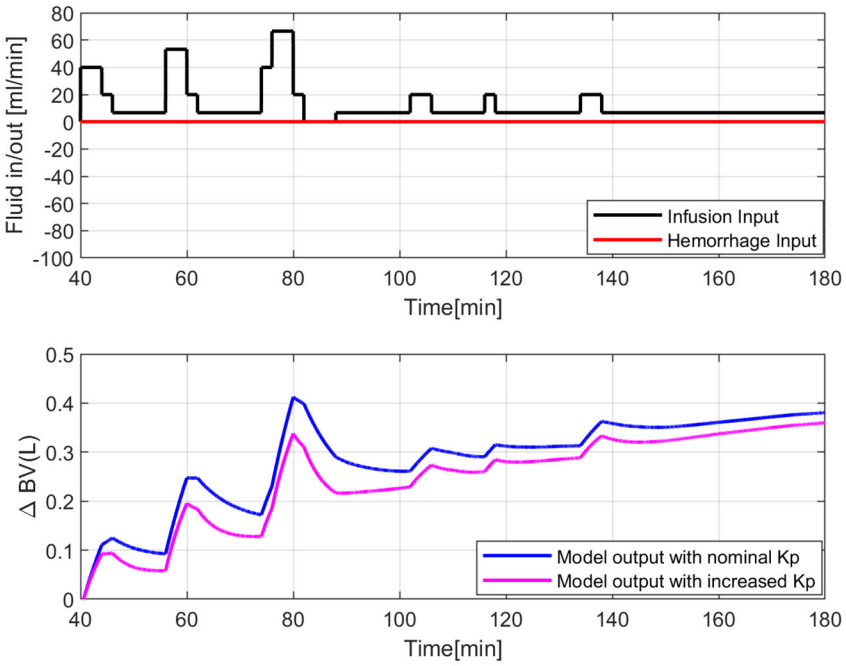


Figure 5. 2: An example scenario for verification of face validity of the BV kinetic model

5.2.3 Step 3: SIA

In order to conduct SIA for the BV kinetic model, we leverage the Laplace transform approach [129] by virtue of the linearity of the BV kinetics model and the relatively small number of parameters involved therein.

Methods and Results

Here we prove that the 3p blood volume model given in Eq. (5.1) is GSI. The general problem formulation for determination of GSI in Eq. (3.1).

Consider the following general state space formulation for a linear time-invariant system:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

$$\mathbf{Y} = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$$

with initial condition $x(t_0) = x_0$, where x are the state variables, Y are the measured quantities, and **u** are the model inputs. Converting the three parameter BV kinetic model Eq. (5.1) to state space form leads to the following state, input, output, and direct transmission matrices (see [162]):

$$A = \begin{bmatrix} -K_p & 0 & \frac{K_p}{(1+\alpha_u)} & \frac{-K_p}{(1+\alpha_v)} \\ -1 & 0 & \frac{1}{(1+\alpha_u)} & \frac{-1}{(1+\alpha_v)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C = [1 \quad 0 \quad 0 \quad 0], D = 0$$

with **u** = [u,v].

The transfer function expression of the state space BV kinetic model can be computed using the following

$$M(s) = C(sI - A)^{-1}B + D$$

Substitution of state, input, and output matrices of the BV kinetic model with block-wise inversion leads to:

$$(sI - A)^{-1} = \begin{bmatrix} E^{-1} & \frac{-1}{s}E^{-1}F \\ 0 & \frac{1}{s}I \end{bmatrix}$$

with the following block assignments:

$$\mathrm{E} = \begin{bmatrix} \mathrm{s} + \mathrm{K_p} & 0 \\ 1 & \mathrm{s} \end{bmatrix}, F = \begin{bmatrix} \frac{-\mathrm{K_p}}{(1+\alpha_u)} & \frac{\mathrm{K_p}}{(1+\alpha_v)} \\ \frac{-1}{(1+\alpha_u)} & \frac{1}{(1+\alpha_v)} \end{bmatrix}.$$

Computing each block:

$$E^{-1} = \frac{1}{s(s+K_P)}\begin{bmatrix} s & 0 \\ -1 & s+K_P \end{bmatrix}$$

$$\frac{-1}{s}E^{-1}F = \frac{-1}{s^2(s+K_P)}\begin{bmatrix} s & 0 \\ -1 & s+K_P \end{bmatrix}\begin{bmatrix} \frac{-\mathrm{K_p}}{(1+\alpha_u)} & \frac{\mathrm{K_p}}{(1+\alpha_v)} \\ \frac{-1}{(1+\alpha_u)} & \frac{1}{(1+\alpha_v)} \end{bmatrix}$$

$$= \frac{1}{s^2(s+K_P)}\begin{bmatrix} \frac{\mathrm{sK_P}}{(1+\alpha_u)} & -\frac{\mathrm{sK_P}}{(1+\alpha_v)} \\ \frac{-\mathrm{K_P}}{(1+\alpha_u)} + \frac{s+K_P}{(1+\alpha_u)} & \frac{\mathrm{K_P}}{(1+\alpha_v)} - \frac{s+K_P}{(1+\alpha_v)} \end{bmatrix},$$

we obtain

$$(sI - A)^{-1} = \frac{1}{s(s+K_P)}\begin{bmatrix} \mathrm{s} & 0 & \frac{\mathrm{K_p}}{(1+\alpha_u)} & \frac{-\mathrm{K_p}}{(1+\alpha_v)} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Therefore, the transfer function representation of $\mathcal{M}$ is

$$\mathcal{M}(\theta) = \mathrm{C(sI - A)^{-1}B} = \frac{1}{s(s+K_P)}\left[s + \frac{K_P}{(1+\alpha_u)} \quad -\mathrm{s} - \frac{K_P}{(1+\alpha_v)}\right]$$

For $\mathcal{M}$ to be GSI we need to show Eq. (3.1) (repeated below) holds.

$$\mathcal{M}(\theta) = \mathcal{M}(\tilde{\theta}) \Rightarrow \theta = \tilde{\theta}$$

Since entries of $\mathcal{M}$ are in canonical form, for the first entry we have:

$$\frac{s + \left(\frac{K_p}{1 + \alpha_u}\right)}{s^2 + sK_p} = \frac{s + \left(\frac{\widetilde{K_p}}{1 + \widetilde{\alpha_u}}\right)}{s^2 + s\widetilde{K_p}}$$

$$\Rightarrow K_p = \widetilde{K_p} \text{ and } \alpha_u = \widetilde{\alpha_u}$$

and similarly from the second entry:

$$\alpha_v = \widetilde{\alpha_v}$$

as required.

5.2.4 Step 4: PSA

Because the BV kinetic model is intended for prediction of blood volume response during fluid resuscitation and hemorrhage, we leverage the PSA step of the framework to satisfy the requirements consistent with this COU. In doing so we tailored the PSA step towards two goals: 1) identification of any potentially insensitive model parameters and 2) quantification of a calibration window during which the model shows adequate sensitivity to the parameters. The second goal is solely designed to enable model predictive capability assessment and model validation on independent data in the final stages of the framework, and therefore requires consideration of the specific data available.

5.2.4.1 Experimental Protocol

The experimental data used in this work were collected retrospectively from 22 conscious sheep undergoing intravenous blood loss and fluid infusion. The experimental protocol was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Texas Medical Branch at Galveston, with adherence to National Institutes of Health guidelines for care and use of laboratory animals [164].The measurements included the rates of hemorrhage and fluid infusion, and BV in Figure (5.3). . All 22 animals received lactated Ringers (LR) solution. The duration of study for each animal was 180 min. After the baseline data were recorded, an initial hemorrhage (25 mL/kg) was performed over 15 min. Fluid infusion was started 15 min after the end of the hemorrhage and continued for 150 min.

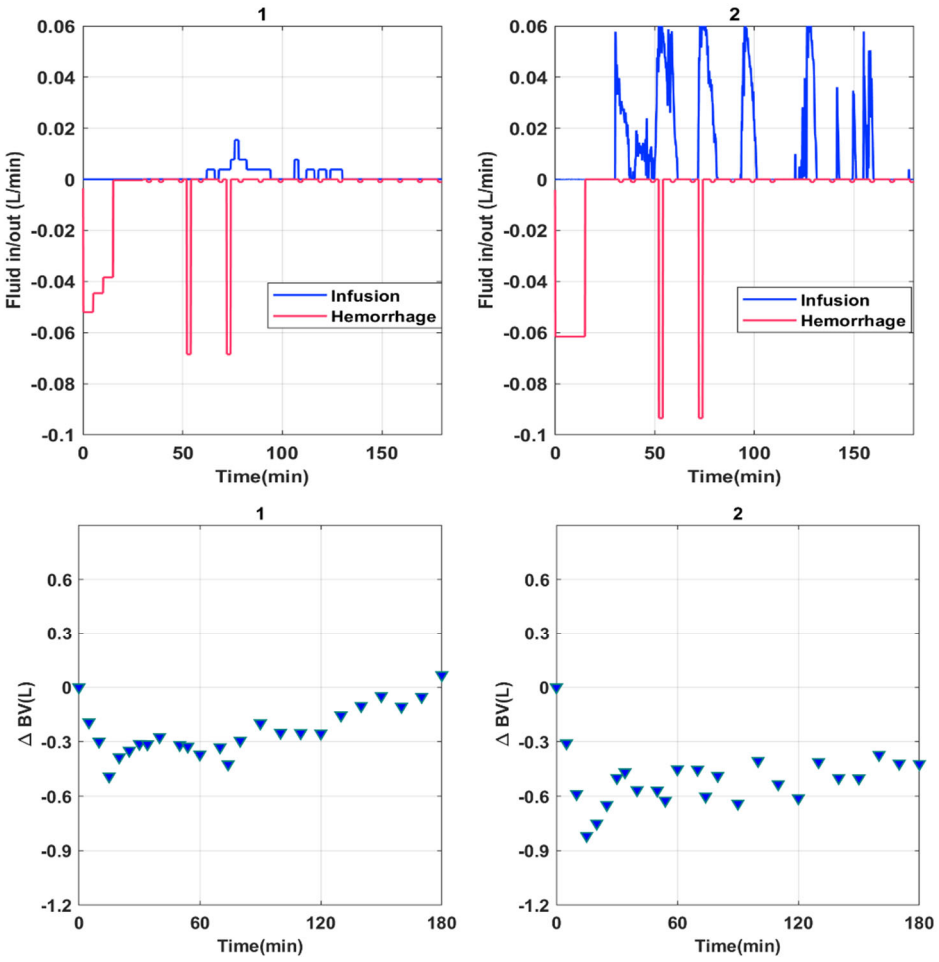


Figure 5. 3: Model fluid input (a) and change in blood volume (b) for subjects 1 and 2. Urine output was negligible compared to hemorrhage and fluid infusion and was not included as a model input for data fitting.

Second and third hemorrhage (5 mL/kg) were performed 50 and 70 min after the start of the initial hemorrhage, each of which lasted for 5 min. While hemorrhage protocol remained constant across the subjects, fluid infusion was varied based on predetermined rates and algorithms as described in [164]. In each animal, baseline BV was measured via indocyanine green dye (ICG). Hematocrit, the ratio between the red blood cell volume (RBCV) and BV, was measured before

and throughout the experiment at 5 to 10 min intervals and was used to measure the fractional change in BV. For more details about the experimental protocol refer to [164].

5.2.4.2 PSA Method

We used the singular value decomposition (SVD) method to evaluate the sensitivity of parameters [131]. The first step allowing for SVD to be leveraged was to convert the mathematical model to linear regression form as in Eq. (5.2):

$$\Delta\ddot{V}_B(t) - \left[\dot{U}(t) - \dot{V}(t)\right] = \begin{bmatrix} \frac{K_p}{1+\alpha_u} & \frac{K_p}{1+\alpha_v} & K_p \end{bmatrix} \begin{bmatrix} U(t) \\ -V(t) \\ -\Delta\dot{V}_B(t) \end{bmatrix} \tag{5.2}$$

For each subject, the regressor matrix (also the same as the sensitivity matrix discussed in chapter 3 because the model is linear) is a matrix with rows corresponding to experimental measurements for $U(t)$, $V(t)$ and $\Delta\dot{V}_B(t)$at N distinct time instants, respectively. SVD was applied to the regressor matrix for each subject to compute the three singular values of the regressor matrix and the corresponding principal directions as determined by the right singular vector. If the SVD results demonstrate substantially small singular values in the direction of certain parameter(s) relative to the others (i.e., the regressor matrix is not of full rank and the estimation problem is ill conditioned), consistently across all animals, this indicates unidentifiability or low identifiability of the associated parameter(s). Those parameters must then be set as physiologically relevant constant values. In this case, we would reapply the SVD to the new mathematical model in order to evaluate the data quality for calibration of the remaining parameters.

