



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

---

# Identifiability of parameters in PBPK models

Identifiability analysis using the profile likelihood method for model parameters in physiologically based pharmacokinetic models

Master's thesis in Engineering Mathematics and Computational Science

SIMON WATANABE



MASTER'S THESIS 2019:NN

# Identifiability of parameters in PBPK models

Identifiability analysis using the profile likelihood method for model parameters in physiologically based pharmacokinetic models

Simon Watanabe



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences  
*Applied Mathematics and Statistics*  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Gothenburg, Sweden 2019

Identifiability of parameters in PBPK models  
Identifiability analysis using the profile likelihood method for model parameters in  
physiologically based pharmacokinetic models  
Simon Watanabe

© Simon Watanabe, 2019.

Supervisors: Umberto Picchini, Department of Mathematical Sciences  
Teodor Erngren, AstraZeneca, Early RIA DMPK  
Rikard Johansson, AstraZeneca, Early RIA DMPK  
David Janzén, AstraZeneca, Early CVRM DMPK  
Elin Boger, AstraZeneca, Early RIA DMPK  
Examiner: David Bolin, Department of Mathematical Sciences

Master's Thesis 2019:NN  
Department of Mathematical Sciences  
Applied Mathematics and Statistics  
Chalmers University of Technology  
SE-412 96 Gothenburg  
Telephone +46 31 772 1000

Typeset in L<sup>A</sup>T<sub>E</sub>X  
Printed by [Name of printing company]  
Gothenburg, Sweden 2019

Identifiability of parameters in PBPK models  
Identifiability analysis using the profile likelihood method for model parameters in  
physiologically based pharmacokinetic models  
SIMON WATANABE  
Department of Mathematical Sciences  
Chalmers University of Technology

## Abstract

In the field of pharmacologics, physiologically-based pharmacokinetic (PBPK) models can be used for predicting the pharmacokinetics of a drug compound in the body. These models are often a system of ordinary differential equations (ODEs) that describe the transport of a drug between different compartments of the body. The models depend on several parameters, some of which cannot be measured experimentally and instead these parameters are often estimated from experimental data using maximum likelihood. However, in many applications in systems biology, estimates will suffer from unidentifiability issues, meaning that well-determined estimates cannot be inferred from the data [17].

This problem comes in two forms, structural unidentifiability and practical unidentifiability, both of which can be analyzed with the profile likelihood method developed by Raue et.al [14]. The profile likelihood method is a numerical method for calculating likelihood-based confidence intervals of the parameters, which are then used to assess identifiability.

In this project the profile likelihood method is implemented in MATLAB and used to perform identifiability analysis on key model parameters for three PBPK models using simulated data. Thus, the results of this project are both a showcase of the profile likelihood method and an analysis of the identifiability of parameters in some specific models used for pulmonary drug delivery.

The results indicate that if very precise measurements could be taken then all parameters considered would be identifiable. When a reasonable measurement error is applied on the simulated data the same is not true. Some parameters, such as the in-vivo pulmonary permeability and deposition fraction will remain identifiable, but most other parameters will suffer from practical unidentifiability. With a reasonable measurement error the identifiability of most model parameters will also be dependent on the particular error realization. To address these issues, additional data is considered by observing how the uncertainty in parameter estimates impacts observables. By this method (also suggested by Raue [14]) additional measurements are introduced in an effective manner to potentially resolve unidentifiabilities.

Keywords: structural unidentifiability, practical unidentifiability, pulmonary drug delivery, maximum likelihood estimation.



## Acknowledgements

I would like to express my gratitude towards my supervisors without whom this project would not be possible. A special thanks to Teo and Umberto for the continuous support and patience during the process of working with project. I would also like to thank Eleonor Ehrman at Dfind for putting me in touch with the people at AstraZeneca.

Simon Watanabe, Gothenburg, May 2019



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theory</b>	<b>3</b>
2.1	Simulation of data . . . . .	3
2.2	Estimation . . . . .	4
2.3	Confidence intervals . . . . .	5
2.4	Identifiability . . . . .	6
2.4.1	Structural unidentifiability . . . . .	7
2.4.2	Practical unidentifiability . . . . .	7
<b>3</b>	<b>Methods</b>	<b>9</b>
3.1	Global estimation . . . . .	9
3.1.1	Sensitivity equations . . . . .	10
3.2	Profile likelihood . . . . .	11
3.3	Determining additional measurement points . . . . .	13
<b>4</b>	<b>Implementation</b>	<b>15</b>
4.1	ODE-systems . . . . .	15
4.2	Optimizers . . . . .	15
4.3	Profile likelihood . . . . .	16
4.3.1	Adaptive step size . . . . .	16
4.3.2	Constrained optimization . . . . .	18
4.3.3	Start point . . . . .	18
4.3.4	Flowchart . . . . .	19
4.3.5	Confidence intervals . . . . .	20
4.4	Test problem . . . . .	20
<b>5</b>	<b>Models</b>	<b>25</b>
5.1	Small molecules . . . . .	25
5.1.1	Neutral molecules . . . . .	27
5.1.2	Basic molecules . . . . .	29
5.2	Large molecules . . . . .	31
5.2.1	Reflection coefficients . . . . .	36
5.3	Dosing . . . . .	37
5.4	Experimental measurements . . . . .	37
5.4.1	Additional observables . . . . .	38

5.5	Parameters	39
5.5.1	Parameter ranges	40
5.5.2	Priors	41
<b>6</b>	<b>Results</b>	<b>43</b>
6.1	Neutral molecules model	44
6.2	Basic molecules model	53
6.2.1	Simplified case	53
6.2.2	Unique case	56
6.3	Large molecules model	58
6.3.1	Simplified case: $\sigma_{v,tb} \equiv \sigma_{v,org}$	59
6.3.2	Unique case	66
<b>7</b>	<b>Discussion and Conclusion</b>	<b>69</b>
	<b>Bibliography</b>	<b>73</b>
<b>A</b>	<b>Test problems</b>	<b>I</b>
A.1	Problem 1: A structurally unidentifiable model	I
A.2	Problem 2	IV
<b>B</b>	<b>Calculation of reasonable measurement error</b>	<b>VII</b>
<b>C</b>	<b>Neutral molecules</b>	<b>X</b>
C.1	Analysis for Low measurement error	X
C.2	Additional Figures and Tables - Reasonable measurement error	XV
<b>D</b>	<b>Basic molecules</b>	<b>XXIV</b>
D.1	Simplified case - Low measurement error	XXIV
D.2	Additional Figures and Tables - Reasonable measurement error	XXIX
<b>E</b>	<b>Large molecules</b>	<b>XXXI</b>
E.1	Simplified case - Low measurement error	XXXI
E.2	Additional Figures and Tables - Simplified case	XXXV
E.3	Unique case - Low measurement error	XXXVII
E.4	Additional Figures and Tables - Unique case	XLI

# 1

## Introduction

In pharmacological research, the ability to make useful and accurate predictions of the pharmacokinetics for drug-compounds is of importance. To do this one can use physiologically based pharmacokinetic (PBPK) modeling to define a set of equations that describe the concentration of the compound in different compartments of the body and the simultaneous transport between these compartments. The set of equations are often a system of non-linear ordinary differential equations (ODEs). Although many of the model parameters can be found in literature or can be measured in vitro/in vivo, some needs to be estimated from data, commonly with the use of maximum likelihood.

In order for the estimate to be well-determined from the measured data, it needs to be identifiable, but in many cases for large and complex models such as the ones used in pre-clinical assessment of drug-compounds in pharmacology, this is not the case[11]. Instead, estimates suffer from unidentifiability of two forms: *structural* and *practical*[14][15][11]. *Structural unidentifiability* is a property of the model structure for a given set of input and output functions, and if it is present, a subset of the model parameters cannot be inferred from data, regardless of quality and size of the data. *Practical unidentifiability* occurs when model parameters cannot be determined due to insufficient amount and/or quality of available data but could be estimated in principle. Factors that affects this includes noise level of the measurement errors, biological variability, study occasion variability, number of measurements and when and what is measured.

Obtaining reliable parameter estimates is important to make reliable predictions from any model and possibly better understand the phenomenon under study. The main goal of this project is therefore to assess the identifiability of key model parameters for three different PBPK models. In other words, the question we ask is “can we obtain well-determined and reliable estimates of these parameters given a model and data?”.

The models were developed by Teodor Erngren and Elin Boger at AstraZeneca (Early RIA DMPK). To address the above question simulated data was used, meaning that the analysis in this thesis provides an a priori assessment of parameter identifiability that can be used to guideline future projects. The simulated data should reflect plausible experimental data, this of course relies on the assumption that the

models used here are suitable to describe the biological process which is measured, but it also depends on defining a reasonable noise distribution that incorporates the measurement error for real life experiments. To obtain plausible simulated data we therefore use an experimental data set from [9] to calculate reasonable measurement errors.

Our objective is to see which key model parameters can be estimated for a given model and a simulated data set. To be able to do this we make use of the profile likelihood method, developed mainly by Raue et.al. [14]. The profile likelihood method is a numerical method for calculation of confidence intervals of the estimates which then determine if they are identifiable. Moreover, it gives information about the nature of the unidentifiability (when present) as in *practical* or *structural* and it is possible to extract what additional information is needed to possibly resolve any unidentifiabilities. The main motivating factor for using this method is that it is an easy-to-interpret numerical method, as when the models are complex and large it is not often possible or reasonable to make analytical calculations. Furthermore, it makes fewer assumptions than most other existing methods to detect non-identifiability and allows for asymmetrical confidence intervals[14]. The profile likelihood method is implemented in MATLAB r2017b as a part of this project.

The three models considered will represent different types of molecules, where we have two models for small molecules and one for large molecules. For the small molecules, one model represents neutral molecules and one basic molecules. The overall structure of all models are similar and are based on the PBPK model for lung deposition developed by Boger et.al. in [5]. It is of note that no model development was considered in this project, the models were given and no changes were made to them.

# 2

## Theory

This chapter introduces the definition of the likelihood based confidence intervals, which in turn is used to define the identifiability of a parameter. To define these concepts, an overview of the mathematical models, later defined in chapter 5, with some useful notations is given first.

The ODE systems describing the three models that will be considered (given by equations (5.2)-(5.28)) can be summarized as

$$\frac{d\mathbf{x}(t; \boldsymbol{\lambda})}{dt} = \mathbf{f}(\mathbf{x}(t; \boldsymbol{\lambda}), \boldsymbol{\lambda}), \quad (2.1)$$

$$\mathbf{x}(0, \boldsymbol{\lambda}) = \mathbf{x}_0(\boldsymbol{\lambda}) \quad (2.2)$$

where  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_P]$  is the vector of parameters considered for estimation (defined later in section 5.5) with length  $P = \#\text{parameters}$  and  $\mathbf{x}(t; \boldsymbol{\lambda})$  are the trajectories of the unobserved states of the system evaluated with parameter values set to  $\boldsymbol{\lambda}$ .  $P$  in our case is mostly in the range of 6-10. The function  $\mathbf{f}$  is continuous and differentiable w.r.t. the parameters  $\boldsymbol{\lambda}$ . Moreover, it is generally a non-linear function of the unobserved states  $\mathbf{x}$ .

The states in  $\mathbf{x}$  cannot be observed directly, instead the observables are given by functions  $g_k(\mathbf{x}(t; \boldsymbol{\lambda}))$ ,  $k = 1, \dots, K$  (in this case given by equations (5.29)-(5.34)). For example,  $g_k(\mathbf{x}(t; \boldsymbol{\lambda}))$  can be a linear combination of some of the states in  $\mathbf{x}$ . Furthermore, we assume that only noisy versions of the observables are possible to measure. The data, denoted by  $\mathbf{y}$ , is then given by discrete noisy measurements of  $g_k(\mathbf{x}(t; \boldsymbol{\lambda}))$ ,  $k = 1, \dots, K$ , where  $K < \dim(\mathbf{x})$ .

If some parameter transform is used, e.g. log-parametrization then  $\boldsymbol{\lambda}$  is replaced by  $\log(\boldsymbol{\lambda})$ . In our case,  $\log_{10}$ -parametrization was used since all parameters are assumed to be positive and to simplify the notation  $\boldsymbol{\theta} \equiv \log_{10}(\boldsymbol{\lambda})$  is introduced.

### 2.1 Simulation of data

Simulation of data is achieved by first solving the system of ODEs in equation (2.1) with  $\boldsymbol{\theta}$  fixed to some true values  $\boldsymbol{\theta}_{true}$ . The solution  $\mathbf{x}(t; \boldsymbol{\theta}_{true})$  is then used to produce

simulated measurements. To do this the proper distribution of measurement errors needs to be established. There are two common ways to do this by either Gaussian or log-normal error distributions.

With additive Gaussian errors, measurements for observable  $k$  at time  $t_i$  is given by

$$y_k(t_i) = g_k(\mathbf{x}(t_i; \boldsymbol{\theta}_{true})) + \epsilon_{ki} \quad (2.3)$$

where  $\epsilon_{ki}$  are independent and identically distributed Gaussian error terms with standard deviations  $\sigma_{ki}$  as

$$\epsilon_{ki} \sim \mathcal{N}(0, \sigma_{ki}), \quad k = 1, \dots, K, \quad i = 1, \dots, N_k \quad (2.4)$$

$K$  is the number of observables and  $N_k$  is the number of time points for observable  $k$ .

As will be shown later in chapter 5, the observables in our case are concentrations and thus non-negative, which implies the use of log-normal measurement error for the data  $\mathbf{y}$ , where

$$\mathbf{y} = [y_1(t_1), y_1(t_2), \dots, y_{k-1}(t_{N_{k-1}}), y_k(t_1), \dots, y_K(t_{N_K})]$$

as

$$\log(y_k(t_i)) \sim \mathcal{N}(\log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta}_{true}))), \sigma_{ki}) \quad (2.5)$$

which is to say that

$$\begin{aligned} \log(y_k(t_i)) &= \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta}_{true}))) + \epsilon_{ki} \\ \epsilon_{ki} &\sim \mathcal{N}(0, \sigma_{ki}), \end{aligned} \quad (2.6)$$

or equivalently

$$\begin{aligned} y_k(t_i) &= g_k(\mathbf{x}(t_i; \boldsymbol{\theta}_{true})) \cdot e^{\epsilon_{ki}} = g_k(\mathbf{x}(t_i; \boldsymbol{\theta}_{true})) \cdot e_{ki} \\ e_{ki} &\sim \text{log}\mathcal{N}(0, \sigma_{ki}), \quad k = 1, \dots, K, \quad i = 1, \dots, N_k. \end{aligned} \quad (2.7)$$

In this project only simulated measurement data was considered, and the standard deviations  $\sigma_{ki}$  were assumed to be known, i.e.  $\sigma_{ki}$  were not considered as parameters for estimation.