Next, we proactively accounted for the COU of the model being prediction of change in BV in response to hemorrhage and fluid resuscitation and the fact that this type of model use will require independent data not used in the calibration process for validation purposes [2], [151] . The experimental data was split into calibration and validation datasets based on the results of

SVD analysis of the regressor matrix. Specifically, a time T_c was identified based on the magnitude of the singular values and their associated principal direction as determined by the right singular vector, and then data for the time interval [0 min, T_c] was used for calibration, with the remaining experimental data [T_c, 180 min] saved for model validation. It is also important to select T_c with considerations of overall experimental protocol to enable independent model evaluation under inputs and conditions that were not included to the calibration process.

5.2.4.3 PSA results

Results of the SVD analysis are provided in Figure (5.4) (a), (b), and (c) which correspond to the three singular values. Consistently across all subjects, the singular values' principal directions were aligned with the axes in parameter space; the first (largest) singular value corresponded to α_v, the second to K_p, and the third to α_u. The singular values are of the same order of magnitude indicating model's output sensitivity to the changes in all three parameters. As such the order of the model is appropriate and should not be reduced at this step.

The x-axis of the plots represents the amount of experimental data used in constructing the regressor matrix: t=180 min means that the regressor matrix used all the experimental data, t=50 min means that the regressor matrix used only the data up to 50 min. Therefore, increasing data is used as t increases, which in turn increases the singular values. The results reveal the general order of identifiability to be $\alpha_v > \alpha_u$, across all subjects. This is not to say which parameters are identifiable but rather their relative identifiability. Based on these results, we chose T_C=50 min, that is, the calibration window as 0 min to 50 min, because this was the smallest time window containing both hemorrhage and fluid infusion and having relatively large singular values for majority of subjects in the study. The motivation for choosing the smallest possible time window for calibration is to leave the largest possible time window (50 min to 180 min) for model

validation. An alternative choice of calibration window of 0 min to 80 min could have been made based on the fact that singular values do not increase substantially after 80 min. However, with this choice there would be no hemorrhage in the validation window (see Figure 5.4).

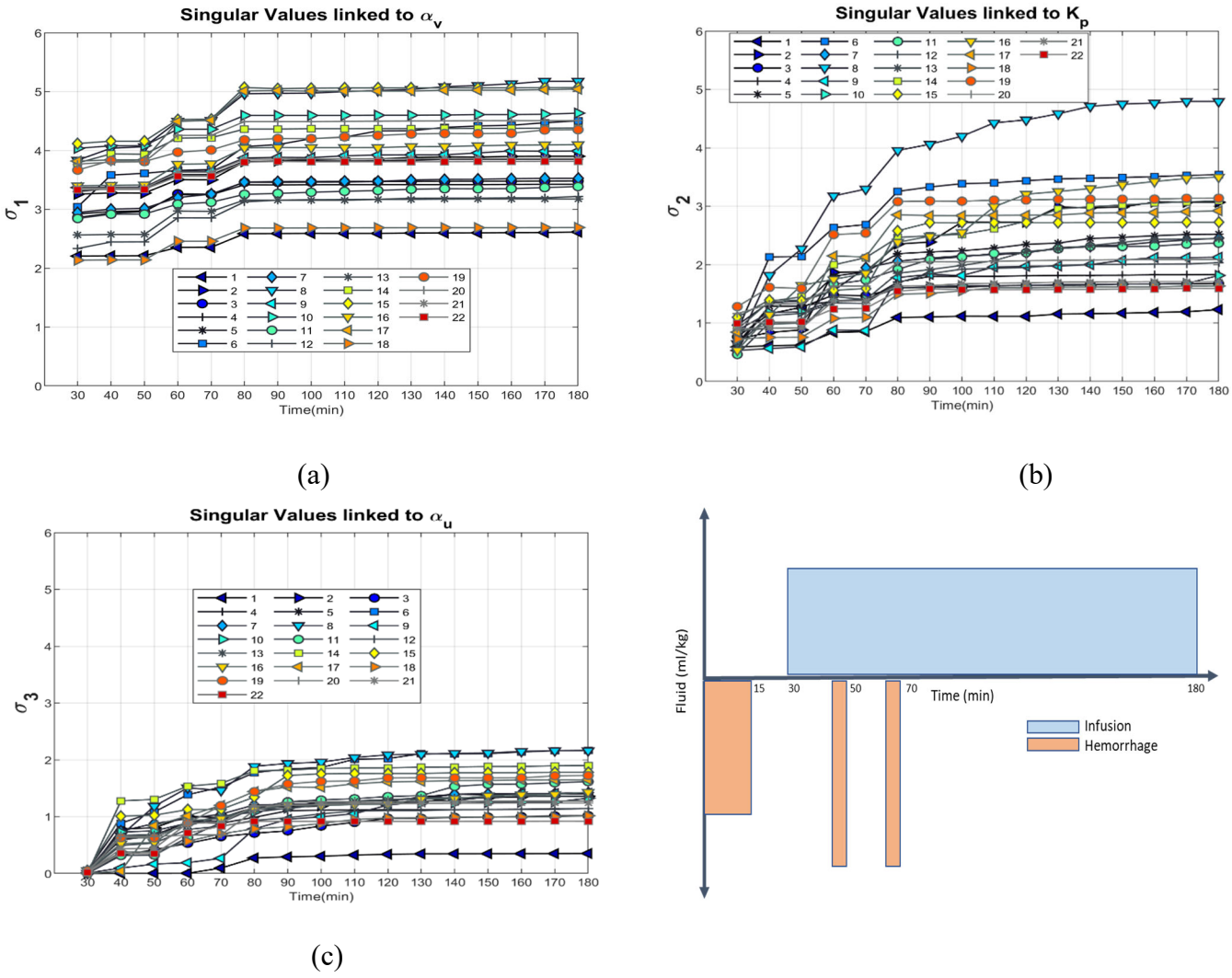


Figure 5. 4: Entire cohort singular values for a) α_v b) α_u c) K_p. d) Experimental protocol used for data collection depicting timing of fluid hemorrhage and infusion. X-axis relates to the amount of experimental data under consideration, for example T_c=50 min corresponds to considering all data between 0 and 50 mins. The calibration window of 0 to 50min was selected because it was the largest time window that would allow for evaluation of model prediction in an independent data segment (i.e., 50 to 180 min) containing both infusion and hemorrhage while still having relatively large singular values for majority of subjects in the study. Overall relative order of identifiability is observed to be $\alpha_v > \alpha_u$.

5.2.4.4 Calibration

Next, model calibration was performed, using only the data in the range [0 min, T_c min] for each animal. The Levenberg-Marquardt algorithm was used to solve the nonlinear least squares optimization problem in Eq. (5.3) with a proportional error model to account for the uncertainty associated with the BV measurements [165]:

$$\theta^* = \{\alpha_u^*, \alpha_v^*, K_p^*\} = \arg\min_{\theta} \left\| \left(\frac{\Delta V_{B,m}(\theta,t) - \Delta V_{B,e}(t)}{|\Delta V_{B,m}(\theta,t)|} \right) \right\|_2 \tag{5.3}$$

where $\Delta V_{B,m}(\theta, t)$ is the model predicted change in BV at time t using parameters and $\Delta V_{B,e}(t)$ is the experimentally measured change in BV at time t. The quality of fit was evaluated using the root-mean-squared normalized error (RMSNE) as defined by:

$$\text{RMSNE} = \left(|\overline{\Delta V_{B,e}(t_i)}| \right)^{-1} \sqrt{\frac{\sum_{i=1}^{N} (\Delta V_{B,m}(t_i) - \Delta V_{B,e}(t_i))^2}{N}} \tag{5.4}$$

5.2.5 Step 5: CQE

5.2.5.1 Parameter physiological relevance

Table (5.1) lists the values of the identified parameters after calibration to the 0 min to 50 min data, and the corresponding RMSNE. For all subjects except Subjects 3, 13, and 16, α_u was not identifiable based on calibration results. This was confirmed in the subsequent PIA stage.

5.2.5.2 Fit Quality Evaluation

The average fit quality for the 3p model was RMSNE=12.6% which passed the prespecified average acceptance criterion of 20% despite α_u being unidentifiable.

Table 5. 1:Calibration Results for 3p model. Inf = infinity, that is, the solver failed to converge for this parameter. α_u was found to be unidentifiable for all parameters except for subject 3, 13, and 16. However, other parameters in these subjects were calibrated to values beyond physiological range.

Subject #	α_u[.]	α_v[.]	K_p min^{-1}	RMNSE%
1	inf	1.39	0.14	9.74
2	inf	0.53	0.11	11.26
3	-0.43	inf	0.04	24.59
4	inf	0.86	0.13	3.93
5	inf	0.92	0.10	4.87
6	inf	2.40	0.11	25.14
7	inf	4.69	0.12	29.91
8	inf	1.71	0.07	14.55
9	inf	0.69	0.18	6.20
10	inf	0.50	0.06	10.21
11	inf	1.89	0.10	11.52
12	inf	3.14	0.06	12.93
13	0.30	1.11	0.67	22.65
14	inf	0.41	0.15	11.55
15	inf	1.08	0.11	10.00
16	2.32	0.57	0.70	5.42
17	inf	1.11	0.11	7.31
18	inf	1.97	0.17	12.15
19	inf	2.21	0.08	17.75
20	inf	1.66	0.10	9.82
21	inf	1.09	0.25	9.21
22	inf	0.96	0.12	7.33

5.2.6 Step 6: PIA

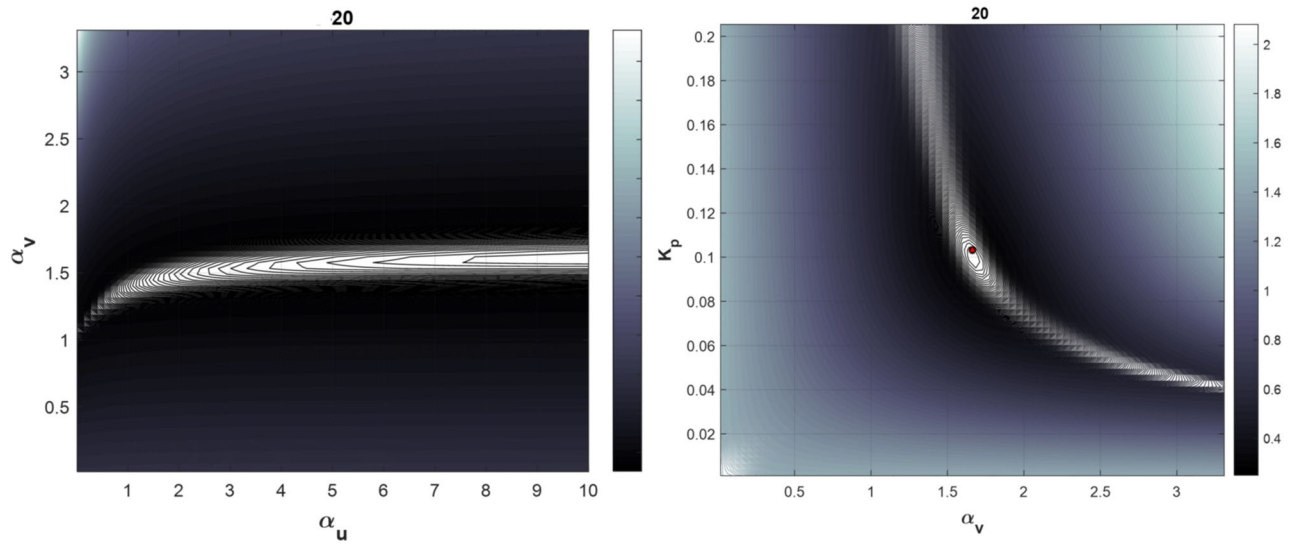
5.2.6.1 Methods

Reliability of the parameters should be evaluated after parameters have been estimated. In this step, we first visualized the cost function contours for each subject, in parameter space, and then calculated the asymptotic 95% confidence region of the calibrated parameters using the parameter covariance matrix computed during the calibration stage [129]. This enables visual determination of correlation between parameters and the potential associated impact on parameter identifiability. If the contour region is unbounded in direction of one particular parameter resulting in unbounded confidence region, this indicates that that parameter was practically unidentifiable [137], in which case we fix it to a physiologically reasonable value and repeat everything from Step 5.

5.2.6.2 Results

PIA demonstrated an unbounded contour in the direction of α_u (see Figure (5.5) (a) for example with Subject 20) and bounded regions containing identified values of α_v and K_p (see Figure (5.5) (b)). For these subjects, the error function continues to decrease as α_u increases and no global minimum exists. For Subjects 3, 13, and 16, a finite value of α_u was found but α_v was found to be unidentifiable or K_p took values significantly beyond cohort average. Since α_u was unidentifiable for nearly all subjects, we set it to be 3 based on previous studies [166]. The new model, with two parameters to be calibrated K_p and α_v, will be referred to as the 2p model. The calibration process was then repeated for the 2p model; results are provided in Table (5.2). With the 2p model, calibrated α_v and K_p were within reasonable physiological ranges and with physiological relevance for all subjects except Subjects 3 and 16. These subjects were considered

subjects for which the mathematical model fails (see discussion in Section 5.3) and were excluded from further analysis. The fitted mathematical models closely matched the experimental data, as seen in Figure (5.6) (see RMSNE column in Table (5.2) for quantitative measure of fit).



(a) (b)

Figure 5. 5: Cost function visualization for the Subject 20 for the 3p model. (a) Cost function as a function of α_u and $\alpha_v$at fixed K_p. The direction of unbounded ellipses representing cost function contours demonstrates unidentifiability of α_u (b) Cost function as a function of K_p and α_v at fixed α_u values.

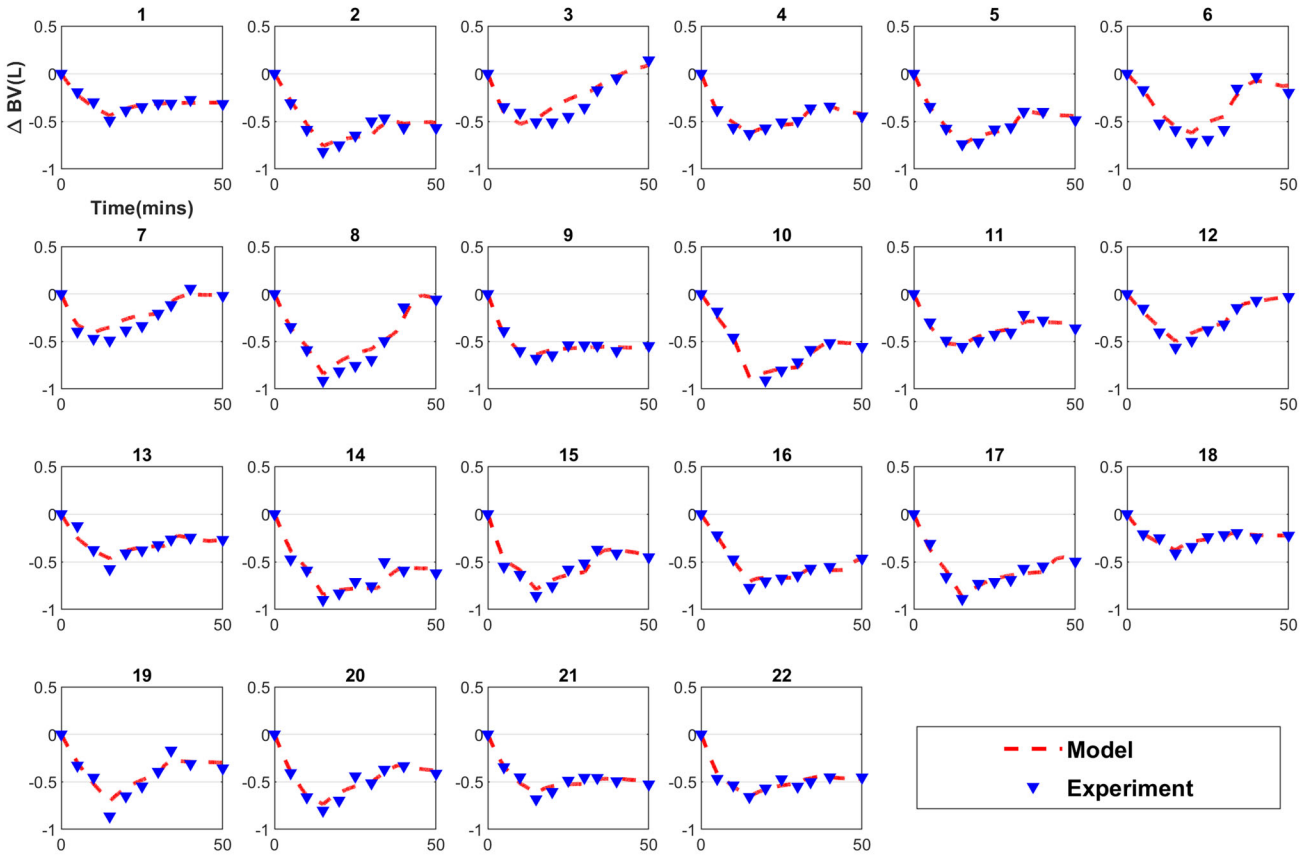


Figure 5. 6: Calibration results for the2p model.

5.2.7 Step 7: UQ

5.2.7.1 Model UQ

We accounted for the uncertainty in the calibrated model parameters by assuming the calibrated parameters for any subject were multivariate normally distributed with covariance matrix as calculated in Step 6. We propagated this uncertainty through the mathematical model when running the validation simulations (Monte Carlo with 10,000 samples). In subjects whose extremes of confidence regions resulted in physiologically implausible parameter values (e.g., $K_p < 0$), sampling resulted in a small number of samples outside the plausible region, which were

excluded. The 95% confidence regions for the calibrated model parameters are plotted in Figure (5.7). The corresponding 95% confidence intervals for each model parameter are listed in Table (5.2). The uncertainty in each subject's identified parameter values varies considerably between the subjects; it is very small for many animals, but it is considerable for others (e.g., Subject 7). However, the resultant uncertainty in model predictions of BV within the validation time window (50 mins to 180 mins) was within acceptable physiological bounds (see later Figure (5.8)), allowing us to progress to experiment UQ and then validation.

Table 5. 2: Uncertainty quantification of the parameters for the 2p model after fixing α_u to 3 . The high and low values represent 95% confidence intervals for the calibrated parameters. Subject 3 and 16 continued to have parameter values beyond reasonable physiological range.

Subject#	α_u[.]	α_v[.]	Kp min^{-1}	RMNSE%	CI α_v low	CI α_v high	CI Kp low	CI Kp high
1	3.00	1.40	0.14	9.73	1.05	1.75	0.07	0.22
2	3.00	0.47	0.13	12.42	0.31	0.64	0.00	0.27
3	3.00	inf	0.06	33.93	inf	inf	0.01	0.11
4	3.00	0.75	0.16	5.50	0.68	0.82	0.12	0.21
5	3.00	0.81	0.13	6.00	0.71	0.91	0.08	0.17
6	3.00	1.81	0.13	25.37	1.30	2.32	0.03	0.23
7	3.00	3.20	0.16	30.78	1.52	4.89	0.00	0.32
8	3.00	1.35	0.09	15.87	1.11	1.59	0.06	0.13
9	3.00	0.68	0.19	6.27	0.56	0.81	0.07	0.31
10	3.00	0.43	0.07	10.61	0.23	0.62	0.00	0.15
11	3.00	1.76	0.11	11.71	1.17	2.35	0.05	0.18
12	3.00	2.29	0.07	15.91	2.15	2.44	0.06	0.09
13	3.00	1.44	0.18	18.41	0.99	1.90	0.02	0.34
14	3.00	0.35	0.21	13.13	0.19	0.51	0.00	0.47
15	3.00	0.94	0.12	11.31	0.75	1.14	0.06	0.19
16	3.00	0.59	0.59	5.49	0.53	0.65	0.30	0.88
17	3.00	1.05	0.12	7.80	0.81	1.30	0.06	0.18
18	3.00	1.85	0.19	12.47	1.50	2.21	0.10	0.27
19	3.00	2.10	0.10	19.12	1.04	3.16	0.02	0.17
20	3.00	1.51	0.11	10.57	1.22	1.80	0.07	0.16
21	3.00	1.03	0.27	9.86	0.86	1.21	0.10	0.44
22	3.00	0.91	0.13	7.12	0.73	1.10	0.06	0.20

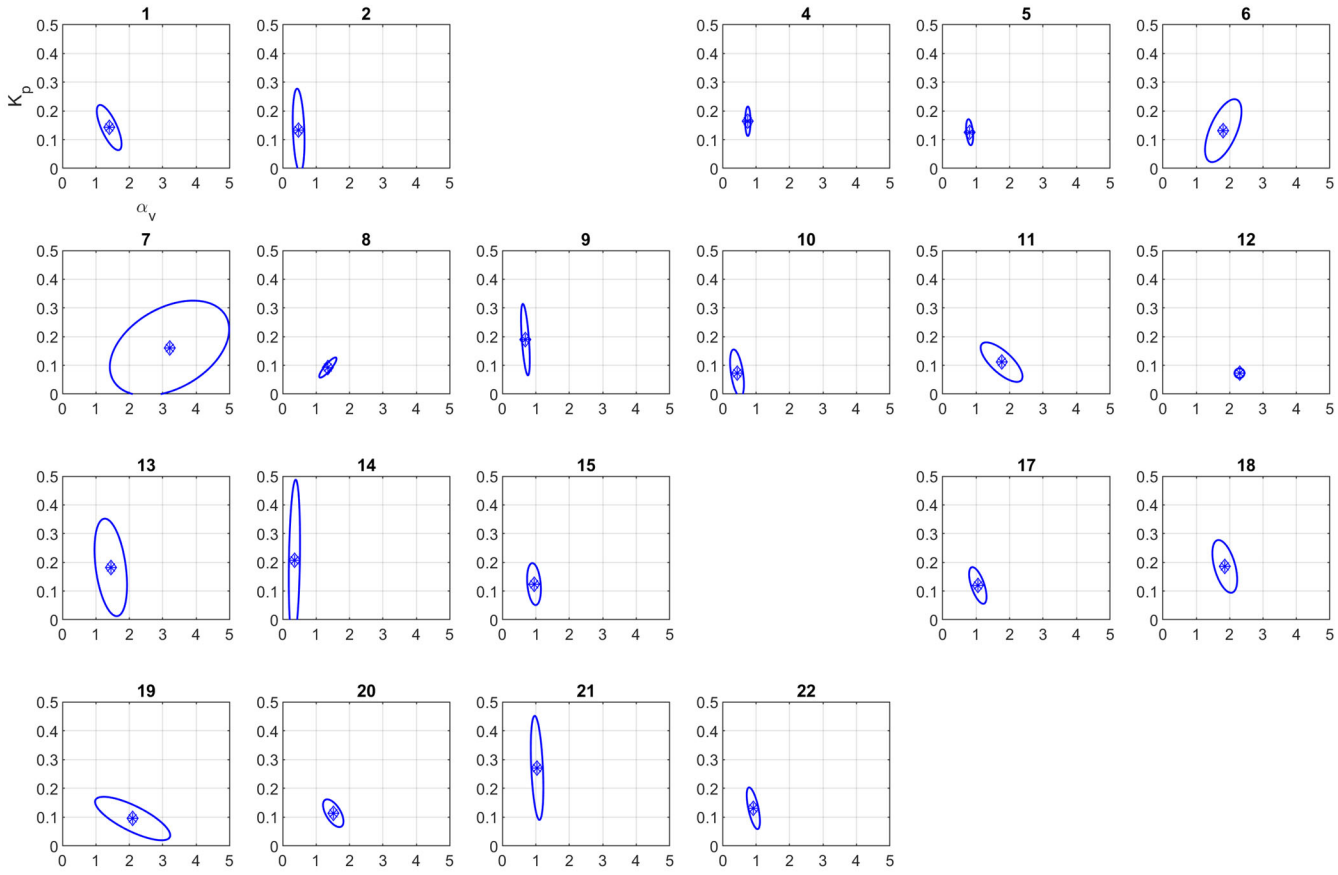


Figure 5. 7: 95% confidence regions for the calibrated parameters using the 2p model. Subjects 3 and 16 were excluded because calibrated parameters were outside physiological range (see the discussion section)

5.2.7.2 Experiment UQ

In order for the validation results to be meaningful, experimental uncertainty must be quantified and compared with computational uncertainty. Uncertainty in the change in BV measurements may be quantified by evaluating the standard deviation of the baseline BV (BV_0) which is based on ICG measurement of plasma volume (PV_0) and hematocrit (Hct_0). Since the animal study was not designed to have repeated measurements of BV at multiple points during the course of the experiment, the standard deviation of such measurements cannot be determined based on the available data. While accuracy of BV measurements for the ICG technique has been studied against gold standards such as radiolabeled albumin technique [165], published literature on repeatability or reproducibility of BV measurement is scarce [167] and few studies that have attempted it have focused on the specific BV measurement technique used in our experimental study. However, it is possible to extract BV measurement standard deviation and quantify its proportionality to changes in BV from published literature on hemodialysis studies [165]. Experimental uncertainty quantified based on this reference, which used a validated method against the ICG technique, yielded a proportionality constant of 0.2 between measurements of change in BV and their standard deviation. The experimental uncertainties are depicted as part of Figure (5.7) and are acceptable for the COU of the model.