## 2.2 Estimation

Given a data set of  $y_k(t_i)$ ,  $k = 1, \dots, K$ ,  $i = 1, \dots, N_k$  and assuming Gaussian errors the likelihood is formulated as

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{k=1}^K \prod_{i=1}^{N_k} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp\left(-\frac{1}{2}\left(\frac{y_k(t_i) - g_k(\mathbf{x}(t_i; \boldsymbol{\theta}))}{\sigma_{ki}}\right)^2\right) \quad (2.8)$$

or for log-normal errors

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{k=1}^K \prod_{i=1}^{N_k} \frac{1}{y_k(t_i)} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp\left(-\frac{1}{2}\left(\frac{\log(y_k(t_i)) - \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}}\right)^2\right). \quad (2.9)$$

To estimate the parameters  $\boldsymbol{\theta}$ , maximum likelihood is applied[14][10][15][11], i.e. solving the optimization problem

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{y}) \quad (2.10)$$

but in practice it is often easier to solve the equivalent problem of minimizing the negative log-likelihood as

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} [-2 \log(L(\boldsymbol{\theta}|\mathbf{y}))] \quad (2.11)$$

where the factor 2 is multiplied to simplify the expression in the following equations. For equations (2.8) and (2.9) this leads to

$$-2 \log(L(\boldsymbol{\theta}|\mathbf{y})) = \sum_{k=1}^K \sum_{i=1}^{N_k} \left(\frac{y_k(t_i) - g_k(\mathbf{x}(t_i; \boldsymbol{\theta}))}{\sigma_{ki}}\right)^2 - \text{const.} \quad (2.12)$$

and

$$-2 \log(L(\boldsymbol{\theta}|\mathbf{y})) = \sum_{k=1}^K \sum_{i=1}^{N_k} \left(\frac{\log(y_k(t_i)) - \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}}\right)^2 - \text{const.} \quad (2.13)$$

respectively. In this thesis mostly log-normally distributed errors are considered and therefore the following definition is introduced:

$$\xi^2(\boldsymbol{\theta}) \equiv \sum_{k=1}^K \sum_{i=1}^{N_k} \left(\frac{\log(y_k(t_i)) - \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}}\right)^2. \quad (2.14)$$

Thus the parameter estimates are found by solving the optimization problem

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \xi^2(\boldsymbol{\theta}). \quad (2.15)$$

## 2.3 Confidence intervals

The main objective of this thesis is to analyze the identifiability of parameters using the profile likelihood method. To do this, the definition of finite-sample confidence intervals (also called likelihood based confidence intervals) are introduced here [14][15][10]. These are defined by a confidence region in the parameter space where the likelihood, or equivalently  $\xi^2$ , stays within a threshold  $\Delta_\alpha$ . The definition of the confidence region is

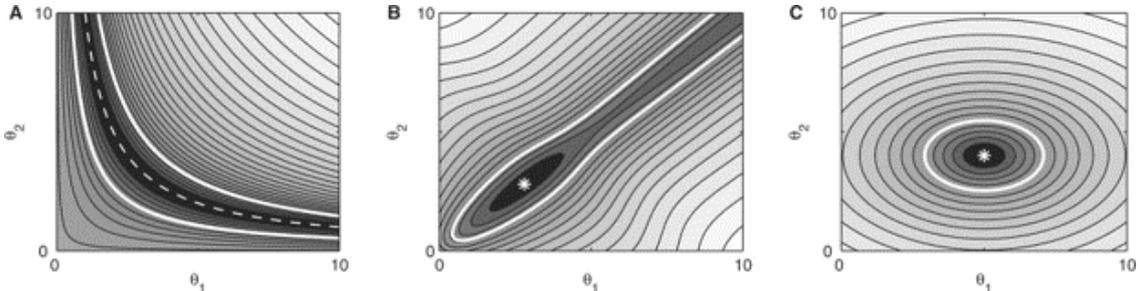
$$\{\boldsymbol{\theta} | \xi^2(\boldsymbol{\theta}) - \xi^2(\hat{\boldsymbol{\theta}}) < \Delta_\alpha\}, \quad \text{with } \Delta_\alpha = \chi^2(\alpha, df). \quad (2.16)$$

That is, all parameter values  $\boldsymbol{\theta}$  such that  $\xi^2(\boldsymbol{\theta})$  is within  $\Delta_\alpha$  of the optimum  $\xi^2(\hat{\boldsymbol{\theta}})$ . The confidence intervals are then given by the borders of the confidence region. Here  $\Delta_\alpha$  is the  $1 - \alpha$  quantile of the  $\chi^2$  distribution and  $df = 1$  and  $df = P$  represents point-wise and simultaneous confidence intervals respectively. In this thesis only point-wise confidence intervals for the parameter estimates is considered.

With enough data and a suitable model the likelihood will be informative and have a sharp peak at  $\hat{\boldsymbol{\theta}}$  which leads to a small restricted confidence region and consequently narrow confidence intervals. With less informative data, as in less measurements or higher measurement error, the likelihood will also be less informative and more dispersed, hence leading to broader confidence intervals or even infinitely broad confidence intervals. The nature of the confidence intervals defines the identifiability of the parameters, which are explained in the following section.

## 2.4 Identifiability

For a parameter to be identifiable it must have a finite confidence interval  $[c^-, c^+]$ , with  $c^-, c^+ \in \mathbb{R}$  [14]. For a two-dimensional example see right panel of Figure 2.1. In the case of unidentifiability a distinction is made between structural and practical unidentifiability.



**Figure 2.1:** Contour plots of  $\xi^2(\boldsymbol{\theta})$  for a two-dimensional parameter space, shown on non-logarithmic scale for illustrative purposes. Shades from black to white correspond to low and high values of  $\xi^2$ , respectively. Thick white lines display likelihood-based confidence regions and white stars the optimal parameter estimates  $\hat{\boldsymbol{\theta}}$ . Left panel: a structural non-identifiability along the functional relation  $\mathbf{h}(\boldsymbol{\theta}) = \theta_1 \cdot \theta_2 - 10 = 0$  (dashed line). The likelihood-based confidence region is infinitely extended. Middle panel: a practical non-identifiability. The likelihood-based confidence region is infinitely extended for  $\theta_1 \rightarrow \infty$  and  $\theta_2 \rightarrow \infty$ , lower confidence bounds can be derived. Right panel: both parameters identifiable. Figure taken from *Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood*, with permission from A.Raue and Bioinformatics Journal.

### 2.4.1 Structural unidentifiability

Structural unidentifiability is independent of the amount and quality of data and is a property of the model structure for a given set of input and output functions. Thus, given a perfect data set, any structurally unidentifiable parameters are still not well-determined. This may be due to ambiguous parameterisation which can be characterized as functional relations  $\mathbf{h}(\boldsymbol{\theta}_{sub}) = 0$ , between a subset of parameters  $\boldsymbol{\theta}_{sub} \subset \boldsymbol{\theta}$  where the observables remains unchanged[14][15][11], i.e.

$$\mathbf{h}(\boldsymbol{\theta}_{sub}) = 0 \iff g_k(\mathbf{x}(t, \boldsymbol{\theta})) = g_k(\mathbf{x}(t, \hat{\boldsymbol{\theta}})), \quad \forall k = 1, \dots, K, \quad \forall t \quad (2.17)$$

The result is that all parameters  $\theta_j \in \boldsymbol{\theta}_{sub}$  are structurally unidentifiable since varying the parameters according to the functional relations  $\mathbf{h}$  makes no difference in the observed values  $g_k(t_i, \boldsymbol{\theta})$  and hence the value of  $\xi^2(\boldsymbol{\theta})$  is kept constant. Therefore an infinite range of parameter values can be found along the functional relations that are equally good and confidence intervals for these parameters are infinitely wide. Another example of structural unidentifiability occurs when the input/output functions are such that certain parts of the model structure doesn't affect the observables. Parameters incorporated in these parts of the model structure are then structurally unidentifiable. In a two-dimensional parameter space a structural unidentifiability can be visualized as a perfectly flat valley, infinitely extended along the functional relationship  $\mathbf{h}$ , see for example left panel in figure 2.1. The remedy for structural unidentifiability would be to measure additional states, i.e. increasing the number of observables and/or perturb the system in another way by changing the input for the model[14][15].

### 2.4.2 Practical unidentifiability

A parameter estimate is practically unidentifiable if the confidence interval extends infinitely in either direction but the likelihood has a unique optimum[14][15][11]. This means that the parameter value of a practically unidentifiable parameter can be infinitely increased and/or decreased without exceeding the threshold  $\Delta_\alpha$ , but unlike structural unidentifiability an optimum exists. The confidence interval in this case need not be infinite in both directions as the parameter estimate can have lower or upper bounds but not both, see for example middle panel in figure 2.1 where a lower bound can be obtained. In a two-dimensional parameter space practical unidentifiability can be visualized as an infinitely extended relatively flat valley along which  $\xi^2(\boldsymbol{\theta})$  never exceeds the threshold  $\Delta_\alpha$ . As with structural unidentifiability, measurements of additional states can introduce more information about the parameter to remedy the problem, but since practical unidentifiability arises due to insufficient data, it is also possible to remedy this by increasing the amount of measurements of the current observables or reduce the measurement error for the current measurements[15][14].



# 3

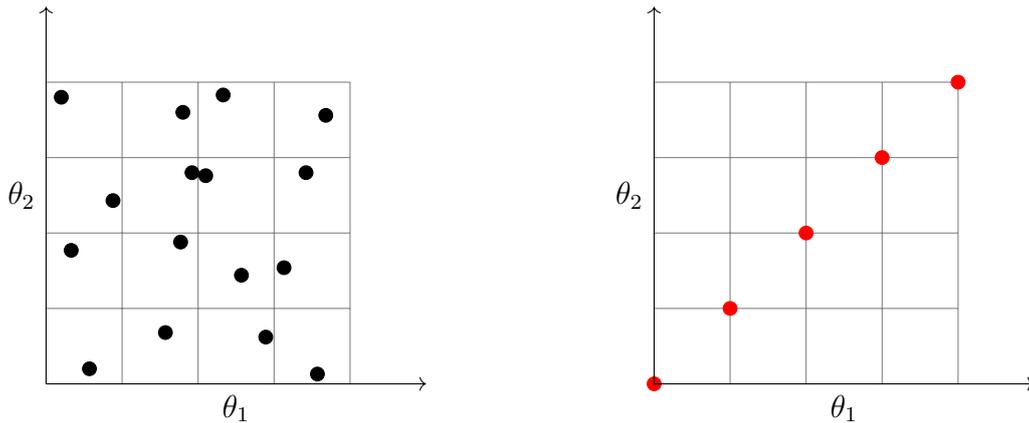
## Methods

To calculate the likelihood-based confidence intervals the profile likelihood method will be used[14]. This method is based on exploring the likelihood-based confidence region in the direction of least increase, starting from the optimal parameter values. Thus, for unidentifiable parameters it will explore the parameter space along the infinitely extended relations where  $\xi^2(\boldsymbol{\theta})$  never exceeds the threshold. The optimum is found by global optimization of the problem defined in equation (2.15).

### 3.1 Global estimation

The global estimation to find the optimum of  $\xi^2(\boldsymbol{\theta})$  is performed first and this estimate is then used as starting point for the profile likelihood calculation. Since strictly speaking even so called global algorithms are local, as otherwise all common optimization methods only ensure a local optimum is found, the optimization algorithm is started from  $M$  different start points. The integer  $M$  is chosen sufficiently large to ensure the global optimum is found from at least one of the start points. Moreover, since all parameters are  $\log_{10}$ -transformed this greatly improves optimization efficiency and also ensures positivity of all parameters.

The start points should be chosen within reasonable ranges of the parameter values. For  $P$  parameters, this can be achieved by calculating a  $P$ -dimensional grid over the parameter space bounded by parameter ranges and randomly choosing one start point in each box. For a 2-dimensional example see left panel of figure 3.1. The easiest way of doing this assumes that all axes are divided into equal number of intervals  $N_d$ . However, this method leads to the number of start points increasing exponentially with the number of parameters as  $N_d^{\#\text{parameters}}$ , and the number of start points gets unreasonably large for 6-10 parameters, even with a low  $N_d$ . Instead, only taking the diagonal elements in the  $P$ -dimensional grid as start points leads to the number of start points being equal to  $N_d$ . To further simplify things the diagonal crossings are chosen as start points, instead of randomly choosing start points inside each diagonal box. Then for parameter ranges defined by lower and upper limits,  $\boldsymbol{l}$



**Figure 3.1:** How multiple start points are chosen for the global optimization. The figures are presented on non-logarithmic scale for illustrative reasons. Left: Start points taken randomly inside each box of the grid. Right: Start points taken on the diagonal crossings of the grid. The right side method leads to considerably fewer start points but does not explore the parameter space as thoroughly as the left side method. The number of start points for the left method on the other hand increases exponentially with number of parameters.

and  $\mathbf{u}$ , the start points  $\theta_{start,i}$ ,  $i = 1, \dots, M$  can be calculated by

$$\theta_{start,i} = \log_{10} \mathbf{l} + \frac{i-1}{M-1} \cdot (\log_{10} \mathbf{u} - \log_{10} \mathbf{l}), \quad i = 1, \dots, M \quad (3.1)$$

and the number of start points is then equal to  $N_d + 1$ , for an illustration of this see right panel of figure 3.1. For all results in this thesis equation (3.1) was used to calculate start points for the global estimation.

### 3.1.1 Sensitivity equations

For most deterministic optimization approaches the derivatives of the objective function  $\xi^2(\boldsymbol{\theta})$  (given by equation (2.14)) with respect to the parameters  $\theta_p$  are needed[4]. To obtain these the sensitivities  $\frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p}$  needs to be calculated, which is usually done in one of two ways:

- By solving the sensitivity equations, given by the ODEs

$$\frac{d}{dt} \frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p} + \frac{\partial \mathbf{f}}{\partial \theta_p} \quad (3.2)$$

- or by finite difference approximation,

$$\frac{d\mathbf{x}(t; \boldsymbol{\theta})}{d\theta_p} \approx \frac{\mathbf{x}(t; \boldsymbol{\theta}) - \mathbf{x}(t; \boldsymbol{\theta} + h \cdot \mathbf{e}_p)}{h} \quad (3.3)$$

Finite difference approximation is already implemented in most software, including MATLABs *lsqnonlin*, and can be used for any general problem. However, the approach using sensitivity equations is much faster and gives more reliable results and hence this is what will be used [4].