5.2.8 Step 8: Model Validation

The above steps ensure that the parameters of the mathematical model are reliable, their uncertainty is adequately quantified, and propagated for quantitative comparison with experimental uncertainty. We only then proceeded to assess the predictive capability of the SSPM based on its COU which is to provide subject-specific prediction of changes to BV response to hemorrhage and fluid resuscitation to support pre-clinical safety assessment. Given this COU, two validation tests were formulated: (i) assessment of prediction of responsiveness to fluid therapy at the conclusion of the study (i.e., 180 min), in which general responsiveness of BV was defined by rise of BV above normovolumia ($\Delta V_{B,e} = 0$) in response to subject-specific fluid therapy (i.e. irrespective of variability in fluid therapy); and (ii) assessment of prediction of responsiveness to fluid therapy defined by BV restoration to normovolemic state at subject-specific time instant of cumulative fluid infusion equal to cumulative hemorrhage. For each validation test, we evaluated the probability that the mathematical model's binary classification of responsiveness/non-responsiveness was in agreement with that observed experimentally.

Figure (5.8) plots the results of each subject's calibrated model for the entire 180 min (see red dashed line), together with the corresponding experimental data (triangles). The first 50 min corresponds to model calibration (same data as Figure (5.6)), whereas the 50 min to 180 min results represent model validation. The parameter's quantified uncertainty discussed in the previous section was also propagated through the mathematical model using Monte Carlo sampling (N=10,000) to derive the uncertainty in the model predictions. These are represented in Figure (5.8) as a 95% confidence interval in model output (red shaded region) and can be compared to

the 95% confidence interval for the experimental measurements (blue shaded region) which is based on the values discussed in the previous section.

Finally, the results of the binary tests to evaluate the mathematical model's prediction of fluid responsiveness for critical time points during therapy while accounting for the uncertainty on both the prediction and the experiment are presented in Table (5.3).

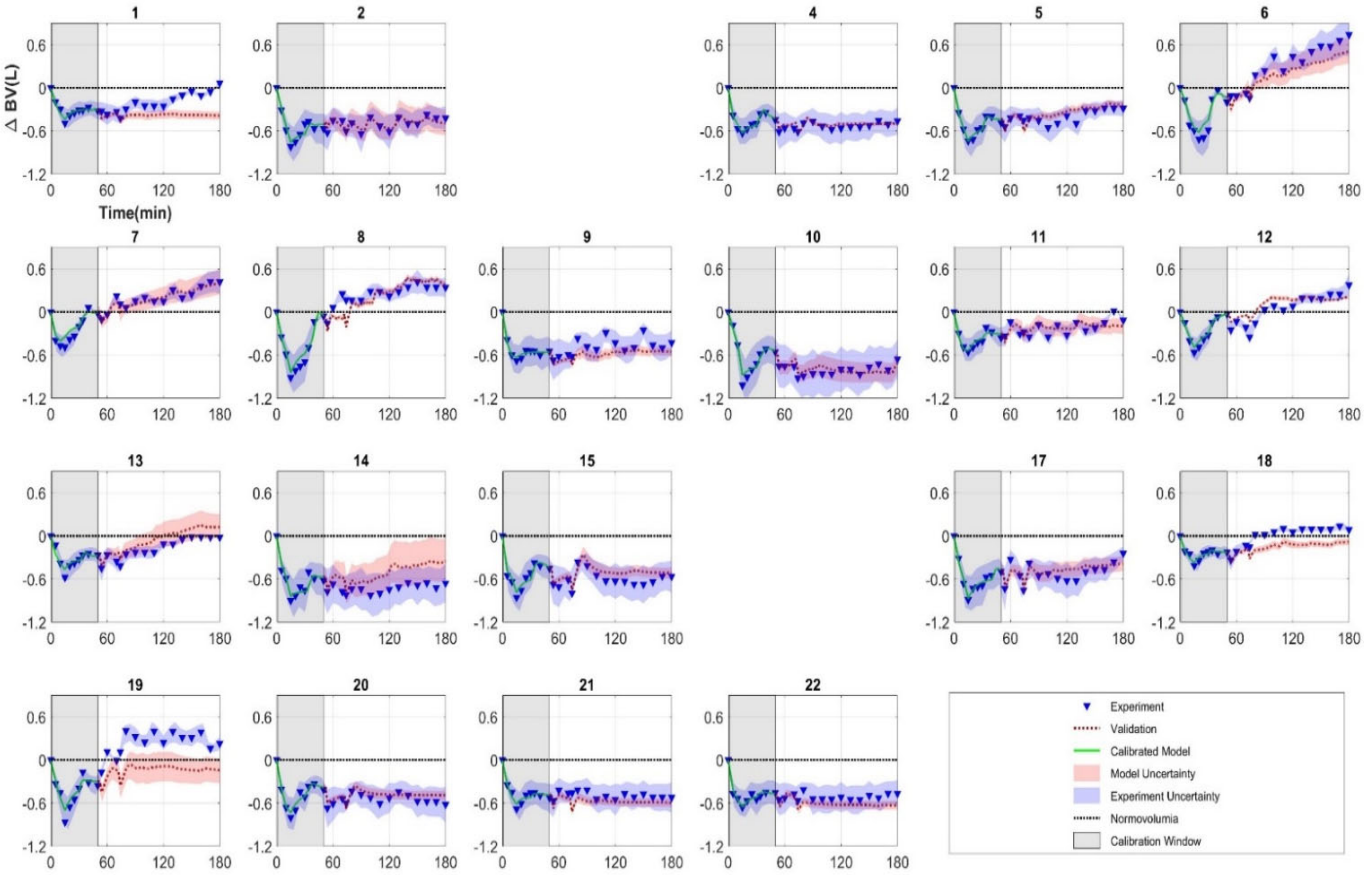


Figure 5. 8: Subject-specific model calibration and validation. Each plot represents a different animal (Subjects 3 and 16 were excluded because parameters were not identifiable for these subjects). Model was calibrated to experimental data between 0 and 50 mins (model: green line, experiment: triangles). Model was then simulated from 50 mins to 180 mins (red dashed line) and was compared against experimental measurements (triangles) (model validation). Shaded regions represent 95% confidence intervals.

Table 5. 3: Results of quantitative validation tests. $P_{agreement,180}$ *is the probability of the model and experiment both being greater than 0, or both being less than 0, at t=180 mins. T* is the time at which cumulative infusion was equal to cumulative hemorrhage (no such time exists for some subjects).* $P_{agreement,T^*}$ *is the probability of the model and experiment both being greater than 0, or both being less than 0, at t=T**

Subject	$P_{agreement,180}$	T* (min)	$P_{agreement,T^*}$
1	0.00	-	n/a
2	1.00	96	1.00
4	1.00	-	n/a
5	1.00	108	1.00
6	1.00	78	1.00
7	1.00	84	1.00
8	1.00	72	0.02
9	1.00	-	n/a
10	1.00	-	n/a
11	1.00	180	1.00
12	1.00	90	1.00
13	0.04	102	1.00
14	0.98	84	1.00
15	1.00	-	n/a
17	1.00	174	1.00
18	0.00	174	0.00
19	0.04	177	0.04
20	1.00	-	n/a
21	1.00	-	n/a
22	1.00	-	n/a

In the validation tests, 82% and 75% of the subject- specific mathematical models were able to correctly predict blood volume response when predictive capability was evaluated at 180 min and at the time when amount of infused fluid equals fluid loss.

5.3 Discussion

We have evaluated the credibility of an SSPM model of BV changes in response to hemorrhage and fluid infusion, using the proposed novel framework. This SSPM provided a case study in which a complete (all 8 steps) application of the proposed framework was demonstrated successfully. The framework was successful in establishing the credibility of a mathematical model to avoid problems such as parameter unidentifiability and helped identify a validation window with sufficient information to allow predictive capability evaluation. As both issues are commonly encountered in subject-specific modeling the framework also demonstrates broader applicability for SSPMs with COUs related to prediction of physiological variables such as blood volume and cardiac output. Furthermore, the framework proved effective in avoiding resource intensive steps such as collection of additional data in order to enhance parameter and model prediction reliability. This was accomplished by an analysis of calibration results and PIA as prescribed by the framework to remedy the practically unidentifiable parameter α_u by fixing it and still maintaining acceptable predictive capability of the SSPM without having to resort to additional data collection.

As the first step of the framework, we characterized the COU of the model to be for prediction of change in BV in response to fluid challenge or hemorrhage. This characterization proved to be consequential in determining the rigor of credibility stages in the later stages. For example, it necessitates splitting the calibration data into two segments using a calibration window such that independent data may be used for the purpose of validating the model. This COU also emphasizes the importance of conducting a rigorous *a priori* and *a posteriori* identifiability

analysis for the purpose of determining uniqueness and reliability of parameter estimates to be used for prediction of the physiological response.

Once the SSPM COU is characterized and the model is formulated, it needs to be evaluated for face validity in Step 2 of the framework. This step was conducted successfully on the BV kinetic model by first looking at simulation output of the model in response to episodes of fluid infusion and hemorrhage. Furthermore, the impact of change of model parameters was qualitatively assessed and determined to be physiologically relevant.

In Step 3 toward evaluating the credibility of this SSPM, the fundamental soundness of model structure and whether it allowed for unique parameter identification was evaluated by SIA. Our proof included in this work demonstrated the mathematical model is globally structurally identifiable, meaning it is theoretically possible to design an experiment in such a way that the resulting post-calibration parameters are unique

In Step 4, the model output sensitivity to parameters was successfully evaluated and leveraged to rank the parameters in the order of their identifiability and in turn identification of a calibration window for the purpose of model validation. This step was necessary as the COU of the model is for predictive purposes and redundant or inconsequential parameters may need to be omitted to improve predictive capability. However its main utility for this SSPM and its associated COU was for quantification of a calibration window. It is often desirable to qualify a subset of data for reliable calibration particularly in the case of SSPMs where a portion of the data is needed to define the parameters associated with each subject. This enables independent assessment of the mathematical model using the remaining data, as the ultimate test of predictive capability of the mathematical model. Furthermore, depending on the number of model parameters to be estimated

and the associated optimization method, model calibration can be computationally expensive and take a long time to finish. In such circumstances, model order needed to be assessed for any potential opportunity to be reduced (i.e., by fixing insensitive parameters).

To gain insight for the order of parameter sensitivity and potential to quantify a calibration window, we leveraged SVD of the regressor matrix by reformulating the model into standard linear regression form. The calibration window was selected so that: (i) it would provide sufficient information to maximize the identifiability of the model parameters; and (ii) the remaining validation segment would include hemorrhage and infusion inputs distinct from those included in the calibration window to allow for evaluation of the mathematical model's predictive capability.

Utilization of SVD in this pre-calibration identifiability analysis was successful in gaining such insight on relative identifiability of parameters and selecting a suitable calibration window. SVD was applied to the regressor matrix for the entire duration of the experiment (T=180 min in Figure (5.4). For the duration of experiment, the singular values associated with α_v were the largest. This is potentially due to the experimental protocol allocating the initial 15 min of the experiment to hemorrhage only followed by 15 min of zero-input period before infusion could start at 30 min. Based on these results and that remaining hemorrhage inputs were at 50 min and 70 min, a calibration window between 0 min and 50 min was selected and we proceeded to calibrate the mathematical model using data in that window.

As part of Step 5, a quadratic cost function was used for nonlinear least square optimization for model calibration. This is consistent with the choice of the cost function and optimization method for most studies on the mathematical models for hemorrhage and fluid resuscitation [166], [168]. The cost function utilized in our study assumed that the experimental noise was proportional

to the magnitude of the true change in BV. This assumption is supported by published literature [165]. We designated a prespecified average RMSNE acceptance criterion of 20%.

The resulting calibrated parameters in Table (5.1) demonstrated that despite α_u having associated with nonzero singular values (i.e., the regressor matrix at 50 min having a full rank for all subjects) this parameter is identifiable at T_C =180 (results not shown) but may be unidentifiable at T_C=50 min for most of the subjects. Moreover, for the few subjects for which α_u was identified (Subjects 3, 13, and 16), α_v was unidentifiable for Subject 3 and the identified value for K_p took values significantly beyond the cohort mean for Subjects 13 and 16. The calibration step was therefore repeated after fixing α_u to an average value of 3 based on previous work [169]. Subjects 3 and 16 continued to have implausible parameter values, so they were deemed as outliers and examples of subjects for which the mathematical model fails to provide a valid BV prediction.

We suspect that the presence of physiological mechanism currently unaccounted for in the mathematical model may have resulted in its failure in these few subjects. Closer inspection of the experimental protocol for Subject 3 revealed that despite relatively small amount of infused fluid prior to T_C=50 min, the BV was restored to baseline at this time. This could be due to the activation of additional unmodeled compensatory mechanisms responsible for fluid shift resulting in higher sensitivity to fluid gain and lower sensitivity to fluid loss in the vascular compartment compared with other subjects. This in turn could have impacted the estimation of α_v as by definition this parameter is the ratio of fluid loss of intravascular and interstitial compartments. For subject 16, significantly high K_p value could be due to an unmodeled mechanism that influences fluid shift between intravascular and interstitial compartments. In other words, a physiological mechanism affecting balance of oncotic and hydrostatic pressure such as lymphatic flow [170], [171] could make rate of fluid loss different than fluid gain. Currently mathematical model assumes fluid shift

between the compartments without consideration of fluid shift direction, i.e., K_p is the same for intravascular fluid gain and loss. As such, the model structure may not adequately estimate K_p for this subject.

In Step 6 PIA was conducted to investigate the reliability of the parameter estimates. We first confirmed the identifiability of model parameters using cost function contour plots for the 3p model depicted in Figure (5.5), which confirms the previous calibration and pre-calibration identifiability results by showing unbounded contours in the direction of α_u and bounded contours for the other model parameters. We also visualized contour plots for all subjects using the 2p model, to verify they exhibited bounded contours (results not shown).

Step 7 involved uncertainty quantification of parameter estimates (Figure (5.8)). The results showed significant variability in both size and direction of the asymptotic 95% confidence region amongst subjects, which was expected considering the variability in subject-specific fluid infusion profiles magnified by the inherent inter- and intra-subject variabilities. Despite this variability, the general order of identifiability previously stated in the pre-calibration step could be verified by comparing the confidence intervals of the parameters in Table (5.2) or the shape of the ellipsoidal region. Propagation of uncertainty was carried out via the Monte Carlo parameter sampling based on the uncertainty quantified in the previous step.

In Step 8 and final step we conducted SSPM validation by assessing the predictive capability for the SSPM using qualitative and quantitative validation approaches considering the uncertainty quantified as part of Step 7 and depicted in Figure (5.7). The qualitative comparison of model prediction to experiment show varying degrees of predictive capability. Model predictions mostly aligned with experiments in subjects who did not reach normovolumia in the course of the therapy. There were instances of mathematical model under predicting (Subjects 1,

18, and 19) and over predicting (Subjects 13 and 14) the BV response. For quantitative validation and predictive capability assessment, we were interested in the segment of the time series independent of calibration, i.e., from T_c=50 min to T=180 min. As such, two physiologically relevant validation tests were developed at two critical times within this window.

When replacement fluid is infused to subjects under hemorrhagic shock, it is critical to avoid both under-infusion and over-infusion with the goal of maintaining BV at an ideal level which can be the subject's normovolumic state (i.e., restoration to baseline BV). As such, it is essential for the mathematical model to be able to predict two physiologically relevant scenarios: (i) whether fluid infusion regimen will result in the achievement of normovolumic state at the conclusion of fluid therapy (T=180 min); and (ii) whether normovolumia is achieved at the critical time when equal volume of fluid is infused compared to blood loss. The binary validation test results indicate that model predictions agree with the experimental results in 82% and 75% of the subjects for scenario i and ii, respectively (Table (5.3)).

For the first scenario, the mathematical model failed in predicting volumetric state in 4 Subjects (1, 13, 18, and 19). While for two Subjects (1 and 19) the predictions were drastically different than experimental results, the predictions for the other two Subjects (13 and 18) were very close to the observed BV and missed the binary test by small margin. In fact, the small difference between predictions and experimental results in these two subjects may not be deemed physiologically significant.

The second scenario was evaluated in 12 subjects only, as the rest of the subjects did not have protocols to allow for infusion volume to ever equal the preset hemorrhage volume. In this group, the mathematical model failed in three Subjects (8, 18, and 19). For Subjects 18 and 19, the time when volume of infusion equaled that of hemorrhage was very close to 180 min (174 min and

177 min, respectively). Thus, for these subjects, this scenario did not offer new information compared to the previous test and only confirms the previous scenario results.

In this work, we have demonstrated the utility of the proposed framework by conducting the complete process of validating a subject-specific BV kinetics model for the purpose of volume status prediction and demonstrated the mathematical model's utility for this context of use. Completing the framework in the prescribed order allowed for a comprehensive and streamlined assessment of the SSPMs credibility. In particular, having an explicitly defined COU with no *ad hoc* changes allowed for a credibility activity process in the later steps (e.g., Steps 3,4 and 6) with rigor commensurate with the COU of the model. Resource intensive steps such as model calibration and data collection were either streamlined or avoided for the purpose of model validation by following rigorous identifiability credibility activities yielding information as to how to proceed towards the final stages of the flowchart enabling model validation to complete with a solid conclusion. For example, calibration was streamlined by the preceding step 4 which allowed for quantification of calibration window with adequate information for parameter estimation while at the same time leaving sufficient independent data for validation purposes. Data collection was avoided by ensuring that despite fixing of one unidentifiable parameter the PIA, UQ and validation ultimately were not significantly affected, and the credibility assessment process could be completed using the entire steps prescribed by the flowchart.

It should be noted that for some animal subjects the study protocols did not allow for evaluation of model performance at the time cumulative infusion was equal to cumulative hemorrhage. Furthermore, fluid responsiveness in our work has been defined based on the changes in BV measurements. In pre-clinical and clinical setting, measured physiologic variables such as cardiac output or blood pressure are normally used for assessment of fluid responsiveness. That

said, clinical relevance of BV measurements and potential insight they offer towards subject resuscitation management is well documented[172]–[174].

Furthermore, the lack of experiments intended and designed for model validation also often results in absent or incomplete collection of information necessary for quantification of experimental uncertainty. In addition to the neglected accuracy specifications for instruments, data processing for such experiments often fails to account for measurement sensitivities for particular experimental parameters. Although our experimental uncertainty quantification is justified by the referenced literature which used a validated measurement technique, we acknowledge that differences in experimental conditions, subjects, and conduct could render the derived proportionality constant inaccurate. Potential mischaracterization of this value could alter the conclusion of the validation test.

5.4 Chapter Conclusion

The proposed framework was successfully applied in its entirety towards credibility assessment of the SSPM. The framework proved effective in identifying the type and rigor of credibility activities to be conducted based on the SSPMs COU. The framework successfully identified model limitations and provided direction as to how to address them so that the credibility assessment process could proceed and be completed.

Chapter 6: Dissertation Conclusion and Future Work

The urgent need for a framework for credibility assessment of SSPMs was established as part of the FDA and medical device industry meeting held in 2015 and further solidified as part of our published literature review of SSPMs. The literature review further identified the credibility activities currently being leveraged in other industry domains for credibility assessment of SSPMs. In this work, we proposed a novel framework to order the identified credibility activities in such a way to be maximally efficient towards credibility assessment of SSPMs. The proposed framework instills discipline in the model development and evaluation process and avoids *ad hoc* steps to save time and resources. Furthermore, the framework has been designed to address challenges and limitations such as data scarcity specific to SSPMs. Lastly, the framework enables efficient diagnosis of common SSPM or data weaknesses and provides a pathway for the modeler towards rectifying the issue(s) at hand. The utility of the proposed framework was demonstrated and verified in two case studies. First, the framework was applied to an SSPM of CO_2 gas exchange intended to predict physiological parameters such as lung volume and cardiac output in pediatric patients. Second, the framework was applied to an SSPM of blood volume response to fluid resuscitation intended to predict blood volume response to episodes of fluid infusion and hemorrhage. In both cases, the framework proved its merits.

The proposed framework has certain limitations which may be addressed as part of future research focusing on credibility assessment of the SSPMs.. The proposed framework should be regarded as a starting point with potential to be expanded and equipped with more detailed steps to investigate the credibility of the SSPM depending on the complexity level of the SSPM and its COU. For example, the framework in its current form, may not be directly applicable to complex SSPMs with numerous parameters. One potential area that may be explored is inclusion of

specific steps focusing on model selection and model parsimony determination which was not evaluated in detail as part of the case studies due to the nature of the investigated SSPMs. The proposed framework could also be enhanced to provide more detailed information to the investigators with regards to the type and quality of data required for evaluation of the SSPM. The basic framework proposed currently only provides general direction about additional or more informative data to address SSPM weaknesses such as lack of practical identifiability. Future work should be focused on additional and specific steps informing particular aspects of experimental design so as to allow for the SSPM to be practically identifiable and model parameters estimated with acceptable level of reliability. Furthermore, since PCLCs and CDS medical devices may have different risk profiles, the required rigor of credibility activities for the SSPM to be leveraged for their evaluation may vary. Future work may focus on isolating an application of, for example PCLC devices, and studying particular SSPM credibility activities commensurate with the risk associated with using the SSPM based on the unique performance attributes of PCLC devices under evaluation. Lastly, the present work may be extended to synthetic and population-based SSPMs to generate cohort of synthetic subjects to assess SAS performance under subject variability.

Glossary

Calibration Quality Evaluation: process of evaluating calibration results by confirming physiological relevance of the parameters and evaluating the quality of fit.

Clinical Decision Support System: A medical device intended to provide a therapeutic or diagnostic decision recommendation to the user.

Computational Patient Model: A representation of patient physiology in computational environment.

Context of Use: A statement that defines the specific role and scope of the computational model used to address a question of interest.

Credibility activity: An activity that generates evidence for model credibility.

Credibility assessment: The process of evaluating whether the evidence generated by the credibility activity establish model credibility.

Face validity: Activity that determines whether model outputs are consistent with expectations of domain experts.

Globally Structurally Identifiable: A model whose all parameters are unique globally

Locally Structurally Identifiable: A model whose all parameters are unique locally

Model calibration: The model building process of tuning or optimizing parameters in a mathematical model to minimize the difference between model outputs and real-world data.

Model credibility: The trust, based on all available evidence, in the predictive capability of the mathematical model.

Model formulation: Information on how the equations of a mathematical model are derived, including basic simplifying assumptions.

Model parsimony determination: The process of determining the complexity (including the order) of a mathematical model and determining if a simpler mathematical model can be selected (e.g., by fixing a subset of parameters) for the same level of agreement with data.

Model validation: The process of determining the degree to which a model or a simulation is an accurate representation of the real world.

Semi-Autonomous System: A medical device that enables some degree of automation in patient care either by directly adjusting therapy (in case of PCLC) or by recommending therapy (in case of CDS).

Sensitivity Analysis: The process of determining how a change in a model input (e.g., parameters or initial conditions) affects model outputs.
Structural Identifiability Analysis: The process of determining the uniqueness of parameter estimates given the structure of the model and under a theoretical assumption that the calibration data are perfect.

Subject-Specific Physiological Model: A CMP whose parameters are calibrated to subject-specific data.

Parameter physiological relevance: Activity that determines whether the calibrated model parameters make physical sense and are consistent with their physical value expectation.

Precalibration Sensitivity Analysis: A sensitivity analysis conducted prior to obtaining calibrated parameter values and with nominal parameter.

Practical Identifiability Analysis: The process of determining the reliability of parameter estimates given the experimental data used for model calibration.

Physiological Closed-Loop Controller: A medical device that incorporates physiological sensor(s) for automatic manipulation of a physiological variable through actuation of therapy that is conventionally made by a clinician.

Uncertainty quantification: The process of determining the uncertainty in model inputs (e.g., parameters or initial conditions) and computing the resultant uncertainty in model outputs.

Bibliography

[1] B. Parvinian, C. Scully, H. Wiyor, A. Kumar, and S. Weininger, "Regulatory Considerations for Physiological Closed-Loop Controlled Medical Devices Used for Automated Critical Care: Food and Drug Administration Workshop Discussion Topics," *Anesth Analg*, 2017.