The sensitivities  $\frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p}$  describe the dependency of the state variables on the parameter values, and by taking the time derivative and applying the chain rule and Clairauts theorem equation (3.2) is obtained:

$$\frac{d}{dt} \frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p} = \frac{\partial}{\partial \theta_p} \frac{d\mathbf{x}(t; \boldsymbol{\theta})}{dt} = \frac{\partial}{\partial \theta_p} \mathbf{f}(\mathbf{x}(t; \boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\mathbf{x}(t; \boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p} + \frac{\partial \mathbf{f}(\mathbf{x}(t; \boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (3.4)$$

These ODEs can then be solved together with the rest of the ODE-system. An efficient implementation of this using the CVODES solver is the AMICI (Advanced Multilanguage Interface for CVODES and IDAS) implementation used in this thesis [8]. When the solution  $\frac{\partial \mathbf{x}(t; \boldsymbol{\theta})}{\partial \theta_p}$  has been found, the derivatives of the objective function  $\frac{\partial \xi^2(\boldsymbol{\theta})}{\partial \theta_p}$  can be found by application of the chain rule

$$\begin{aligned} \frac{\partial \xi^2(\boldsymbol{\theta})}{\partial \theta_p} &= \frac{\partial}{\partial \theta_p} \sum_{k=1}^K \sum_{i=1}^{N_k} \left( \frac{\log(y_k(t_i)) - \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}} \right)^2 = \\ &- 2 \sum_{k=1}^K \sum_{i=1}^{N_k} \left( \frac{\log(y_k(t_i)) - \log(g_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}} \right) \cdot \frac{1}{g_k(\mathbf{x}(t_i; \boldsymbol{\theta})) \sigma_{ki}} \frac{\partial g_k(\mathbf{x}(t_i; \boldsymbol{\theta}))}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t, \boldsymbol{\theta})}{\partial \theta_p}. \end{aligned} \quad (3.5)$$

Furthermore, with AMICI it is possible to obtain the derivatives  $\frac{\partial g_k(\mathbf{x}(t_i; \boldsymbol{\theta}))}{\partial \theta_p}$  directly, skipping the last application of the chain rule in equation (3.5).

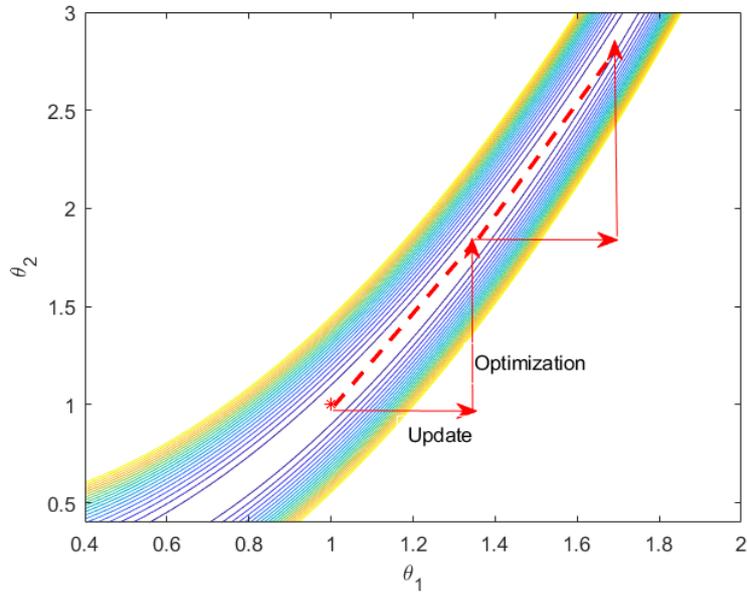
## 3.2 Profile likelihood

To obtain the desired confidence intervals and perform identifiability analysis the profile likelihood (PL) approach suggested by Raue et.al is utilized [14]. The general idea of this method is to explore the likelihood in the direction of least increase by fixating one parameter at a time and re-optimizing the remaining parameters, conditionally to the value of the fixed one.

The PL for  $\theta_j$ , denoted by  $\xi_{PL}^2(\theta_j)$ , is calculated by iteratively solving the constrained optimization problem

$$\xi_{PL}^2(\theta_j) = \min_{\theta_{p \neq j}} [\xi^2(\boldsymbol{\theta})], \quad (3.6)$$

with  $\theta_j$  fixated to different values. Starting at the global optimum  $\theta_j = \hat{\theta}_j$  the PL is calculated analogously in increasing and decreasing direction of  $\theta_j$ -values, see figure 3.2 for an illustration in two-dimensions.



**Figure 3.2:** Illustration of PL calculation for parameter  $\theta_1$  in increasing direction. The parameter value of  $\theta_1$  is updated and then constrained optimization is performed with  $\theta_1$  fixated to its updated value. The dashed line is the PL. The calculation in decreasing direction is analogous.

The PL re-optimizes for all parameters except  $\theta_j$  and hence explores  $\xi^2$  in the direction of least increase w.r.t. all other parameters conditionally on  $\theta_j$  fixed to its current value. Consequently for any structurally unidentifiable parameters  $\theta_j \in \boldsymbol{\theta}_{sub} \subset \boldsymbol{\theta}$  the profile likelihood is flat, as the direction of least increase would be along the functional relations  $\mathbf{h}(\boldsymbol{\theta}_{sub}) = 0$ . For practically unidentifiable parameters the profile likelihood will have a minimum, but it will never exceed the threshold  $\Delta_\alpha$  in increasing and/or decreasing direction. If a parameter is identifiable the profile likelihood exceeds  $\Delta_\alpha$  in both increasing and decreasing direction and these points of passover represent the likelihood based confidence intervals. It has been shown that the profile likelihood explores the likelihood in the desired way to detect unidentifiabilities[14].

The PL given in equation (3.6) is a one-dimensional function, but each step of calculating  $\xi_{PL}^2(\theta_j)$  is generally a multi-dimensional problem. Some notation is now introduced to make this clearer. Let  $l = 1, \dots, L$  denote the iterations in calculation of profile likelihood for a parameter  $\theta_j$ . Then, this results in a matrix of parameter values

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^l, \dots, \boldsymbol{\theta}^L], \quad (3.7)$$

where each  $\boldsymbol{\theta}^l$  is a solution to the constrained optimization problem in equation (3.6) with  $\theta_j^l$  fixated to some value  $C_l$ , which are monotonously increasing as  $C_1 < \dots < C_l < \dots < C_L$ . The corresponding objective function values

$$\boldsymbol{\Xi}^2 = [\xi^2(\boldsymbol{\theta}^1), \dots, \xi^2(\boldsymbol{\theta}^l), \dots, \xi^2(\boldsymbol{\theta}^L)] \quad (3.8)$$

determines whether or not the parameter is identifiable or practically/structurally unidentifiable. The function  $\xi_{PL}^2(\theta_j)$  is then defined as the mapping  $\mathbb{R} \rightarrow \mathbb{R}_+$  of  $\theta_j$  values in  $\Theta$  to objective function values in  $\Xi$ .

### 3.3 Determining additional measurement points

Let the PL calculation of a parameter result in  $\Theta$  as in equation (3.7). Then the trajectories of observables given by functions  $g_k$  along the PL of this parameter, that is

$$g_k(\mathbf{x}(t; \boldsymbol{\theta}^l)), \quad l = 1, \dots, L, \quad (3.9)$$

can give information about what new measurements are needed in the case of unidentifiability[14][15].

For structural unidentifiability the observables  $g_k$  are unaffected by any change in the parameter along the functional relationship  $\mathbf{h}(\boldsymbol{\theta}_{sub})$ . Since the PL calculation explores the parameter space in the least increase of  $\xi^2$ , the parameter values  $\Theta$  along the PL of the structurally unidentifiable parameters are such that

$$\mathbf{h}(\boldsymbol{\theta}_{sub}^l) = 0, \quad l = 1, \dots, L \quad (3.10)$$

and

$$g_k(\mathbf{x}(t; \boldsymbol{\theta}^1)) = \dots = g_k(\mathbf{x}(t; \boldsymbol{\theta}^l)) = \dots = g_k(\mathbf{x}(t; \boldsymbol{\theta}^L)), \quad \forall t. \quad (3.11)$$

Hence, the trajectories  $g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))$  along the PL show no variation. Thus the structural unidentifiability is independent of accuracy and amount of data, even with an infinite amount of perfectly measured data points the parameter value can be perturbed along  $\mathbf{h}$  without affecting the observed states  $g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))$ . The only remedy is therefore to introduce qualitatively new measurements, e.g. by measuring some of the previously unobserved model states[15].

However, for a practically unidentifiable parameter the uncertainty in the parameter impacts the trajectories  $g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))$ ,  $l = 1, \dots, L$  and they will reveal some variability. The spots with largest variability in the trajectories can be thought of as places where the uncertainty in the parameter affects the observables the most and additional measurements at these places are likely to efficiently introduce information about the parameter compared to measurements at places with less variability[14]. To identify these places easily one can calculate a measure of variability for the trajectories  $g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))$  at each time point. In the following results the range of log of observables will be used as

$$\text{range}(\log(g_k(\mathbf{x}(t)))) = \max_l \log(g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))) - \min_l \log(g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))). \quad (3.12)$$

If the parameter values  $\boldsymbol{\theta}^l$  along the PL of some practically unidentifiable parameter gives rise to large variability in the trajectories  $g_k(\mathbf{x}(t; \boldsymbol{\theta}^l))$  at some time interval,

then the range in equation (3.12) will also be large. Additional measurements at time intervals where equation (3.12) is large are then most likely to effectively resolve practical unidentifiability. Moreover, since the range is calculated on the logarithm of the trajectories it can be compared to the log-normal measurement error. If the variability given by equation (3.12) is much smaller than the measurement error at all times, then any new information introduced by additional measurements will be outweighed by the measurement error.

# 4

## Implementation

All programming was done in MATLAB R2017b.

### 4.1 ODE-systems

The implementation of the ODE-systems in MATLAB was given at the start of the project. However, due to the complexity and size of the models, a more efficient ODE-solver than the standard solvers provided by MATLAB was used. The ODE-solver used in this thesis was the CVODES-solver which can be accessed through MATLAB with AMICI and as previously mentioned it also gives an efficient implementation for solving the sensitivity equations[8][1]. To use AMICI with MATLAB, the ODEs need to be defined using symbolic variables but otherwise no changes need to be made.

### 4.2 Optimizers

For the global optimization the MATLAB function *lsqnonlin* with the *trust-region-reflective* algorithm was used unless otherwise noted. For problems where all parameters are unbounded the same function with the *levenberg-marquardt* algorithm is recommended. An implementation for using the function *fmincon* with the models in this project has been made as well, but it is significantly slower than *lsqnonlin* for most cases. All these functions are deterministic algorithms. Moreover, *lsqnonlin* is a non-linear least square solver and therefore suitable for the problems considered in this thesis. For more information about these functions see MATLABs online documentation[3]. A version of particle swarm optimization (PSO) has also been implemented as a global optimizer and can be used when *lsqnonlin* gives unsatisfactory results. PSO is a stochastic optimization algorithm and therefore doesn't guarantee that the point estimate found is an optimum but is more efficient in exploring the parameter space. Another advantage of PSO is that it doesn't get stuck as easily in local optima like the other algorithms. PSO is however often slower to use.

As mentioned in section 3.1 the global optimization relies on multiple start points for local algorithms. To perform global optimization using multiple start points the MATLAB function *multistart* was used with custom start points set as described in section 3.1.

### 4.3 Profile likelihood

The profile likelihood for each parameter estimate  $\theta_j$  is calculated by iteratively solving the constrained optimization problem in equation (3.6) with  $\theta_j$  fixed to monotonously increasing/decreasing values starting from the global estimate  $\hat{\theta}_j$ . Let the current iteration be denoted by  $l$  and consider calculation of PL in increasing direction, then the algorithm consists of two repeated steps.

1. Take an incremental step  $\theta_{step,j}^l$  in increasing direction of  $\theta_j$  conditionally on all other parameters kept fixed. That is, calculate  $C_l = \theta_j^{l-1} + \theta_{step,j}^l$ .
2. Solve the constrained optimization problem

$$\begin{aligned} \min_{\boldsymbol{\theta}} \xi^2(\boldsymbol{\theta}) \\ \text{s.t. } \theta_j = C_l \end{aligned} \tag{4.1}$$

and set  $\boldsymbol{\theta}^l$  as the solution.

Repeat until  $\xi^2(\boldsymbol{\theta}^l) - \xi^2(\hat{\boldsymbol{\theta}})$  exceeds the threshold  $\Delta_\alpha$ ,  $C_l$  exceeds a upper/lower bound of  $\theta_j$  or maximum number of steps is reached. The calculation in decreasing direction is analogous, with  $C_l = \theta_j^{l-1} - \theta_{step,j}^l$  in step 1. To do this efficiently the step sizes  $\theta_j^l$  at each iteration needs to be chosen in an adaptive manner and a well-performing optimization routine to solve equation (4.1) has to be chosen.

The profile likelihood  $\xi_{PL}^2(\theta_j)$  as given in equation (3.6) is then the function  $\mathbb{R} \rightarrow \mathbb{R}_+$  given by the solutions to equation (4.1).