[2] B. Parvinian *et al.*, "Credibility evidence for computational patient models used in the development of physiological closed-loop controlled devices for critical care medicine," *Front. Physiol.*, 2019.

[3] N. R. Marques *et al.*, "Automated closed-loop resuscitation of multiple hemorrhages: a comparison between fuzzy logic and decision table controllers in a sheep model," *Disaster Mil. Med.*, vol. 3, p. 1, 2017.

[4] J. Rinehart, C. Lee, C. Canales, A. Kong, Z. Kain, and M. Cannesson, "Closed-loop fluid administration compared to anesthesiologist management for hemodynamic optimization and resuscitation during surgery: An in vivo study," *Anesth. Analg.*, vol. 117, no. 5, pp. 1119–1129, 2013.

[5] J. Salinas *et al.*, "Closed-loop and decision-assist resuscitation of burn patients.," *J. Trauma*, vol. 64, no. 4 Suppl, pp. S321–S332, 2008.

[6] E. Morozoff, J. A. Smyth, and M. Saif, "Applying Computer Models to Realize Closed-Loop Neonatal Oxygen Therapy," *Anesth. Analg.*, vol. 124, no. 1, pp. 95–103, 2017.

[7] E. P. Morozoff and M. Saif, "OXYGEN THERAPY CONTROL OF NEONATES – PART I: A MODEL OF NEONATAL OXYGEN TRANSPORT," *Control Intell. Syst.*, vol. 36, no. 3, 2008.

[8] F. T. Tehrani and S. Abbasi, "Evaluation of a computerized system for mechanical ventilation of infants," *J. Clin. Monit. Comput.*, 2009.

[9] N. R. Marques *et al.*, "Physician-Directed Versus Computerized Closed-Loop Control of Blood Pressure Using Phenylephrine in a Swine Model," *Anesth Analg*, vol. 125, no. 1, pp. 110–116, 2017.

[10] M. M. Silva, L. Paz, T. Wigren, and T. Mendonça, "Performance of an adaptive controller for the neuromuscular blockade based on inversion of a Wiener model," *Asian J. Control*, vol. 17, no. 4, pp. 1136–1147, 2015.

[11] S. Bibian, G. A. Dumont, M. Huzmezan, and C. R. Ries, "Patient Variability and Uncertainty Quantification in Anesthesia: Part I – Pkpd Modeling and Identification," *IFAC Proc. Vol.*, vol. 39, no. 18, pp. 549–554, 2006.

[12] W. N. *et al.*, "Robust closed-loop control of induction and maintenance of propofol anesthesia in children," *Paediatr. Anaesth.*, vol. 23, no. 8, pp. 712–719, 2013.

[13] J. O. Hahn, G. A. Dumont, and J. M. Ansermino, "Observer-Based Strategies for Anesthesia Drug Concentration Estimation," *Cdn.Intechopen.Com*, 2006.

[14] O. Faris and J. Shuren, "An FDA Viewpoint on Unique Considerations for Medical-Device Clinical Trials," *N Engl J Med*, vol. 376, no. 14, pp. 1350–1357, 2017.

[15] M. L. Neal and J. B. Bassingthwaighte, "Subject-specific model estimation of cardiac output and blood volume during hemorrhage," *Cardiovasc Eng*, vol. 7, no. 3, pp. 97–120, 2007.

[16] A. D. Marquis, A. Arnold, C. Dean-Bernhoft, B. E. Carlson, and M. S. Olufsen, "Practical identifiability and uncertainty quantification of a pulsatile cardiovascular model," *Math. Biosci.*, vol. 304, pp. 9–24, 2018.

[17] H. Mirinejad *et al.*, “Evaluation of Fluid Resuscitation Control Algorithms via a Hardware-in-the-Loop Test Bed,” *IEEE Trans. Biomed. Eng.*, vol. PP, no. c, p. 1, 2019.
[18] B. Parvinian, R. Bighamian, C. G. Scully, J. O. Hahn, and P. Pathmanathan, “Credibility Assessment of a Subject-Specific Mathematical Model of Blood Volume Kinetics for Prediction of Physiological Response to Hemorrhagic Shock and Fluid Resuscitation,” *Front. Physiol.*, vol. 12, no. September, pp. 1–14, 2021.
[19] R. Bighamian, C.-S. Kim, A. T. Reisner, and J.-O. Hahn, “Closed-Loop Fluid Resuscitation Control Via Blood Volume Estimation,” *J. Dyn. Syst. Meas. Control*, vol. 138, no. 11, p. 111005, 2016.
[20] R. Bighamian, B. Parvinian, C. G. Scully, G. Kramer, and J.-O. Hahn, “Control-oriented physiological modeling of hemodynamic responses to blood volume perturbation,” *Control Eng. Pract.*, vol. 73, no. April 2018, pp. 149–160, Apr. 2018.
[21] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, “In Silico Preclinical Trials: A Proof of Concept in Closed-Loop Control of Type 1 Diabetes,” *J. Diabetes Sci. Technol.*, vol. 3, no. 1, pp. 44–55, 2009.
[22] C. Jean-Quartier, F. Jeanquartier, I. Jurisica, and A. Holzinger, “In silico cancer research towards 3R,” *BMC Cancer*, vol. 18, no. 1, pp. 1–12, 2018.
[23] M. Viceconti, F. Pappalardo, B. Rodriguez, M. Horner, and F. Musuamba, “In silico trials : Veri fi cation , validation and uncertainty quanti fi cation of predictive models used in the regulatory evaluation of biomedical products,” no. September 2019, 2020.
[24] A. Deva *et al.*, “Interpretability of Epidemiological Models : The Curse of Non-Identifiability,” pp. 1–8, 2021.
[25] A. Capaldi, S. Behrend, B. Berman, J. Smith, J. Wright, and A. L. Lloyd, “Parameter estimation and uncertainty quantification for an epidemic model,” *Math. Biosci. Eng.*, vol. 9, no. 3, pp. 553–576, 2012.
[26] G. Massonis, J. R. Banga, and A. F. Villaverde, “Structural identifiability and observability ofcompartmental models ofthe COVID-19 pandemic,” no. January, 2020.
[27] A. D. Jackson *et al.*, “Validation and use of a parametric model for projecting cystic fibrosis survivorship beyond observed data: A birth cohort analysis,” *Thorax*, vol. 66, no. 8, pp. 674–679, 2011.
[28] A. Raue *et al.*, “Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood,” *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, Aug. 2009.
[29] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou, “Structural Identifiability of Dynamic Systems Biology Models,” *PLoS Comput. Biol.*, vol. 12, no. 10, pp. 1–22, 2016.
[30] M. P. Saccomani and K. Thomaseth, “The Union between Structural and Practical Identifiability Makes Strength in Reducing Oncological Model Complexity: A Case Study,” *Complexity*, vol. 2018, 2018.
[31] A. F. Brouwer, R. Meza, and M. C. Eisenberg, “A Systematic Approach to Determining the Identifiability of Multistage Carcinogenesis Models,” *Risk Anal.*, vol. 37, no. 7, pp. 1375–1387, 2017.
[32] P. Pathmanathan, M. S. Shotwell, D. J. Gavaghan, J. M. Cordeiro, and R. A. Gray, “Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology,” *Prog. Biophys. Mol. Biol.*, vol. 117, no. 1, pp. 4–18, 2015.
[33] P. Pathmanathan, J. M. Cordeiro, and R. A. Gray, “Comprehensive uncertainty

quantification and sensitivity analysis for cardiac action potential models," *Front. Physiol.*, vol. 10, no. JUN, pp. 1–20, 2019.
[34] J. Doherty and R. J. Hunt, "Two statistics for evaluating parameter identifiability and error reduction," *J. Hydrol.*, vol. 366, no. 1–4, pp. 119–127, 2009.
[35] C. J. Barnes, "The art of catchment modeling: What is a good model?," *Environ. Int.*, vol. 21, no. 5, pp. 747–751, 1995.
[36] T. A. Watson, J. E. Doherty, and S. Christensen, "Parameter and predictive outcomes of model simplification," *Water Resour. Res.*, vol. 49, no. 7, pp. 3952–3977, 2013.
[37] R. Muñoz-Tamayo *et al.*, "Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling?," *Animal*, vol. 12, no. 4, pp. 701–712, 2018.
[38] L. Puillet, O. Martin, M. Tichit, and D. Sauvant, "Simple representation of physiological regulations in a model of lactating female: Application to the dairy goat," *Animal*, vol. 2, no. 2, pp. 235–246, 2008.
[39] E. Carson, *Modelling methodology for physiology and medicine.* Elsevier Science Publishi, 2013.
[40] J. Batzel, M. Bachar, and F. Kappel, "Mathematical Modeling and Validation in Physiology," in *Springer*, 2013.
[41] A. Dehghani, N. Malagutti, and R. A. Kennedy, "Robust control design for automatic regulation of blood pressure," *IET Control Theory Appl.*, vol. 7, no. 3, pp. 387–396, 2013.
[42] S. J. Silva, T. A. Scardovelli, H. N. Martucci, S. R. M. S. Boschi, and A. P. Silva, "Mean arterial pressure-embedded model for real-time simulation and controller validation," *Electron. Lett.*, vol. 53, no. 7, pp. 450–452, 2017.
[43] J. Rinehart *et al.*, "Evaluation of a novel closed-loop fluid-administration system based on dynamic predictors of fluid responsiveness: An in silico simulation study," *Crit. Care*, vol. 15, no. 6, p. R278, 2011.
[44] T. Wassar *et al.*, "Automatic Control of Arterial Pressure for Hypotensive Patients Using Phenylephrine," *Int. J. Model. Simul.*, vol. 34, no. 4, 2014.
[45] C. M. Held and R. J. Roy, "Multiple drug hemodynamic control by means of a supervisory-fuzzy rule- based adaptive control system: Validation on a model," *IEEE Trans. Biomed. Eng.*, 1995.
[46] J. W. Huang and R. J. Roy, "Multiple-drug hemodynamic control using fuzzy decision theory.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 2, pp. 213–28, 1998.
[47] R. R. Rao, B. W. Bequette, and R. J. Roy, "Simultaneous regulation of hemodynamic and anesthetic states: a simulation study," *Ann. Biomed. Eng.*, vol. 28, no. 1, pp. 71–84, 2000.
[48] J. F. Martin, N. T. Smith, M. L. Quinn, and A. M. Schneider, "Supervisory Adaptive Control of Arterial Pressure During Cardiac Surgery," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 4, pp. 389–393, 1992.
[49] K. Kashihara, T. Kawada, K. Uemura, M. Sugimachi, and K. Sunagawa, "Adaptive predictive control of arterial blood pressure based on a neural network during acute hypotension," *Ann. Biomed. Eng.*, vol. 32, no. 10, pp. 1365–1383, Oct. 2004.
[50] R. Bighamian, C.-S. Kim, A. T. Reisner, and J.-O. Hahn, "Closed-Loop Fluid Resuscitation Control Via Blood Volume Estimation," *J. Dyn. Syst. Meas. Control*, vol. 138, no. 11, p. 111005, 2016.
[51] J. Rinehart, C. Lee, M. Cannesson, and G. Dumont, "Closed-loop fluid resuscitation: Robustness against weight and cardiac contractility variations," *Anesth. Analg.*, vol. 117,