#### 4.3.1 Adaptive step size

The step size in  $\theta_j$  should be taken such that the increase in  $\xi^2$  is sufficiently small to obtain a smooth profile but large enough to be efficient. Since the value of the profile likelihood is not known until the optimization problem in equation (4.1) is solved it is assumed that the increase in  $\xi^2$  is small for small steps. Let  $\boldsymbol{\theta}_{step}^l$  be a vector of same length as  $\boldsymbol{\theta}^l$  with the  $j$ th element set to  $\theta_{step,j}^l$ , i.e.  $\boldsymbol{\theta}_{step}^l = [0, \dots, \theta_{step,j}^l, \dots, 0]$ , then we assume that:

$$\xi_{PL}^2(\theta_j^{l-1} + \theta_{step,j}^l) \approx \xi^2(\boldsymbol{\theta}^{l-1} + \boldsymbol{\theta}_{step}^l) \quad \text{when } \|\boldsymbol{\theta}_{step}^l\| = |\theta_{step,j}^l| \text{ is small,} \tag{4.2}$$

Using the assumption in equation (4.2) the step size  $|\theta_{step,j}^l|$  should be chosen such that it fulfills the condition

$$\xi^2(\boldsymbol{\theta}^{l-1} + \boldsymbol{\theta}_{step}^l) - \xi^2(\boldsymbol{\theta}^{l-1}) \approx q \cdot \Delta_\alpha \quad (4.3)$$

where  $q \in (0, 1]$  [14]. This way at least  $\frac{1}{q}$  steps are needed in each direction for identifiable parameters. However, for unidentifiable parameters the likelihood is almost flat and the condition is never fulfilled. Moreover if local minimas exist the condition is unsuitable as  $\xi^2(\boldsymbol{\theta}^{l-1} + \boldsymbol{\theta}_{step}^l) < \xi^2(\boldsymbol{\theta}^{l-1})$  when  $\boldsymbol{\theta}^{l-1}$  is in the neighbourhood of a minimum. In practice therefore the following condition is used

$$|(\xi^2(\boldsymbol{\theta}^{l-1} + \boldsymbol{\theta}_{step}^l) - \xi^2(\boldsymbol{\theta}^{l-1}))| \leq q \cdot \Delta_\alpha. \quad (4.4)$$

To get a suitable  $\theta_{step,j}^l$  two methods have been implemented: Direct method and slope method.

The idea of the direct method is that a large step size is often favoured for efficacy. Therefore a max step size  $\eta_{max} > 0$  along with a minimal step size  $\eta_{min} > 0$  is set, then do the following at each iteration  $l$  of the PL calculation for parameter  $\theta_j$ :

1. Initialize by  $i = 0, \eta^i = \eta_{max}$
2. Set  $\theta_{step,j}^l = \eta^i$
3.
  - If: (4.4) is satisfied Or:  $\eta^i \leq \eta_{min}$   
 $\Rightarrow \theta_{step,j}^l = \eta^i$  and terminate.
  - Else: Decrease step size by  $f \cdot \eta^i, i \leftarrow i + 1$ .
4. Return to 2.

where  $f \in (0, 1)$ . The direct method ensures that a larger step size is favoured but the efficacy is dependent on the user-defined  $\eta_{max}$ . Choosing the max step size is a matter of guessing and a good value would be such that equation (4.4) is fulfilled immediately without having to decrease the step size, but since this is often not possible to know beforehand a sufficiently large value should be favoured to allow the algorithm to find a suitable step size. In this case,  $\eta_{max}$  is chosen as

$$\begin{aligned} \eta_{max}^l &= \log_{10}(\hat{\lambda}_j + l \cdot 0.2 \cdot \hat{\lambda}_j) - \log_{10}(\hat{\lambda}_j + (l-1) \cdot 0.2 \cdot \hat{\lambda}_j) \\ &= \log_{10}\left(\frac{1 + l \cdot 0.2}{1 + (l-1) \cdot 0.2}\right), \end{aligned} \quad (4.5)$$

with  $\lambda_j$  being the untransformed parameter value, such that  $\hat{\theta}_j = \log_{10}(\hat{\lambda}_j)$  and  $\theta_j^{l-1} = \log_{10}(\lambda_j^{l-1})$ . This way each iteration and parameter has a unique  $\eta_{max}$  and equation (4.5) represents a max step size of 20% of the untransformed estimate, if only max steps has been taken. This was seen to work well in all cases considered later.

The slope method is an improved version of the direct method which approximates the slope at the last point  $\boldsymbol{\theta}^{l-1}$  in the direction of  $\theta_j$  to calculate the step size  $\theta_{step,j}^l$ . Starting from  $\boldsymbol{\theta}^{l-1}$  do the following:

## 4. Implementation

---

1. Take a small step of size  $\eta_{min}$  in the direction of  $\theta_j$  so that  $\theta_{step}^{min} = [0, \dots, \eta_{min}, \dots, 0]$ .
2. Approximate the slope  $k$  by  $k \approx \frac{\xi^2(\theta^{l-1} + \theta_{step}^{min}) - \xi^2(\theta^{l-1})}{\eta_{min}}$ .
  - If  $k = 0$ : Set  $\eta_{new} = \eta_{max}$
  - Else if  $k > 0$ : Set  $\eta_{new} = \frac{q \cdot \Delta\alpha}{k}$
  - Else if  $k < 0$ : Set  $\eta_{new} = \frac{-q \cdot \Delta\alpha}{k}$
3. If  $\eta_{new} > \eta_{max}$ : Set  $\eta_{new} = \eta_{max}$ .
4. Set  $i = 0$  and  $\eta^i = \eta_{new}$ .
5. Set  $\theta_{step,j}^l = \eta^i$
6.
  - If: (4.4) is satisfied Or:  $\eta^i < \eta_{min}$   
 $\Rightarrow \theta_{step,j}^l = \eta^i$  and terminate.
  - Else: Decrease step size by  $f \cdot \eta^i$ ,  $i \leftarrow i + 1$ .
7. Return to 5.

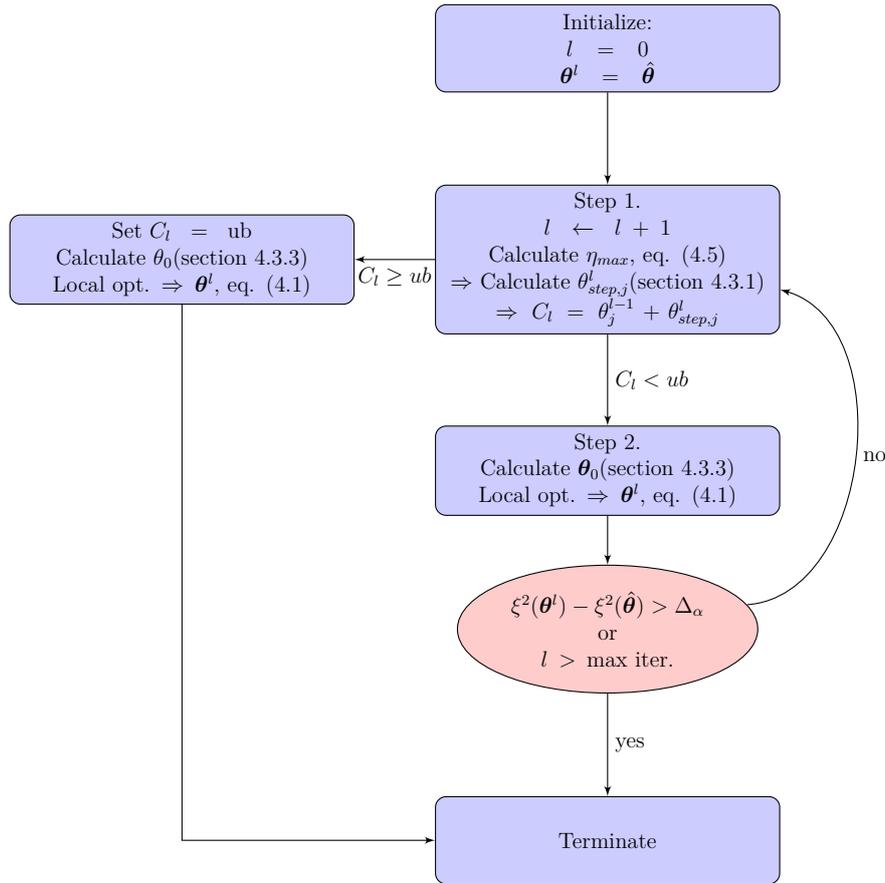
Compared to the direct method, the slope method calculates a reasonable guess at a step size that fulfills the condition in equation (4.4) instead of starting from  $\eta_{max}$  which significantly reduces the number of necessary iterations of decreasing the step size when  $\eta_{max}$  is too large to fulfill condition (4.4).

### 4.3.2 Constrained optimization

The local re-optimization problem at iteration  $l$  is formulated by equation (4.1) where  $C_l$  is calculated from the solution at the previous iteration  $\theta^{l-1}$  and the step size  $\theta_{step,j}^l$ . Since this problem needs to be solved multiple times for each parameter an efficient solver that can handle constrained optimization is needed. The implementation of profile-likelihood in this project allows the user to pass a general script for the constrained optimization.

### 4.3.3 Start point

The choice of starting point  $\theta_0$  for the local optimization can aid the optimizer regardless of what algorithm is used and two ways have been implemented. Let  $\theta^{l-1}$  be the solution to the local optimization problem at iteration  $l - 1$  of the PL calculation, then for the starting point at the next iteration  $l$  one can have



**Figure 4.1:** Flowchart describing the implementation for calculation of profile likelihood of parameter  $\theta_j$  in increasing direction. The calculation in decreasing direction is analogous.  $\hat{\theta}$  is the global estimate,  $l$  denotes the current iteration,  $\eta_{max}^l$  is the maximal step size at iteration  $l$  calculated by equation (4.5),  $\theta_{step,j}^l$  is calculated by either direct or slope method described in section 4.3.1,  $\theta_0$  is the start point calculated by either 0th order or 1st order proposal as described in section 4.3.3,  $\Delta_\alpha$  is the threshold given by the  $1 - \alpha$  quantile of  $\chi_{df}^2$  distribution and  $ub$  is the upper boundary for the parameter  $\theta_j$ .

1. 0th order proposal: the optimal point at the last iteration  $\theta_0 = \theta^{l-1}$ , or
2. 1st order proposal: the linear extrapolation  $\theta_0 = \theta^{l-1} + \frac{C_l - C_{l-1}}{C_{l-1} - C_{l-2}}(\theta^{l-1} - \theta^{l-2})$  based on the two previous optimal points.

The 1st order proposal is almost always computationally more efficient as it uses additional information reducing the risk of slow convergence of the optimization algorithm or a convergence towards a non-global minimum[12].

#### 4.3.4 Flowchart

Figure 4.1 gives an overview of the profile likelihood calculation.

### 4.3.5 Confidence intervals

The PL calculation for a parameter  $\theta_j$  gives the function  $\xi_{PL}^2(\theta_j)$  at values  $\theta_j = C_l$ ,  $l = 1, \dots, L$ . The likelihood-based confidence interval (CI) is then based on if the values of  $\xi_{PL}^2$  in increasing and decreasing direction exceeds the threshold  $\Delta_\alpha$  or not. If it exceeds the threshold then the point of passover represents the upper or lower bound for the CI. This point of passover is obtained by spline interpolation.

## 4.4 Test problem

To validate the performance of the program it is tested on three problems where analytical solutions of the identifiability is known. One of these is given here and the other two can be found in appendix A.

The model is a one-compartment pharmacokinetic model with first order absorption given by:

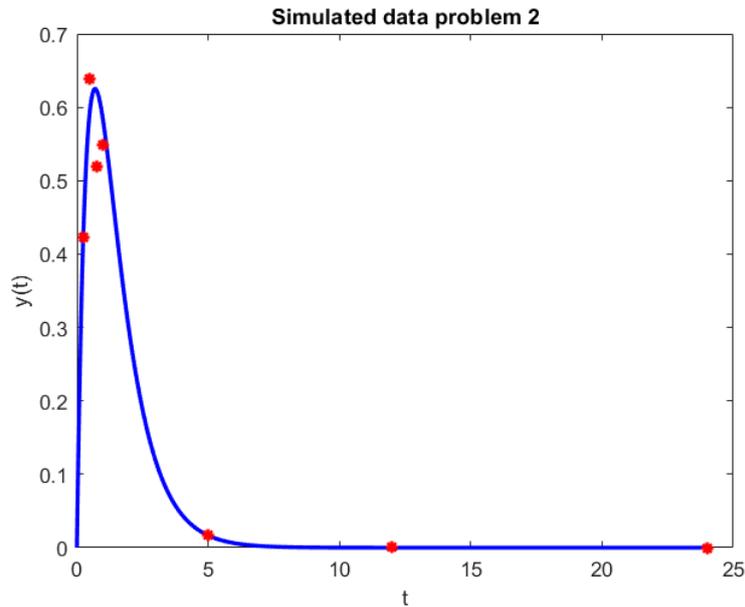
$$\begin{aligned} \frac{dx_1(t)}{dt} &= -k_a x_1(t), \quad x_1(0) = F \cdot dose \\ \frac{dx_2(t)}{dt} &= k_a x_1(t) - CL \frac{x_2(t)}{V}, \quad x_2(0) = 0 \\ y(t_i) &= \frac{x_2(t_i)}{V} \cdot e_i, \quad e_i \sim \log\mathcal{N}(0, \sigma_i). \end{aligned} \quad (4.6)$$

where  $x_1(t)$  is the amount of available drug,  $x_2(t)$  the amount of drug in the central compartment and  $y(t_i)$  is the measurement of the concentration in the central compartment at time  $t_i$ . The parameters are  $k_a, CL, V$  and  $F$ , where  $k_a$  is a rate constant,  $F$  the bioavailability,  $CL$  the clearance rate and  $V$  is volume. Here  $dose$  is assumed to be a known constant. Since  $F$  represents a fraction its domain is  $[0, 1]$ .  $k_a, CL$  and  $V$  are all assumed to be positive. In this case  $F, CL$  and  $V$  cannot be identified. This is since both  $CL$  and  $V$  are unknown and hence they cannot be separated from the ratio  $\frac{CL}{V}$ , and similarly for  $F$  and  $V$ . Meaning that the ratios  $\frac{CL}{V}$  and  $\frac{F}{V}$  can be identified, or equivalently  $\frac{CL}{F}$  and  $\frac{F}{V}$  can be identified, but not the parameters by themselves [16].

Taking measurements at time points as

$$\mathbf{t}_{obs} = [0.25, 0.5, 0.75, 1, 5, 12, 24] \quad (4.7)$$

in hours, and assuming log-normal error with  $\sigma_i = 0.2$ ,  $\forall i$ , the simulated data in figure 4.2 is obtained. The true parameter values were set as in Table 4.1. For estimation purposes  $\log_{10}$  parametrization is used for the problem as all parameters are positive, thus the only bound that needs to be considered is the upper boundary of  $F$ . *lsqnonlin* with *trust-region-reflective* is used for the optimization since it can handle bounded constraints. The parameter estimates are given in Table 4.1.