no. 5, pp. 1110–1118, 2013.
[52] J. B. Slate and L. C. Sheppard, “A Model-Based Adaptive Blood Pressure Controller,” *IFAC Proc. Vol.*, vol. 15, no. 4, pp. 1437–1442, Jun. 1982.
[53] Chin-Te Chen, Wen-Li Lin, Te-Son Kuo, and Cheng-Yi Wang, “Adaptive control of arterial blood pressure with a learning controller based on multilayer neural networks,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, pp. 601–609, 1997.
[54] N. Malagutti, A. Dehghani, and R. A. Kennedy, “Robust control design for automatic regulation of blood pressure,” *IET Control Theory Appl.*, vol. 7, no. 3, pp. 387–396, 2013.
[55] M. M. Polycarpou and J. Y. Conway, “Indirect Adaptive Nonlinear Control of Drug Delivery Systems,” *Science (80-.).*, vol. 43, no. 6, pp. 849–856, 1998.
[56] Y. Gao and M. J. Er, “An intelligent adaptive control scheme for postsurgical blood pressure regulation.,” *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 475–483, 2005.
[57] T. Luspay and K. Grigoriadis, “Robust linear parameter-varying control of blood pressure using vasoactive drugs,” *Int. J. Control*, vol. 88, no. 10, pp. 2013–2029, 2015.
[58] G. L. Coté, R. Durai, and B. Zoghi, “Nonlinear closed-loop control system for intracranial pressure regulation.,” *Ann. Biomed. Eng.*, vol. 23, pp. 760–771, 1995.
[59] J. F. Martin, a M. Schneider, M. L. Quinn, and N. T. Smith, “Improved safety and efficacy in adaptive control of arterial blood pressure through the use of a supervisor.,” *IEEE Trans. Biomed. Eng.*, vol. 39, no. 4, pp. 381–8, 1992.
[60] E. A. Woodruff, J. F. Martin, and M. Omens, “A model for the design and evaluation of algorithms for closed-loop cardiovascular therapy,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 8, pp. 694–705, 1997.
[61] M. Wysocki *et al.*, “Autonomous control of oxygenation.,” *Respir. Physiol. Neurobiol.*, vol. 28, no. 1, pp. 77–88, Jun. 2009.
[62] J. Urzua, E. Sauma, A. Cipriano, and M. Guarini, “A Model of internal control may improve the response time of an automatic arterial pressure controller.”
[63] C. N. Nguyen *et al.*, “The benefits of using Guyton’s model in a hypotensive control system,” *Comput Methods Programs Biomed*, vol. 89, no. 2, pp. 153–161, 2008.
[64] C. W. Frei, M. Derighetti, M. Morari, A. H. Glattfelder, and A. Zbinden, “Improved Regulation of Mean Arterial Blood Pressure During Anesthesia Through Estimates of Surgery Effects,” *{IEEE} {T}rans. {B}iomed. {E}ng.*, vol. 47, no. 11, pp. 1456–1464, 2000.
[65] C. Yu, R. J. Roy, H. Kaufman, and B. W. Bequette, “Multiple-model adaptive predictive control of mean arterial pressure and cardiac output,” *IEEE Trans. Biomed. Eng.*, vol. 39, no. 8, pp. 765–778, 1992.
[66] E. P. Martinoni *et al.*, “Model-based control of mechanical ventilation: design and clinical validation.,” *Br. J. Anaesth.*, vol. 92, no. 6, pp. 800–807, 2004.
[67] A. Pomprapa, S. Leonhardt, and B. J. E. Misgeld, “Optimal learning control of oxygen saturation using a policy iteration algorithm and a proof-of-concept in an interconnecting three-tank system,” *Control Eng. Pract.*, vol. 59, pp. 194–203, 2017.
[68] Y. Jiang, Q. Sun, P. Tan, and Z. Chen, “Modeling and Simulation of an Electronic Oxygen Regulator Based on All-Coefficient Adaptive Control,” *J. Dyn. Syst. Meas. Control*, vol. 138, no. 8, p. 081010, 2016.
[69] E. P. Morozoff and M. Saif, “OXYGEN THERAPY CONTROL OF NEONATES: PART II – EVALUATING MANUAL, PID AND FUZZY LOGIC CONTROLLER DESIGNS,” *Control Intell. Syst.*, vol. 36, no. 3, 2008.

[70] F. T. Tehrani, “Mathematical analysis and computer simulation of the respiratory system in the newborn infant,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 5, pp. 475–481, 1993.
[71] M. G. Iobbi, A. K. Simonds, and R. J. Dickinson, “Oximetry feedback flow control simulation for oxygen therapy,” *J. Clin. Monit. Comput.*, 2007.
[72] E. P. Morozoff and M. Saif, “Oxygen Therapy Control of Neonates – Part I : a Model of Neonatal Oxygen Transport,” vol. 36, no. 3, 2008.
[73] F. T. Tehrani, “A closed-loop system for control of the fraction of inspired oxygen and the positive end-expiratory pressure in mechanical ventilation,” *Comput Biol Med*, vol. 42, no. 11, pp. 1150–1156, 2012.
[74] C. Yu, “Improvement in Arteral Oxygen Control Using Multiple-Model Adaptive Control Procedures,” no. 8, 1987.
[75] E. P. Martinoni *et al.*, “Model-based control of mechanical ventilation: design and clinical validation.,” *Br. J. Anaesth.*, vol. 92, no. 6, pp. 800–807, 2004.
[76] J. O. Hahn, G. A. Dumont, and J. M. Ansermino, “System identification and closed-loop control of end-tidal CO2 in mechanically ventilated patients,” *IEEE Trans. Inf. Technol. Biomed.*, 2012.
[77] C.-S. Kim, J. M. Ansermino, and J.-O. Hahn, “A Comparative Data-Based Modeling Study on Respiratory CO2 Gas Exchange during Mechanical Ventilation.,” *Front. Bioeng. Biotechnol.*, vol. 4, p. 8, 2016.
[78] O. Flechelles *et al.*, “Simulations for mechanical ventilation in children: review and future prospects,” *Crit Care Res Pr.*, vol. 2013, no. 1, pp. 1–8, 2013.
[79] F. T. Tehrani, “A closed-loop system for control of the fraction of inspired oxygen and the positive end-expiratory pressure in mechanical ventilation,” *Comput. Biol. Med.*, 2012.
[80] T. L. Fernando, J. S. Packer, and J. F. Cade, “A CLOSED-LOOP SYSTEM FOR CONTROLLING BLOOD OXYGEN AND CARBON DIOXIDE LEVELS IN MECHANICALLY VENTILATED PATIENTS,” *ControlEng. Pract.*, vol. 3, no. 10, pp. 433–1440, 1995.
[81] A. Pomprapa, B. Misgeld, V. Sorgato, A. Stollenwerk, M. Walter, and S. Leonhardt, “ROBUST CONTROL OF END-TIDAL CO_2 USING THE H∞ LOOP-SHAPING APPROACH,” *Acta Polytech.*, vol. 53, no. 6, 2013.
[82] A. Sano, K. Ohkubo, H. Ohmorif, and M. Kikuchi, “ADAPTIVE REGULATION OF ARTERIAL GAS PRESSURES BY USING ROBUST ADAPTIVE CONTROL SCHEME A.,” no. December, 1988.
[83] W. F. Fincham and F. T. Tehrani, “A MATHEMATICAL MODEL OF THE HUMAN RESPIRATORY SYSTEM.”
[84] V. C. Rideout, “Mathematical and computer modeling of physiological systems,” *Prentice Hall Biophys. Bioeng. Ser.*, pp. xv, 261 p., 1991.
[85] F. T. Tehrani, “Mathematical analysis and computer simulation of the respiratory system in the newborn infant,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 5, pp. 475–481, 1993.
[86] P. E. Morozoff, R. W. Evans, J. A. Smyth, and C. Vh, “AUTOMATIC CONTROL OF BLOOD OXYGEN SATURATION,” no. 1, 1993.
[87] L. Chiari, G. Avanzolini, and M. Ursino, “A comprehensive simulator of the human respiratory system: Validation with experimental and simulated data,” *Ann. Biomed. Eng. 1997 256*, vol. 25, no. 6, pp. 985–999, Nov. 1997.
[88] J. W. Severinghaus, “Simple, accurate equations for human blood O2 dissociation computations,” *https://doi.org/10.1152/jappl.1979.46.3.599*, vol. 46, no. 3, pp. 599–602,

1979.
[89] S. Bibian, G. A. Dumont, M. Huzmezan, and C. R. Ries, “Patient variability and uncertainty quantification in anesthesia: Part II - PKPD uncertainty,” *IFAC Proc. Vol.*, vol. 6, no. PART 1, pp. 555–560, 2006.
[90] J.-O. Hahn, G. A. Dumont, and J. M. Ansermino, “Robust closed-loop control of hypnosis with propofol using WAVCNS index as the controlled variable,” *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 517–524, 2012.
[91] M. Elkfafi, J. S. Shieh, D. A. Linkens, and J. E. Peacock, “Fuzzy logic for auditory evoked response monitoring and control of depth of anaesthesia,” *Fuzzy Sets Syst.*, vol. 100, no. 1–3, pp. 29–43, 1998.
[92] G. A. Dumont, A. Martinez, and J. M. Ansermino, “Robust control of depth of anesthesia,” *Int. J. Adapt. Control Signal Process.*, vol. 23, no. 5, pp. 435–454, May 2009.
[93] C. M. Ionescu, I. Nascu, and R. De Keyser, “Lessons learned from closed loops in engineering: towards a multivariable approach regulating depth of anaesthesia,” *J. Clin. Monit. Comput.*, vol. 28, no. 6, pp. 537–546, 2014.
[94] A. Gentilini, C. Schaniel, M. Morari, C. Bieniok, R. Wymann, and T. Schnider, “A new paradigm for the closed-loop intraoperative administration of analgesics in humans,” *IEEE Trans. Biomed. Eng.*, vol. 49, no. 4, pp. 289–299, 2002.
[95] J. Almeida, T. Mendonça, P. Rocha, and L. Rodrigues, “Controller design for neuromuscular blockade level tracking based on optimal control,” *Control Eng. Pract.*, vol. 59, pp. 151–158, 2017.
[96] Z. T. Zhusubaliyev, A. Medvedev, and M. M. Silva, “Nonlinear dynamics in closed-loop anesthesia: Pharmacokinetic/ pharmacodynamic model under PID-feedback,” *Proc. Am. Control Conf.*, pp. 5496–5501, 2014.
[97] M. Fang, Y. Tao, and Y. Wang, “An enriched simulation environment for evaluation of closed-loop anesthesia,” *J. Clin. Monit. Comput.*, vol. 28, no. 1, pp. 13–26, 2014.
[98] M. M. Silva, J. M. Lemos, A. Coito, B. A. Costa, T. Wigren, and T. Mendonça, “Local identifiability and sensitivity analysis of neuromuscular blockade and depth of hypnosis models,” *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 23–36, 2014.
[99] X. Jin, C. S. Kim, S. T. Shipley, G. A. Dumont, and J. O. Hahn, “Coordinated semi-adaptive closed-loop control for infusion of two interacting medications,” *Int. J. Adapt. Control Signal Process.*, vol. 32, no. 1, pp. 134–146, Jan. 2018.
[100] M. Y. Liberman, S. Ching, J. Chemali, and E. N. Brown, “A closed-loop anesthetic delivery system for real-time control of burst suppression,” *J. Neural Eng.*, vol. 10, no. 4, 2013.
[101] M. B. Westover, S. E. Kim, S. Ching, P. L. Purdon, and E. N. Brown, “Robust control of burst suppression for medical coma,” *J. Neural Eng.*, vol. 12, no. 4, 2015.
[102] E. Kharisov, C. L. Beck, and M. Bloom, *Control of patient response to anesthesia using $\mathscr{L}1$adaptive methods*, vol. 45, no. 18. IFAC, 2012.
[103] L. Merigo *et al.*, “On the Identification of the Propofol PK/PD Model Using BIS Measurements,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 868–873, 2017.
[104] D. A. Linkens and M. Mahfouf, “Generalized predictive control with feedforward (GPCF) for multivariable anaesthesia,” *Int. J. Control*, vol. 56, no. 5, pp. 1039–1057, 1992.
[105] S. Yelneedi, S. Lakshminarayanan, and G. P. Rangaiah, “A comparative study of three advanced controllers for the regulation of hypnosis,” *J. Process Control*, vol. 19, no. 9, pp. 1458–1469, 2009.