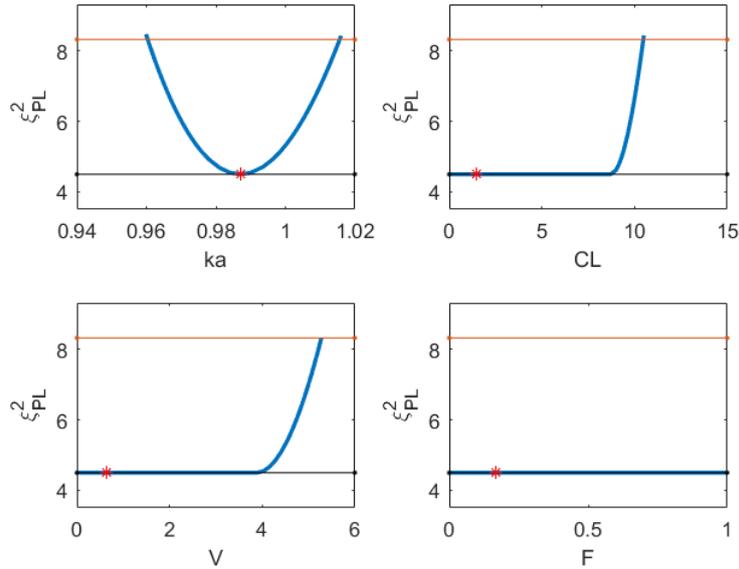
**Figure 4.2:** Simulated model trajectory given by  $\frac{x_2(t)}{V}$  in blue and observed data  $y(t_i)$  in red. The observable was assumed to be log-normal with standard error  $\sigma_i$  set to 0.2 for all measurements.

Parameter	True Value	Estimate	Confidence interval
$k_a$	1	0.99	[0.96, 1.02]
$CL$	4	1.44	(0, 10.47]
$V$	2	0.64	(0, 5.28]
$F$	0.5	0.17	(0, 1)

**Table 4.1:** True parameter values, global estimates and likelihood-based CIs for problem 2. The global estimation was performed with *lsqnonlin* and *trust-region-reflective*. *multistart* was used with 21 startpoints calculated using equation (3.1) with lower and upper bounds of the parameter ranges as:  $k_a$  : [0.0001, 10],  $CL$  : [0.1, 100],  $V$  : [0.1, 10],  $F$  : [0.01, 1]. The data was given as in figure 4.2. The CIs are calculated as the points of passover from the PLs in figure 4.3.

## 4. Implementation

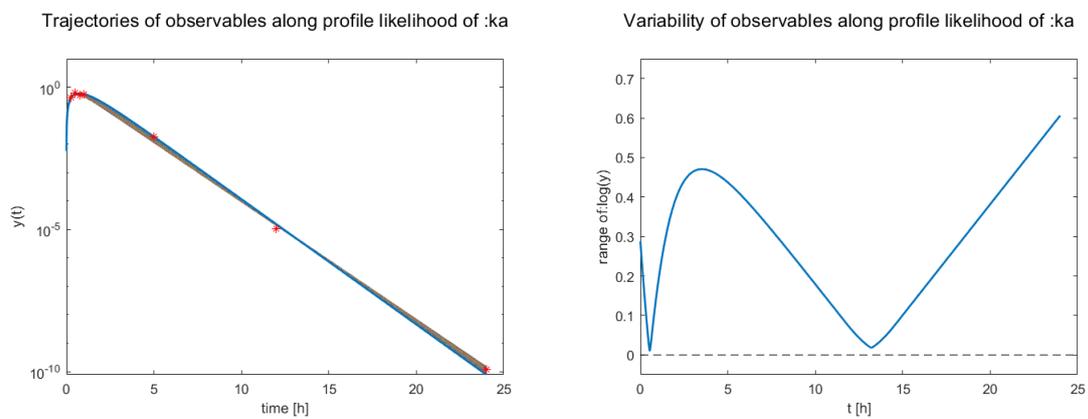
Calculating the profile likelihoods gives figure 4.3 and likelihood-based confidence intervals in Table 4.1. It is clear that  $F$  is structurally unidentifiable. For  $CL$  and  $V$  upper bounds are found but no lower bound greater than 0, implying practical unidentifiability by our definition. This is expected since the ratios  $\frac{F}{V} \equiv a$  and  $\frac{CL}{F} \equiv b$  should be identifiable, and since  $F \leq 1$  then both  $CL$  and  $V$  are bounded from above by  $CL \leq b$  and  $V \leq \frac{1}{a}$ . The parameter  $k_a$  is identifiable.



**Figure 4.3:** Profile likelihoods for test problem 2. The threshold is given by the red line. The data was assumed to have log-normal error distribution with std  $\sigma_i = 0.2$  for all measurements. The measurement times were taken as in equation (4.7). The PLs were calculated with the following settings: the quantile of  $\chi^2$ -distribution  $1 - \alpha$ : 0.95, max steps: 500, max step size: as in equation (4.5), min step size:  $10^{-6}$ , adaptive step size method: slope, q: 0.1, optimizer: *lsqnonlin* with *trust-region-reflective*, start point proposal: 1st order.

Plotting the trajectory of observable  $\frac{x_2(t)}{V}$  for parameter values along the profile likelihood of  $k_a$  it can be seen that the trajectories exhibits some variability, see left figure in 4.4. To assess this more clearly the range of log of the trajectories is calculated as in equation (3.12). Doing this gives the right figure in 4.4. This analysis shows that the uncertainty in  $k_a$  impacts the observables but the resulting variability at a certain time is in general on the same scale as the measurement error. Therefore additional measurements will not greatly improve the results for  $k_a$ , as the influence on the measurements from  $k_a$  will be roughly the same as the measurement error.

Similar analysis of the variability of trajectories can help in determining when additional measurements should be taken to resolve practical unidentifiabilities when these are present.



**Figure 4.4:** Upper: Trajectories of the observable along the profile likelihood of the identifiable parameter  $k_a$ , plotted on log-scale. The trajectories are given by  $y(t; \theta^l)$  in equation (4.6) with parameter values  $\theta^l$  given by the PL of  $k_a$ . The red stars are the observed measurements. Lower: Range of trajectories in the upper panel as given by equation (3.12).



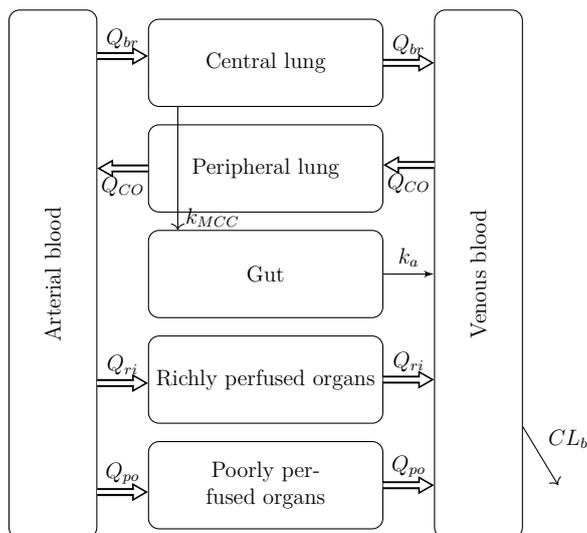
# 5

## Models

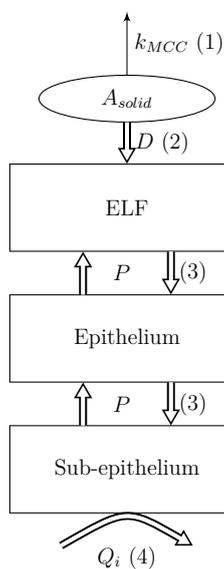
The models used in this project were developed by Teodor Erngren and Elin Boger at AstraZeneca (Early RIA DMPK) and are multi-compartment PBPK models. For small molecules, perfusion rate limited distribution was assumed in all tissues, meaning that the blood flow to the tissues is the limiting process for drug distribution [5]. This assumption is motivated by the fact that small molecules experience little to no resistance when diffusing into or out from tissues. In contrast, distribution of large molecules in tissues is assumed to be limited by endothelial transport. For this reason the model for large molecules also considers the interstitial and vascular spaces for all organs (including the lung).

### 5.1 Small molecules

Two models are considered for small molecules, one for neutral and one for basic compounds. The underlying structure of the models are given by the minimal PBPK model illustrated in figure 5.1. Each of the lung compartments (central and peripheral) are then decomposed into three sub-compartments: the epithelial lining fluid (ELF), the epithelium (ep) and the sub-epithelium (sub) as shown in figure 5.2. The difference between neutral and basic molecules is that basic molecules can be retained in lysosomes inside the lung and organs. Therefore lysosomic compartments are added in connection to the lung and organs for the basic case. In the following ODE-equations *tb* denotes the tracheobronchial region, i.e. central lung compartment and *al* the alveolar region, i.e. the peripheral lung compartment.



**Figure 5.1:** Minimal PBPK model for small molecules.  $Q_{br}$  is the bronchial blood flow,  $Q_{CO}$  the cardiac output,  $Q_{ri}$  and  $Q_{po}$  the blood flows to richly and poorly perfused organs respectively,  $k_a$  the oral absorption constant,  $k_{MCC}$  the mucociliary clearance rate constant and  $CL_b$  the blood clearance rate.



**Figure 5.2:** Schematic overview of the lung compartments in the neutral molecule model. The model accounts for transport by: 1) mucociliary clearance (MCC), 2) drug dissolution, 3) passive diffusion and 4) blood perfusion.  $Q_i$  is the blood flow, given by the bronchial blood flow  $Q_{br}$  or the cardiac output  $Q_{CO}$  in central and peripheral lung, respectively.  $A_{solid}$  is the solid amount of drug administered by DPI,  $k_{MCC}$  the MCC rate constant,  $D$  the diffusion coefficient and  $P$  the permeability.

### 5.1.1 Neutral molecules

Firstly, when drug is administered by inhalation of solid particles (dry powder inhalation (DPI)) the deposited particles are dissolved in the ELF. The change in drug-particle radius due to dissolution in region  $i$ , with  $i \in \{tb, al\}$ , is given by:

$$\begin{aligned}\frac{dr_i(t)}{dt} &= -\frac{D}{\rho r_i(t)}(C_s - C_{ELF,i}(t)f_{u,ELF}), \\ r(0) &= r_{i,0}\end{aligned}\quad (5.1)$$

where  $D$  is the diffusion coefficient,  $\rho$  the particle density,  $C_s$  the solubility,  $C_{ELF,i}$  the drug concentration in the ELF in region  $i$  and  $f_{u,ELF}$  the unbound fraction in the ELF. However, due to singularity issues when  $r_i \rightarrow 0$ , the following variable change

$$a_i(t) = r_i^2(t)$$

is introduced. The change in the squared radius in region  $i$  is then

$$\begin{aligned}\frac{da_i(t)}{dt} &= \frac{da_i(t)}{dr_i} \frac{dr_i(t)}{dt} = 2r_i(t) \frac{dr_i(t)}{dt} \\ &= -\frac{2D}{\rho}(C_s - C_{ELF,i}(t)f_{u,ELF}), \\ a_i(0) &= r_{i,0}^2.\end{aligned}\quad (5.2)$$

Now for the lung-compartments as depicted in figure 5.2, starting with the rate of change of concentration in ELF for region  $i$  ( $i \in tb, al$ ). This is given by

$$\begin{aligned}V_{ELF,i} \frac{dC_{ELF,i}(t)}{dt} &= N_i(t)4\pi D\sqrt{a_i(t)}(C_s - C_{ELF,i}(t)f_{u,ELF}) \\ &\quad - PA_i\left(C_{ELF,i}(t)f_{u,ELF} - \frac{C_{ep,i}(t)}{V_{u,lung}}\right), \\ C_{ELF,i}(0) &= df_i dose_{IT}\end{aligned}\quad (5.3)$$

where  $V_{ELF,i}$  is the volume of the ELF in region  $i$ ,  $C_{ep,i}$  the regional drug concentration in the epithelium,  $N_i$  the number of drug particles in region  $i$ ,  $V_{u,lung}$  the unbound lung volume of distribution and  $A_i$  the regional surface area. The initial condition here is given by the deposition fraction in region  $i$ ,  $df_i$  and the intra-tracheal dose  $dose_{IT}$ . Note that the first term only applies when dry powder inhalation is administrated since  $N_i(t) \equiv 0$  otherwise. The change of  $C_{ep,i}$  is then described by

$$\begin{aligned}V_{ep,i} \frac{dC_{ep,i}(t)}{dt} &= PA_i\left(C_{ELF,i}(t)f_{u,ELF} - \frac{C_{ep,i}(t)}{V_{u,lung}}\right) \\ &\quad - PA_i\left(\frac{C_{ep,i}(t)}{V_{u,lung}} - \frac{C_{sub,i}(t)}{V_{u,lung}}\right), \\ C_{ep,i}(0) &= 0\end{aligned}\quad (5.4)$$

where  $V_{ep,i}$  is the regional epithelial volume and  $C_{sub,i}$  is the regional sub-epithelial drug concentration which in turn is given by

$$\begin{aligned}
 V_{sub,i} \frac{dC_{sub,i}(t)}{dt} &= PA_i \left( \frac{C_{ep,i}(t)}{V_{u,lung}} - \frac{C_{sub,i}(t)}{V_{u,lung}} \right) \\
 &\quad + Q_i \left( C_{blood,i} - \frac{b_p C_{sub,i}(t)}{K_{p,lung}} \right), \\
 C_{sub,i}(0) &= 0
 \end{aligned} \tag{5.5}$$

where  $Q_i$  and  $C_{blood,i}$  are the regional blood flow and concentration, respectively,  $K_{p,lung}$  is the tissue to plasma partition coefficient for the lung and  $b_p$  is the blood to plasma ratio. Since the tracheobronchial region is perfused by the bronchial blood flow with arterial blood we have  $Q_i = Q_{br}$  and  $C_{blood,i} = C_{art}$  for  $i = tb$  while the alveolar region is perfused by the entire cardiac output and venous blood and thus  $Q_i = Q_{CO}$  and  $C_{blood,i} = C_{vein}$  for  $i = al$ .

With DPI administration, particles deposited in the tracheobronchial region ( $N_{tb}$ ) are transported up towards the trachea by mucociliary clearance (MCC). The MCC is described by the ODE

$$\begin{aligned}
 \frac{dN_{tb}(t)}{dt} &= -k_{MCC} N_{tb}(t), \\
 N_{tb}(0) &= N_{tb,0}
 \end{aligned} \tag{5.6}$$

where  $k_{MCC}$  is the MCC rate constant and  $N_{tb,0}$  the number of particles deposited in the tracheobronchial region at  $t = 0$ . In the alveolar region we have no MCC and therefore it is constant over time with  $N_{al}(t) = N_{al,0}$ .

The drug particles removed by MCC are then swallowed and ends up in the gut compartment (see figure 5.1), where the change in drug amount is given by

$$\begin{aligned}
 \frac{dA_{gut}(t)}{dt} &= k_{MCC} N_{tb}(t) m_p(t) - k_a A_{gut}(t) \\
 &= k_{MCC} N_{tb}(t) \frac{4\pi \sqrt{a_{tb}(t)}^3 \rho}{3} - k_a A_{gut}(t), \\
 A_{gut}(0) &= dose_{PO}
 \end{aligned} \tag{5.7}$$

where  $k_a$  is the oral absorption constant and  $dose_{PO}$  is the oral dose.

The remaining compartments in figure 5.1 are for the blood (arterial and venous) and organs excluding the lung which are divided into two groups, richly perfused e.g. liver and spleen, and poorly perfused e.g. skin and bone.

For arterial blood we have the following ODE

$$\begin{aligned}
 V_{art} \frac{dC_{art}(t)}{dt} &= -Q_{CO} \left( C_{art}(t) - \frac{b_p C_{sub,al}(t)}{K_{p,lung}} \right), \\
 C_{art}(0) &= 0
 \end{aligned} \tag{5.8}$$

and for venous blood we have

$$\begin{aligned}
V_{vein} \frac{dC_{vein}(t)}{dt} &= -Q_{ri} \left( C_{vein}(t) - \frac{b_p C_{ri}(t)}{K_{p,ri}} \right) - Q_{po} \left( C_{vein}(t) - \frac{b_p C_{po}(t)}{K_{p,po}} \right) \\
&\quad - Q_{br} \left( C_{vein}(t) - \frac{b_p C_{sub,tb}(t)}{K_{p,lung}} \right) - CL_b C_{vein}(t) + k_a F A_{gut}(t), \quad (5.9) \\
C_{vein}(0) &= \frac{dose_{IV}}{V_{vein}}
\end{aligned}$$

where  $C_{ri}$  and  $C_{po}$  are the concentrations in richly perfused and poorly perfused organs respectively,  $CL_b$  is the blood clearance rate,  $Q_{ri}$  and  $Q_{po}$  the blood flows to richly and poorly perfused organs,  $K_{p,j}$  the plasma to tissue partition coefficient of organ  $j$ ,  $F$  the oral bioavailability and  $dose_{IV}$  the intravenous dose. Lastly the rate of change in  $C_{ri}$  and  $C_{po}$  are given by

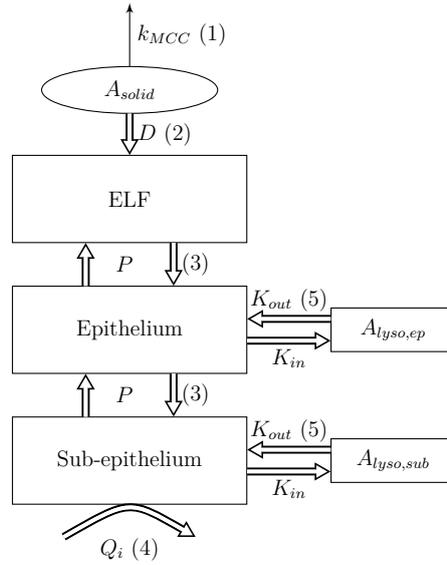
$$\begin{aligned}
V_{ri} \frac{dC_{ri}(t)}{dt} &= Q_{ri} \left( C_{art}(t) - \frac{b_p C_{ri}(t)}{K_{p,ri}} \right), \quad (5.10) \\
C_{ri}(0) &= 0
\end{aligned}$$

and

$$\begin{aligned}
V_{po} \frac{dC_{po}(t)}{dt} &= Q_{po} \left( C_{art}(t) - \frac{b_p C_{po}(t)}{K_{p,po}} \right), \quad (5.11) \\
C_{po}(0) &= 0.
\end{aligned}$$

### 5.1.2 Basic molecules

For basic molecules there are additional states representing amount of drug retained in lysosomes. In the PBPK model this corresponds to adding lysosomal compartments to the epithelium, subepithelium, richly perfused organs and poorly perfused organs with rate constants describing the transport into and out from these compartments, see figure 5.3 for a schematic representation of the compartmental structure in the lung for basic molecules.



**Figure 5.3:** Schematic overview of the lung compartments in the basic molecule model. The model accounts for transport by: 1) mucociliary clearance (MCC), 2) drug dissolution, 3) passive diffusion, 4) blood perfusion and 5) lysosomal trapping.  $Q_i$  is the blood flow, given by the bronchial blood flow  $Q_{br}$  and the cardiac output  $Q_{CO}$  in central and peripheral lung respectively.  $A_{solid}$  is the solid amount of drug administered by DPI,  $k_{MCC}$  the MCC rate constant,  $D$  the diffusion coefficient,  $P$  the permeability and  $K_{in}$  and  $K_{out}$  are rate constants for lysosomal trapping.

The rate of change for the amount of drug in lysosomal compartments in the lung are given by

$$\begin{aligned} \frac{dA_{lyso,i}(t)}{dt} &= K_{in}C_i(t)V_i - K_{out}A_{lyso,i}(t), \\ A_{lyso,i}(0) &= 0 \end{aligned} \quad (5.12)$$

with  $i \in \{eptb, epal, subtb, subal\}$  and where  $C_i$  is the concentration of drug in compartment  $i$ ,  $V_i$  the volume of compartment  $i$  and  $K_{in}$  and  $K_{out}$  are rate constants into and out from lysosomes. For amount of drug in lysosomal compartments in the organs the same equation is used but with different rate constants, i.e.

$$\begin{aligned} \frac{dA_{lyso,i}(t)}{dt} &= K_{in,org}C_i(t)V_i - K_{out,org}A_{lyso,i}(t), \\ A_{lyso,i}(0) &= 0 \end{aligned} \quad (5.13)$$

with  $i \in \{ri, po\}$ . However, as a simplification it can be assumed that the rate-constants in the organs are the same as in the lung, meaning  $K_{in,org} \equiv K_{in}$  and  $K_{out,org} \equiv K_{out}$ , reducing the number of parameters.

In order to accommodate for lysosomic trapping there is need to modify equations (5.4), (5.5), (5.10) and (5.11) such that:

$$\begin{aligned}
V_{ep,i} \frac{dC_{ep,i}(t)}{dt} &= PA_i(C_{ELF,i}(t)f_{u,ELF} - f_{u,shallow}C_{ep,i}(t)) \\
&\quad - PA_i f_{u,shallow}(C_{ep,i}(t) - C_{sub,i}(t)) \\
&\quad - K_{in}C_{ep,i}(t)V_{ep,i} + K_{out}A_{lyso,ep,i}(t), \\
C_{ep,i}(0) &= 0
\end{aligned} \tag{5.14}$$

$$\begin{aligned}
V_{sub,i} \frac{dC_{sub,i}(t)}{dt} &= PA_i f_{u,shallow}(C_{ep,i}(t) - C_{sub,i}(t)) \\
&\quad - K_{in}C_{sub,i}(t)V_{sub,i} + K_{out}A_{lyso,sub,i}(t) \\
&\quad + Q_i \left( C_{blood,i}(t) - \frac{b_p C_{sub,i}(t)}{K_{p,lung}} \right), \\
C_{sub,i}(0) &= 0
\end{aligned} \tag{5.15}$$

for  $i \in \{tb, al\}$  and where the parameter  $f_{u,shallow}$  is defined as:

$$f_{u,shallow} = \frac{1}{V_{u,lung}} \left( 1 + \frac{K_{in}}{K_{out}} \right). \tag{5.16}$$

and

$$\begin{aligned}
V_i \frac{dC_i(t)}{dt} &= Q_i \left( C_{art}(t) - \frac{b_p C_i(t)}{K_{p,sh,i}} \right) - K_{in,org}C_i(t)V_i + K_{out,org}A_{lyso,i}(t), \\
C_i(0) &= 0
\end{aligned} \tag{5.17}$$

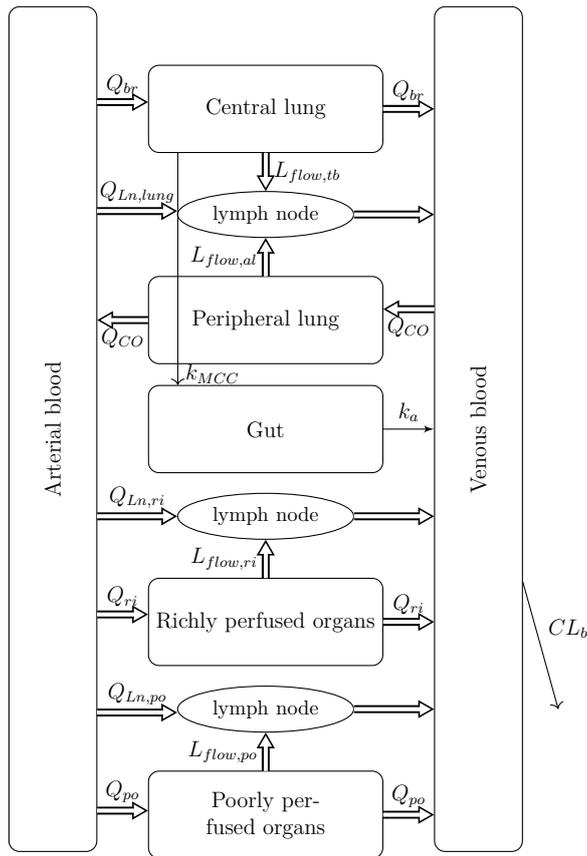
with  $i \in \{ri, po\}$  and where the shallow compartment  $K_{p,sh,i}$  is defined as

$$K_{p,sh,i} = \frac{K_{p,i}}{1 + \frac{K_{in,org}}{K_{out,org}}}. \tag{5.18}$$

The remaining equations are identical to the neutral molecule case.

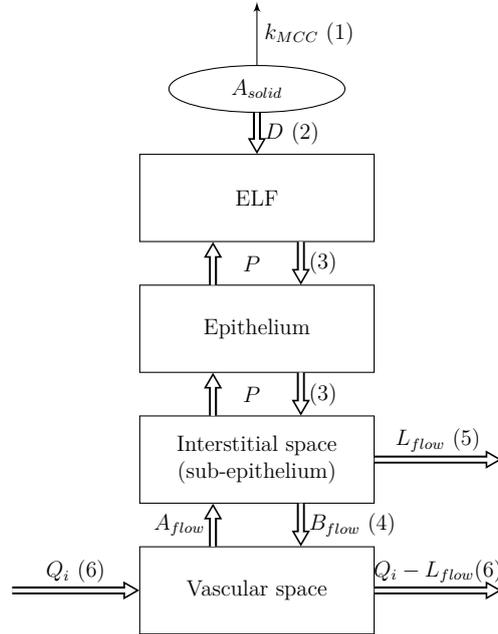
## 5.2 Large molecules

For large molecules the overall structure for the PBPK model is represented by figure 5.4, where the lung-compartments are decomposed into four separate sub-compartments representing the epithelial lining fluid, the epithelium, the sub-epithelium (or interstitial space) and the vascular space, see figure 5.5. Similarly, the other organs (i.e. richly perfused and poorly perfused) are also decomposed into an interstitial space and a vascular space, see figure 5.6 to accommodate for endothelial transport. The interstitial spaces are then connected to lymph nodes which are in turn reconnected to the rest of the system via the venous blood flow, see figure 5.4. The key difference between small and large molecules is that diffusion of small

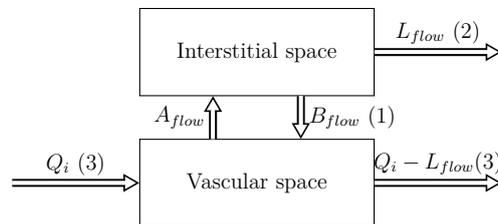


**Figure 5.4:** Minimal PBPK model for large molecules.  $Q_{br}$  is the bronchial blood flow,  $Q_{CO}$  the cardiac output,  $Q_{ri}$  and  $Q_{po}$  the blood flows to richly and poorly perfused organs respectively,  $Q_{Ln,i}$  with  $i \in \{lung, ri, po\}$  are the lymph blood flows,  $L_{flow,i}$  with  $i \in \{tb, al, ri, po\}$  the respective lymphatic drainage,  $k_a$  the oral absorption constant,  $k_{MCC}$  the mucociliary clearance rate and  $CL_b$  the blood clearance rate

molecules is assumed to be rapid compared to the various fluid flows. The distribution of small molecules in tissues (lung, richly and poorly perfused organs) are then governed by the unbound lung volume of distribution  $V_{u,lung}$  and  $K_p$  values. For large molecules on the other hand, the distribution is governed by endothelial transport which is modeled by reflection coefficients  $\sigma$ . However, from a modeling perspective the same equations describing transport between ELF, epithelium and sub-epithelium can be reused in the large molecule case only with  $V_{u,lung}$  as a dummy variable set to 1.



**Figure 5.5:** Schematic overview of the lung compartments in the large molecule model. The model accounts for transport by: 1) mucociliary clearance (MCC), 2) drug dissolution, 3) passive diffusion, 4) endothelial transport, 5) lymphatic drainage and 6) blood perfusion.  $Q_i$  is the blood flow, given by the bronchial blood flow  $Q_{br}$  and the cardiac output  $Q_{CO}$  in central and peripheral lung respectively.  $A_{solid}$  is the solid amount of drug administered by DPI,  $k_{MCC}$  the MCC rate constant,  $D$  the diffusion coefficient,  $P$  the permeability,  $A_{flow}$  and  $B_{flow}$  are the filtration and re-absorption flows respectively and  $L_{flow}$  is the lymphatic drainage.



**Figure 5.6:** Schematic overview of the structure of organs in the large molecule model. The model accounts for transport by: 1) endothelial transport, 2) lymphatic drainage and 3) blood perfusion.  $Q_i$  is the blood flow, given by  $Q_{ri}$  and  $Q_{po}$  in richly and poorly perfused organs respectively,  $A_{flow}$  and  $B_{flow}$  are the filtration and re-absorption flows respectively and  $L_{flow}$  is the lymphatic drainage.