[106] F. Y. Liu and R. B. Northrop, “A new Approach to the Modeling and Control of Postoperative Pain,” *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1147–1158, 1990.
[107] J. A. Méndez, A. Marrero, J. A. Reboso, and A. León, “Adaptive fuzzy predictive controller for anesthesia delivery,” *Control Eng. Pract.*, vol. 46, pp. 1–9, 2016.
[108] M. M. R. F. Struys, T. De Smet, S. Greenwald, A. R. Absalom, S. Bingé, and E. P. Mortier, “Performance Evaluation of Two Published Closed-loop Control Systems Using Bispectral Index Monitoring: A Simulation Study,” *Anesthesiology*, vol. 100, no. 3, pp. 640–647, Mar. 2004.
[109] F. N. Nogueira, T. Mendonça, and P. Rocha, “Nonlinear controller for bispectral index tracking: Robustness and on-line retuning,” *Control Eng. Pract.*, vol. 58, pp. 343–353, 2017.
[110] F. N. Nogueira, T. Mendonça, and P. Rocha, “Controlling the depth of anesthesia by a novel positive control strategy,” *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 87–97, 2014.
[111] H. Puebla and J. Álvarez-Ramírez, “A cascade feedback control approach for hypnosis,” *Ann. Biomed. Eng.*, vol. 33, no. 10, pp. 1449–1463, 2005.
[112] M. Mahfouf, A. J. Asbury, and D. A. Linkens, “Unconstrained and constrained generalised predictive control of depth of anaesthesia during surgery,” *Control Eng. Pract.*, vol. 11, no. 12, pp. 1501–1515, 2003.
[113] A. Norberg *et al.*, “Population volume kinetics predicts retention of 0.9% saline infused in awake and isoflurane-anesthetized volunteers.,” *Anesthesiology*, vol. 107, no. 1, pp. 24–32, 2007.
[114] J. Schüttler and H. Ihmsen, “Population pharmacokinetics of propofol: A multicenter study,” *Anesthesiology*, vol. 92, no. 3, pp. 727–738, 2000.
[115] E. Kharisov, C. L. Beck, and M. Bloom, “Design of L 1 adaptive controllers for human patient anesthesia,” *Control Eng. Pract.*, vol. 44, pp. 65–77, 2015.
[116] H. H. Lin, C. L. Beck, and M. J. Bloom, “On the use of multivariable piecewise-linear models for predicting human response to anesthesia.,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 11, pp. 1876–1887, Nov. 2004.
[117] C. Beck, H. H. Lin, and M. Bloom, “Modeling and control of anesthetic pharmacodynamics,” *Lect. Notes Control Inf. Sci.*, vol. 357, pp. 263–289, 2007.
[118] R. K. Millard, C. R. Monk, and C. Prys-Roberts, “Self-tuning control of hypotension during ENT surgery using a volatile anaesthetic,” *IEE Proc. D Control Theory Appl.*, vol. 135, no. 2, pp. 95–105, 1988.
[119] T. W. Schnider *et al.*, “The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers,” *Anesthesiology*, vol. 88, no. 5, pp. 1170–1182, 1998.
[120] M. Teixeira, T. Mendonça, P. Rocha, and R. Rabiço, “Automatic control of the NMB level in general anaesthesia with a switching total system mass control strategy,” *J. Clin. Monit. Comput.*, vol. 28, no. 6, pp. 501–512, 2014.
[121] M. M. Da Silva, T. Wigren, and T. Mendonca, “Nonlinear identification of a minimal neuromuscular blockade model in anesthesia,” *IEEE Trans. Control Syst. Technol.*, vol. 20, no. 1, pp. 181–188, 2012.
[122] C. Cobelli, E. R. Carson, L. Finkelstein, and M. S. Leaning, “Validation of simple and complex models in physiology and medicine.”
[123] R. L. Summers *et al.*, “Validation of a computational platform for the analysis of the

physiologic mechanisms of a human experimental model of hemorrhage," *Resuscitation*, 2009.
[124] R. L. Hester, R. Iliescu, R. Summers, and T. G. Coleman, "Systems biology and integrative physiological modelling," *J. Physiol.*, vol. 589, no. 5, pp. 1053–1060, Mar. 2011.
[125] C. M. Friedrich, "A model qualification method for mechanistic physiological QSP models to support model-informed drug development," *CPT Pharmacometrics Syst. Pharmacol.*, vol. 5, no. 2, pp. 43–53, Feb. 2016.
[126] E. A. Patterson and M. P. Whelan, "A framework to establish credibility of computational models in biology," *Prog. Biophys. Mol. Biol.*, vol. 129, pp. 13–19, Oct. 2017.
[127] A. F. Villaverde, "Observability and Structural Identifiability of Nonlinear Biological Systems," *Complexity*, vol. 2019, 2019.
[128] A. Raue, J. Karlsson, M. P. Saccomani, M. Jirstrand, and J. Timmer, "Comparison of approaches for parameter identifiability analysis of biological systems," *Bioinformatics*, vol. 30, no. 10, pp. 1440–1448, 2014.
[129] L. Pronzalo and E. Walter, *Identification of parametric models from experimental data*. 1997.
[130] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou, "Structural Identifiability Analysis via Extended Observability and Decomposition," *IFAC-PapersOnLine*, vol. 49, no. 26, pp. 171–177, 2016.
[131] K. J. Keesman, *System Identification - An Itroduction*. 2011.
[132] D. A. Beard, "Assessing the validity and utility of the Guyton model of arterial blood pressure control," *Hypertension*, vol. 72, no. 6, pp. 1272–1273, Dec. 2018.
[133] A. Gábor, A. F. Villaverde, and J. R. Banga, "Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems," *BMC Syst. Biol.*, vol. 11, no. 1, pp. 1–16, 2017.
[134] A. F. Villaverde and J. R. Banga, "Structural Properties of Dynamic Systems Biology Models: Identifiability, Reachability, and Initial Conditions," *Processes*, vol. 5, no. 2, p. 29, 2017.
[135] P. Pathmanathan, M. S. Shotwell, D. J. Gavaghan, J. M. Cordeiro, and R. A. Gray, "Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology," *Progress in Biophysics and Molecular Biology*. 2015.
[136] M. L. Mogensen, *A Physiological Mathematical Model of the Respiratory System*. 2011.
[137] A. Raue *et al.*, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, Aug. 2009.
[138] M. P. Saccomani and K. Thomaseth, "The Union between Structural and Practical Identifiability Makes Strength in Reducing Oncological Model Complexity: A Case Study," *Complexity*, vol. 2018, pp. 1–10, 2018.
[139] A. F. Villaverde, N. D. Evans, M. J. Chappell, and J. R. Banga, "Input-dependent structural identifiability of nonlinear systems," *IEEE Control Syst. Lett.*, vol. 3, no. 2, pp. 272–277, 2019.
[140] D. Joubert, J. D. Stigter, and J. Molenaar, "Assessing the role of initial conditions in the local structural identifiability of large dynamic models," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, 2021.

[141] O. Chis, J. R. Banga, and E. Balsa-Canto, *Methods for checking structural identifiability of nonlinear biosystems: A critical comparison*, vol. 18, no. PART 1. IFAC, 2011.
[142] C. Kreutz, "An easy and efficient approach for testing identifiability," *Bioinformatics*, vol. 34, no. 11, pp. 1913–1921, 2018.
[143] S. M. Verduyn Lunel, "Parameter identifiability of differential delay equations," *Int. J. Adapt. Control Signal Process.*, vol. 15, no. 6, pp. 655–678, 2001.
[144] O. Structural, R. Bellman, L. Angeles, and K. J. Astrom, "K. J. ASTROM Division of Automatic Control Lund Institute of Technology Lund, Sv. eden," *Math. Biosci.*, vol. 7, pp. 329–339, 1970.
[145] M. Pia Saccomani, S. Audoly, G. Bellu, and L. D'Angio, "A new differential algebra algorithm to test identifiability of nonlinear systems with given initial conditions," pp. 3108–3113, Nov. 2002.
[146] T. Maiwald *et al.*, "Driving the model to its limit: Profile likelihood based model reduction," *PLoS One*, vol. 11, no. 9, pp. 1–18, 2016.
[147] M. C. Eisenberg and H. V. Jain, "A confidence building exercise in data and identifiability: Modeling cancer chemotherapy as a case study," *J. Theor. Biol.*, vol. 431, p. 63, Oct. 2017.
[148] J. Vanlier, C. A. Tiemann, P. A. J. Hilbers, and N. A. W. van Riel, "An integrated strategy for prediction uncertainty analysis," *Bioinformatics*, vol. 28, no. 8, p. 1130, Apr. 2012.
[149] W. F. Fincham and F. T. Tehrani, "A mathematical model of the human respiratory system," *J. Biomed. Eng.*, vol. 5, no. 2, pp. 125–133, 1983.
[150] A. Gábor and J. R. Banga, "Robust and efficient parameter estimation in dynamic models of biological systems," *BMC Syst. Biol.*, vol. 9, no. 1, 2015.
[151] B. Parvinian, R. Bighamian, C. G. Scully, J. O. Hahn, and P. Pathmanathan, "Credibility Assessment of a Subject-Specific Mathematical Model of Blood Volume Kinetics for Prediction of Physiological Response to Hemorrhagic Shock and Fluid Resuscitation," *Front. Physiol.*, vol. 12, Sep. 2021.
[152] Y. Liu, W. Chen, P. Arendt, and H. Z. Huang, "Toward a better understanding of model validation metrics," *J. Mech. Des. Trans. ASME*, vol. 133, no. 7, pp. 1–13, 2011.
[153] M. C. Eisenberg and M. A. L. Hayashi, "Determining identifiable parameter combinations using subset profiling," *Math. Biosci.*, vol. 256, pp. 116–126, 2014.
[154] W. L. Oberkampf and T. G. Trucano, "Verification and validation benchmarks," *Nucl. Eng. Des.*, vol. 238, no. 3, pp. 716–743, 2008.
[155] W. L. Oberkampf and C. J. Roy, "Verification and validation in scientific computing," *Verif. Valid. Sci. Comput.*, pp. 1–767, Jan. 2011.
[156] L. F. Shampine and S. Thompson, "Solving DDEs in MATLAB," *Appl. Numer. Math.*, vol. 37, no. 4, pp. 441–458, Jun. 2001.
[157] J. W. Kreit, "Volume capnography in the intensive care unit: Physiological principles, measurements, and calculations," *Ann. Am. Thorac. Soc.*, vol. 16, no. 3, pp. 291–300, 2019.
[158] "ERS Handbook of Paediatric Respiratory Medicine," *ERS Handb. Paediatr. Respir. Med.*, Jan. 2021.
[159] N. Meshkat, C. Er-zhen Kuo, and J. DiStefano, "On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation," *PLoS One*, vol. 9, no. 10, 2014.
[160] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angiò, "DAISY: A new software tool to

test global identifiability of biological and physiological systems," *Comput. Methods Programs Biomed.*, vol. 88, no. 1, pp. 52–61, 2007.
[161] R. Bighamian, C.-S. Kim, A. T. Reisner, and J.-O. Hahn, "Closed-Loop Fluid Resuscitation Control Via Blood Volume Estimation," *J. Dyn. Syst. Meas. Control*, vol. 138, no. 11, p. 111005, 2016.
[162] R. Bighamian, B. Parvinian, C. G. Scully, G. Kramer, and J. O. Hahn, "Control-oriented physiological modeling of hemodynamic responses to blood volume perturbation," *Control Eng. Pract.*, vol. 73, no. December 2017, pp. 149–160, 2018.
[163] J. E. Hall and A. C. Guyton, "Cardiovascular physiology IV," p. 371, 1982.
[164] A. D. Rafie *et al.*, "Hypotensive Resuscitation of Multiple Hemorrhages Using Crystalloid and Colloids," *Shock*, vol. 22, no. 3, pp. 262–269, 2004.
[165] N. Aladangady, T. Leung, K. Costeloe, and D. Delpy, "Measuring circulating blood volume in newborn infants using pulse dye densitometry and indocyanine green," vol. i, pp. 865–871, 2008.
[166] R. Bighamian, A. T. Reisner, and J.-O. Hahn, "A Lumped-Parameter Subject-Specific Model of Blood Volume Response to Fluid Infusion," *Front. Physiol.*, vol. 7, no. August, p. 390, 2016.
[167] T. Imai *et al.*, "Accuracy and repeatability of blood volume measurement by pulse dye densitometry compared to the conventional method using 51Cr-labeled red blood cells," *Intensive Care Med.*, 2000.
[168] R. Bighamian, M. Kinsky, G. Kramer, and J. O. Hahn, "In-human subject-specific evaluation of a control-theoretic plasma volume regulation model," *Comput. Biol. Med.*, vol. 91, no. August, pp. 96–102, 2017.
[169] R. Bighamian, A. T. Reisner, and J. O. Hahn, "A Lumped-Parameter Subject-Specific Model of Blood Volume Response to Fluid Infusion," *Front Physiol*, vol. 7, p. 390, 2016.
[170] A. E. Taylor, "Capillary fluid filtration. Starling forces and lymph flow," *Circulation Research*. 1981.
[171] V. H. Huxley and J. Scallan, "Lymphatic fluid: Exchange mechanisms and regulation," *Journal of Physiology*. 2011.
[172] R. Ramasawmy *et al.*, "Blood volume measurement using cardiovascular magnetic resonance and ferumoxytol : preclinical validation," vol. m, pp. 1–10, 2018.
[173] M. Yu *et al.*, "A prospective randomized trial using blood volume analysis in addition to pulmonary artery catheter, compared with pulmonary artery catheter alone, to guide shock resuscitation in critically ill surgical patients," *Shock*, 2011.
[174] T. A. Manzone, H. Q. Dam, D. Soltis, and V. V Sagar, "Blood volume analysis: a new technique and new clinical interest reinvigorate a classic study," *J Nucl Med Technol*, vol. 35, no. 2, pp. 79, 55–63; quiz 77, 2007.