The compartmental structure for solid particles ( $A_{solid}$ ), ELF and epithelium are the same as for neutral molecules (cmp figures 5.5 and 5.2) and thus the ODEs that describes the rate of change of particle radius, ELF concentration and epithelium concentration given by equations (5.2) - (5.4) are reused but with  $V_{u,lung}$  as a dummy variable set to 1. In the sub-epithelium for large molecules the change in drug concentration is given by:

$$\begin{aligned}
 V_{sub,i} \frac{dC_{sub,i}(t)}{dt} &= ((1 - \sigma_{v,i}) \frac{C_{vasc,i}(t)}{b_p} A_{flow,i}) - ((1 - \sigma_{v,i}) C_{sub,i}(t) B_{flow,i}) \\
 &\quad - ((1 - \sigma_l) C_{sub,i}(t) L_{flow,i}) + PA_i \left( \frac{C_{ep,i}(t)}{V_{u,lung}} - \frac{C_{sub,i}(t)}{V_{u,lung}} \right), \quad (5.19) \\
 C_{sub,i}(0) &= 0, \quad i \in \{tb, al\}
 \end{aligned}$$

where  $\sigma_{v,i}$  is the vascular reflection coefficient in region  $i$ ,  $A_{flow,i}$  the filtration rate in region  $i$ ,  $B_{flow,i}$  the re-absorption rate in region  $i$  and  $L_{flow,i}$  the lymphatic drainage in region  $i$ .

The drug concentration in vascular space of region  $i$ , with  $i \in \{tb, al\}$ , is given by

$$\begin{aligned}
 V_{vasc,tb} \frac{dC_{vasc,tb}(t)}{dt} &= Q_{br} C_{art} - (Q_{br} - L_{flow,tb}) C_{vasc,tb}(t) \\
 &\quad - ((1 - \sigma_{v,tb}) \frac{C_{vasc,tb}(t)}{b_p} A_{flow,tb}) + ((1 - \sigma_{v,tb}) C_{sub,tb}(t) B_{flow,tb}), \\
 C_{vasc,tb}(0) &= 0
 \end{aligned} \quad (5.20)$$

for the central lung and

$$\begin{aligned}
 V_{vasc,al} \frac{dC_{vasc,al}(t)}{dt} &= Q_{CO} C_{vein} - (Q_{CO} - L_{flow,al}) C_{vasc,al}(t) \\
 &\quad - ((1 - \sigma_{v,al}) \frac{C_{vasc,al}(t)}{b_p} A_{flow,al}) + ((1 - \sigma_{v,al}) C_{sub,al}(t) B_{flow,al}), \\
 C_{vasc,al}(0) &= 0
 \end{aligned} \quad (5.21)$$

for the peripheral lung.

As before, MCC applies for particles deposited in central lung and the corresponding ODEs for  $N_{tb}$  and the gut remains unchanged (equations (5.6) and (5.7)). For the blood compartments the change in drug concentration is now given by

$$\begin{aligned}
 V_{art} \frac{dC_{art}(t)}{dt} &= (Q_{CO} - L_{flow,al}) C_{vasc,al}(t) - Q_{CO} C_{art}(t), \\
 C_{art}(0) &= 0
 \end{aligned} \quad (5.22)$$

and

$$\begin{aligned}
V_{vein} \frac{dC_{vein}(t)}{dt} &= (Q_{po} - L_{po})C_{vasc,po}(t) + (Q_{ri} - L_{ri})C_{vasc,ri}(t) \\
&+ (Q_{br} - L_{br})C_{vasc,br}(t) + (Q_{Ln,po} - L_{Ln,po})C_{Ln,vasc,po}(t) \\
&+ (Q_{Ln,ri} - L_{Ln,ri})C_{Ln,vasc,ri}(t) + (Q_{Ln,lung} - L_{Ln,lung})C_{Ln,vasc,lung}(t) \\
&+ (L_{po} + L_{Ln,po})C_{Ln,is,po}(t) + (L_{ri} + L_{Ln,ri})C_{Ln,is,ri}(t) \\
&+ (L_{br} + L_{al} + L_{Ln,lung})C_{Ln,is,lung}(t) - (Q_{po} + Q_{ri} + Q_{br})C_{vein}(t) \\
&- (Q_{Ln,po} + Q_{Ln,ri} + Q_{Ln,lung})C_{vein}(t) \\
&- CL_b C_{vein}(t) + k_a F A_{gut}, \\
C_{vein}(0) &= \frac{dose_{IV}}{V_{vein}}.
\end{aligned} \tag{5.23}$$

The organs are modeled by two compartments, the interstitial and the vascular spaces as in figure 5.6, where the rate of change in drug concentration are given by

$$\begin{aligned}
V_{is,i} \frac{dC_{is,i}(t)}{dt} &= ((1 - \sigma_{v,org}) \frac{C_{vasc,i}(t)}{b_p} A_{flow,i}) \\
&- ((1 - \sigma_{v,org}) C_{is,i}(t) B_{flow,i}) \\
&- ((1 - \sigma_l) C_{is,i}(t) L_{flow,i}), \\
C_{is,i}(0) &= 0,
\end{aligned} \tag{5.24}$$

and

$$\begin{aligned}
V_{vasc,i} \frac{dC_{vasc,i}(t)}{dt} &= Q_i C_{art}(t) \\
&- (Q_i - L_{flow,i}) C_{vasc,i}(t) \\
&+ ((1 - \sigma_{v,org}) C_{is,i}(t) B_{flow,i}) \\
&- ((1 - \sigma_{v,org}) \frac{C_{vasc,i}(t)}{b_p} A_{flow,i}), \\
C_{vasc,i}(0) &= 0
\end{aligned} \tag{5.25}$$

with  $i \in \{ri, po\}$ .

In the large molecule model, lymph nodes are connected to all organs (including the lung), see figure 5.4. The lymph nodes are decomposed into an interstitial lymph space and a vascular lymph space, similar to the organs as in figure 5.6 but with an additional influx to the interstitial space given by the respective lymphatic drainages  $L_{flow,i}$ . The ODEs describing the rate of change in drug concentration in these compartments are given by the following equations, starting with the interstitial lymph space in connection to the richly and poorly perfused organs we have

$$\begin{aligned}
V_{Ln,is,i} \frac{dC_{Ln,is,i}(t)}{dt} &= ((1 - \sigma_l) C_{is,i}(t) L_{flow,i}) + ((1 - \sigma_{v,Ln}) \frac{C_{Ln,vasc,i}(t)}{b_p} A_{flow,Ln,i}) \\
&- ((1 - \sigma_{v,Ln}) C_{Ln,is,i}(t) B_{flow,Ln,i}) - (C_{Ln,is,i}(t) (L_{flow,i} + L_{flow,Ln,i})), \\
C_{Ln,is,i}(0) &= 0,
\end{aligned} \tag{5.26}$$

with  $i \in \{ri, po\}$  and for the interstitial lymph space in connection to the lung we have

$$\begin{aligned}
 V_{Ln, is, lung} \frac{dC_{Ln, is, lung}(t)}{dt} &= ((1 - \sigma_l)C_{sub, tb}(t)L_{flow, tb}) + ((1 - \sigma_l)C_{sub, al}(t)L_{flow, al}) \\
 &\quad + ((1 - \sigma_{v, Ln}) \frac{C_{Ln, vasc, lung}(t)}{b_p} A_{flow, Ln, lung}) \\
 &\quad - ((1 - \sigma_{v, Ln})C_{Ln, is, lung}(t)B_{flow, Ln, lung}) \\
 &\quad - (L_{flow, tb} + L_{flow, al} + L_{flow, Ln, lung})C_{Ln, is, lung}(t), \\
 C_{Ln, is, lung}(0) &= 0.
 \end{aligned} \tag{5.27}$$

Lastly, the change in drug concentration in all vascular lymph spaces is given by:

$$\begin{aligned}
 V_{Ln, vasc, i} \frac{dC_{Ln, vasc, i}(t)}{dt} &= Q_{Ln, i}C_{art}(t) - ((1 - \sigma_{v, Ln}) \frac{C_{Ln, vasc, i}(t)}{b_p} A_{flow, i}) \\
 &\quad + ((1 - \sigma_{v, Ln})C_{Ln, is, i}(t)B_{flow, i}) - (Q_{Ln, i} - L_{flow, Ln, i})C_{Ln, vasc, i}(t), \\
 C_{Ln, vasc, i}(0) &= 0,
 \end{aligned} \tag{5.28}$$

with  $i \in \{ri, po, lung\}$ .

### 5.2.1 Reflection coefficients

The reflection coefficients represent restriction coefficients for endothelial transport dependent on the size of the compound and take on values between 0 and 1. A larger reflection coefficient corresponds to higher restriction of transport into tissues and vice versa. The model accounts for three different modeling cases for the vascular reflection coefficients  $\sigma_{v, i}$ .

1. The most complex case when we have unique reflection coefficients in the peripheral lung, central lung and organs. Meaning that the restriction for endothelial transport are different in all three compartments.
2. A simplified case, assuming reflection coefficients in central lung and organs are the same, i.e.  $\sigma_{v, tb} \equiv \sigma_{v, org}$ .
3. The most simple case when all reflection coefficients in lung and organs are the same, i.e.  $\sigma_{v, tb} \equiv \sigma_{v, al} \equiv \sigma_{v, org}$ .

In this thesis only case 1 and 2 are considered.

### 5.3 Dosing

The models support four possible administration routes for the drug which are incorporated in the initial conditions of the ODE-systems.

1. Intravenous administration (IV):  $C_{vein}(0) = \frac{dose_{IV}}{V_{vein}}$
2. Oral administration (PO):  $A_{gut}(0) = dose_{PO}$
3. Dry powder inhalation (DPI):  $N_i(0) = N_{i,0}$  where  $N_{i,0} = \frac{df_i dose_{DPI}}{\rho \frac{4\pi \sqrt{a_{i,0}}^3}{3}}$ ,  
for  $i \in \{tb, al\}$
4. Intratracheal instillation (IT):  $C_{ELF,i}(0) = df_i dose_{IT}$ , for  $i \in \{tb, al\}$

Oral administration is not considered in this thesis.

### 5.4 Experimental measurements

The equations in sections 5.1.1, 5.1.2 and 5.2 describe the unobserved states  $\mathbf{x}(t; \boldsymbol{\theta})$  for neutral, basic and large molecules respectively. For all three models the observables  $g_k(\mathbf{x}(t; \boldsymbol{\lambda}))$  are given by measuring the plasma concentration

$$C_{plasma}(t) = \frac{C_{vein}(t)}{b_p} \quad (5.29)$$

and the total lung concentration which is given by the sum of amount of drug in all lung compartments + the amount of solid drug particles. Since the models have different compartmental structure in the lung, three different expressions for the total lung concentration are obtained depending on the model.

- For neutral molecules:

$$C_{lung}(t) = \frac{1}{V_{lung}} \sum_{i \in \{tb, al\}} \left( V_{ELF,i} C_{ELF,i}(t) + V_{ep,i} C_{ep,i}(t) + V_{sub,i} C_{sub,i}(t) + N_i(t) \rho \frac{4\pi a_i^{3/2}}{3} \right) \quad (5.30)$$

- For basic molecules lysosomic compartments are also present in the lung and thus:

$$C_{lung}(t) = \frac{1}{V_{lung}} \sum_{i \in \{tb, al\}} \left( V_{ELF,i} C_{ELF,i}(t) + V_{ep,i} C_{ep,i}(t) + A_{lyso,ep,i}(t) + V_{sub,i} C_{sub,i}(t) + A_{lyso,sub,i}(t) + N_i(t) \rho \frac{4\pi a_i^{3/2}}{3} \right) \quad (5.31)$$

- For large molecules there is also a vascular space compartment and therefore:

$$C_{lung}(t) = \frac{1}{V_{lung}} \sum_{i \in \{tb, al\}} \left( V_{ELF,i} C_{ELF,i}(t) + V_{ep,i} C_{ep,i}(t) + V_{sub,i} C_{sub,i}(t) + V_{vasc,i} C_{vasc,i}(t) + N_i(t) \rho \frac{4\pi a_i^{3/2}}{3} \right) \quad (5.32)$$

The simulated measurements of these quantities are then given by applying log-normal errors, as described in section 2.1.

In this thesis, although only simulated data is used, these measurements represents in vivo experiments performed on rats or mice. To obtain reasonable simulated data it is important to know how actual experimental data would be taken and therefore a short motivation on how the measurement times for simulated data were decided is given here.

The procedure for measuring the lung concentration is terminal, meaning that each lung measurement correspond to one animal. The plasma concentration need not be terminal and several measurements of plasma for each animal can be obtained, but no more than 10 measurements per animal is reasonable. In a standard pre-clinical set up, around 7-9 measurements each of lung and plasma are measured. This would correspond to 21-27 animals if three individual rats/mice are measured at each measurement time to account for inter-individual/study occasion variability. The lung and plasma concentration in this case would be measured at the same time points for practical reasons.

In this thesis, a set up of 8 simulated measurements each of plasma and lung at the same time points is used as a starting point. During the identifiability analysis, additional measurements need to be considered for unidentifiable parameters, and in these cases plasma concentration will also be measured at each new measurement time for lung concentration. If only additional measurements of plasma are needed but the current lung measurements are sufficient, we will consider only adding plasma measurements as each new lung measurement would imply the need for a new rat/mice.

#### 5.4.1 Additional observables

In this thesis  $C_{plasma}$  and  $C_{lung}$  are always considered as observables, as these are common measurements in inhalation studies. However, two additional options for measurements will be considered to potentially improve identifiability in estimation studies. The first is the average concentration in ELF given by

$$C_{ELF,avg}(t) = \sum_{i \in \{tb, al\}} \frac{V_{ELF,i} C_{ELF,i}(t)}{V_{ELF,tb} + V_{ELF,al}}. \quad (5.33)$$

The second is the total concentration of drug in some richly perfused organ, e.g. spleen, which will only be considered for large molecules. It is given by

$$C_{ri,tot}(t) = \frac{V_{vasc,ri}C_{vasc,ri} + V_{is,ri}C_{is,ri}}{V_{ri}} \quad (5.34)$$

where  $V_{ri}$  is the total volume of the organ.

As with  $C_{plasma}$  and  $C_{lung}$ , simulated data for these additional observables are obtained by applying log-normal errors.

## 5.5 Parameters

The parameters that are considered for identifiability analysis are difficult, or not possible to measure accurately by experimental methods, or rely on assumptions that certain empirical models are true. For the three different models, also three different sets of parameters are of interest. In the neutral molecule model the following parameters are considered for estimation:

- Diffusion coefficient  $D$  - Can be calculated from Stokes-Einstein equation [13]. Clearly, this relies on the assumption that this equation is true. Furthermore, there is some uncertainty around the viscosity of the ELF, which is an input parameter to the Stokes-Einstein equation. It would therefore be interesting to assess if it is feasible to estimate  $D$  from experimental data.
- Solubility  $C_s$  - Development of experimental methodologies of solubility in biorelevant mediums is ongoing research but there is currently lack of consensus of which dissolution media to use for characterizing the solubility of compounds in ELF [6]. Estimating this parameter is thus of interest until better experimental setups are developed.
- Unbound lung volume of distribution  $V_{u,lung}$  - In order to reduce the number of experiments required prior to starting PBPK modelling, it would be interesting to assess if this parameter can be estimated directly from in vivo data.
- Permeability  $P$  - This parameter represents the in vivo pulmonary permeability which cannot be determined experimentally [6].
- Fraction unbound in ELF  $f_{u,ELF}$  - Currently not possible to measure but experimental methods are emerging.
- Deposition fraction  $df_{tb}$  - Obtained from deposition modeling. Nevertheless interesting to see if feasible to estimate.

Note that only the tracheobronchular (i.e. central lung) deposition fraction,  $df_{tb}$ , is considered as a parameter since  $df_{al} = 1 - df_{tb}$ . One could equivalently have set the

alveolar deposition fraction  $df_{al}$  as the parameter of interest.

For the basic molecule case, rate constants for the lysosomes  $K_{in}$ ,  $K_{out}$ ,  $K_{in,org}$  and  $K_{out,org}$  are also of interest together with all parameters above. There are currently no methods for determining rate constants experimentally.

For the large molecule  $V_{u,lung}$  is a dummy variable set to 1 as previously mentioned, so this parameter is not of interest. Instead all vascular reflection coefficients  $\sigma_{v,i}$  except for in the lymph nodes  $\sigma_{v,Ln}$  are considered together with the remaining parameters in the neutral case.

All parameters of interest are defined as to be positive, moreover the parameters  $f_{u,ELF}$ ,  $df_{tb}$ ,  $\sigma_{v,org}$ ,  $\sigma_{v,al}$  and  $\sigma_{v,tb}$  are restricted to be less than 1. All other physical parameters in the models are treated as constants, obtained by independent experimental measurements or from literature. For a quick summary of all parameters considered for estimation in each model and their domain see Table 5.1.

Parameter	Neutral	Base	Large	Domain
$D$	Yes	Yes	Yes	$(0, \infty)$
$C_s$	Yes	Yes	Yes	$(0, \infty)$
$V_{u,lung}$	Yes	Yes	No	$(0, \infty)$
$P$	Yes	Yes	Yes	$(0, \infty)$
$f_{u,ELF}$	Yes	Yes	Yes	$(0, 1]$
$df_{tb}$	Yes	Yes	Yes	$(0, 1]$
$K_{in}$	-	Yes	-	$(0, \infty)$
$K_{out}$	-	Yes	-	$(0, \infty)$
$K_{in,org}$	-	Yes*	-	$(0, \infty)$
$K_{out,org}$	-	Yes*	-	$(0, \infty)$
$\sigma_{v,org}$	-	-	Yes	$(0, 1]$
$\sigma_{v,al}$	-	-	Yes	$(0, 1]$
$\sigma_{v,tb}$	-	-	Yes*	$(0, 1]$

**Table 5.1:** Table for Parameters of interest for identifiability analysis for each model and their domain. A star (\*) indicates parameters that can be eliminated by assumption as described in section 5.2.1 and 5.1.2.

In the basic molecule model there is a possible simplification by setting rate constants in lung as same as in the organs, and similarly for large molecules, reflection coefficients in the central lung can be assumed to be identical to in the organs. These simplifications will be considered as separate cases during the identifiability analysis.

### 5.5.1 Parameter ranges

To be able to calculate start points for the global estimation procedure, defined in section 3.1 and equation (3.1), the parameter ranges defined by lower limits  $\mathbf{l}$  and

upper limits  $\mathbf{u}$  needs to be chosen. The parameter ranges that were used are given in Table 5.2.

parameter [unit]	$l$ - lower range limit	$u$ - upper range limit
$D$ [ $\frac{dm^2}{h}$ ]	$10^{-6}$	$10^{-2}$
$C_s$ [ $nM$ ]	$10^{-1}$	$10^5$
$P$ [ $\frac{dm}{h}$ ]	$\begin{cases} 10^{-3} \text{ small molecules} \\ 10^{-6} \text{ large molecules} \end{cases}$	$\begin{cases} 10^2 \text{ small molecules} \\ 1 \text{ large molecules} \end{cases}$
$V_{u,lung}$ [ $\frac{mL}{g \text{ lung tissue}}$ ]	1	5000
$f_{u,ELF}$ [-]	0.1	1
$df_{tb}$ [-]	1/3	1
$K_{in}$ & $K_{in,org}$ , [ $\frac{1}{h}$ ]	$10^{-4}$	2
$K_{out}$ & $K_{out,org}$ , [ $\frac{1}{h}$ ]	$10^{-3}$	10
$\sigma_{v,org}$ & $\sigma_{v,al}$ & $\sigma_{v,tb}$ [-]	0.2	1

**Table 5.2:** Parameter ranges used for calculating start points in the global optimization problem.

### 5.5.2 Priors

To aid the global optimizer uniform priors can be set on some or all parameters. In our case priors on the parameters  $D$  and  $C_s$  are set when DPI administration is used. This is mainly to avoid the optimizer to explore the parameter space for very large values of  $D$  and  $C_s$  as the gradient of equations (5.3) and (5.2) gets very large which makes integration problematic. Without priors it was seen that the global optimization often diverged to unreasonably large values. Since reasonable parameter values are known from the ranges defined in Table 5.2, this information will also be used to formulate the priors. A wider range of values for the priors are set than what is defined in Table 5.2 since enforcing highly informative priors defeats the purpose of estimation. With  $\log_{10}$  parametrization the uniform priors are chosen as

$$\log_{10} D \sim U(-10, 1) \quad (5.35)$$

$$\log_{10} C_s \sim U(-6, 6). \quad (5.36)$$

The addition of priors means that the objective function must be altered. In the case of a uniform prior  $U(a, b)$ , the density is  $\frac{1}{b-a}$  and the term  $-2 \log(\frac{1}{b-a})$  is added. Outside the support for the uniform densities the objective function should be infinite, meaning that equation (2.14) needs to be modified as

$$\xi^2(\boldsymbol{\theta}) \equiv \sum_{k=1}^K \sum_{i=1}^{N_k} \left( \frac{\log(y_k(t_i)) - \log(\mathbf{g}_k(\mathbf{x}(t_i; \boldsymbol{\theta})))}{\sigma_{ki}} \right)^2 - 2 \left( \log \left( \frac{1}{11} \cdot I_{(-10,1)}(D) \right) + \log \left( \frac{1}{12} \cdot I_{(-6,6)}(C_s) \right) \right). \quad (5.37)$$

whenever DPI administration is considered and  $D$  and  $C_s$  are estimated. Here  $I(a, b)$  are indicator functions defined as

$$I_{(a,b)}(x) = \begin{cases} 1, & \text{if } x \in (a, b) \\ 0, & \text{otherwise.} \end{cases} \quad (5.38)$$

Since many optimizers does not allow for infinte values, the implementation of equation 5.37 is done such that when the optimizer steps outside the support of any of the priors the  $\xi^2$  value is set to something large, e.g.  $10^{100}$ .

# 6

## Results

As a starting point for the identifiability analysis, simulated data with measurements taken at time points

$$\mathbf{t}^{obs} = [2/60, 5/60, 15/60, 30/60, 1, 3, 6, 24] \quad (6.1)$$

given in hours for both observables  $C_{plasma}(t)$  and  $C_{lung}(t)$  will be used. The error is assumed to be log-normal and first the standard deviations are set to a low value of  $\sigma_{ki} = 0.01$ ,  $\forall k, i$ , to see which parameters are well-determined in an ideal case when very precise measurements can be taken.

The measurement error is then increased to more reasonable values to obtain data that reflects the current practices of in-vivo measurements. These values are obtained from a data set of experimental measurements of  $C_{plasma}$  and  $C_{lung}$  for IT and IV administration in rats [9], see appendix B for details. The reasonable values for the standard deviations obtained from this are

$$\begin{aligned} \sigma_{C_{plasma}}^{IT} &= 0.3159, \text{ for plasma measurements and IT administration} \\ \sigma_{C_{lung}}^{IT} &= 0.2056, \text{ for lung measurements and IT administration} \\ \sigma_{C_{plasma}}^{IV} &= 0.3752, \text{ for plasma measurements and IV administration} \\ \sigma_{C_{lung}}^{IV} &= 0.2609, \text{ for lung measurements and IV administration} \end{aligned} \quad (6.2)$$

and in each of the four cases the standard deviations are assumed to be the same at all times. For DPI administration no data was available so it is assumed to be the same as for IT administration. In the cases when additional observables are considered the corresponding standard deviations are set to 0.3.

As mentioned in chapter 5, three of the parameters are only incorporated in the models for certain administration routes, namely  $D, C_s$  and  $df_{tb}$ . The diffusion coefficient  $D$  and solubility  $C_s$  are a part of the model structure only if there is dissolution of solid particles in the lung, meaning they are only incorporated for DPI-administration. Similarly  $df_{tb}$  is only incorporated in the cases when the drug is administered directly into the lung, i.e. for IT or DPI-administration.

The administration routes that will be addressed here are mainly IT and IV. Moreover, since the different administration routes represent different inputs to the models it is often the case that data from both administration routes is needed or significantly improves the results as will be shown below. Thus the combination of IT

and IV data is also treated. Lastly, since the dissolution process is not incorporated unless DPI administration is used combined IT, IV and DPI data is studied as well.

Unless otherwise noted, the PL calculation is performed with the settings in Table 6.1. The max step size is calculated from equation (4.5).

$\alpha$	0.05
max steps	500
min step size	$10^{-6}$
adaptive step size method	slope
start point proposal	1st order
optimizer	<i>lsqnonlin</i> with <i>trust-region-reflective</i>

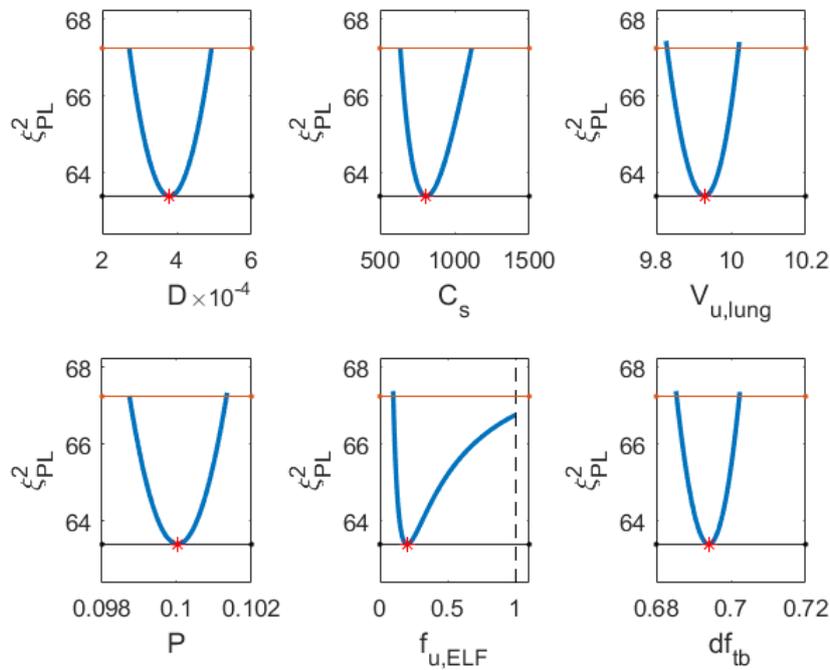
**Table 6.1:** Standard settings for PL algorithm.

The results of profile likelihood calculations and global optimizations for the parameters are presented on the untransformed, i.e. linear, scale for the sake of interpretation. It is therefore important to note that since  $\log_{10}$ -parametrization is used, any unidentifiability in the negative direction (meaning a confidence interval as  $(-\infty, c_i]$ ) will correspond to a parameter allowed infinitesimally close to 0 (meaning a confidence interval as  $(0, 10^{c_i}]$ ).

## 6.1 Neutral molecules model

First the ideal case with precise measurement. A thorough analysis for this case is given in appendix C and a summary of the most important results are given here. First and foremost, there is large uncertainty in the estimate of  $f_{u,ELF}$ , even with very high precision on measurements ( $\sigma_{k,i} = 0.01, \forall k, i$ ). Moreover, only IT data seems to be sufficient to obtain well-determined estimates of all parameters except  $D$  and  $C_s$  if precise measurements are possible. Lastly, all parameters are identifiable with combined IT, IV and DPI data, see Figure 6.1 and corresponding Table 6.2.

Now the log-normal error is increased to more reasonable values by setting the standard deviations as in equation (6.2). First considering only IT or IV administration the results vary with the error realization. For data with only IT-administration, the parameters  $P$  and  $df_{tb}$  remains identifiable but  $V_{u,lung}$  is sometimes missing a lower bound, meaning practical unidentifiability, see for example Figure C.6 in appendix C. For some error realizations  $V_{u,lung}$  is still identifiable but it can either have a narrow CI as depicted in Figure C.7 or exhibit large uncertainty with the lower bound for the CI much farther away from the true value, often with a local minimum as depicted in Figure C.8.



**Figure 6.1:** Profile likelihoods of parameter estimates in the neutral molecule model calculated with IT, IV and DPI data and low measurement error. The data was given by IT, IV and DPI administration separately where  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  were measured with  $t_i$  as in equation (6.1). The log-normal standard deviations were set to 0.01 at all measurements. All parameters are identifiable.

Low measurement error - IT/IV/DPI data			
Parameter	True value	Estimate	CI
$D$	0.000305	0.000381	[0.000273, 0.000493]
$C_s$	1000	811	[636, 1114]
$V_{u,lung}$	10	9.93	[9.83, 10.02]
$P$	0.1	0.100	[0.0987, 0.101]
$f_{u,ELF}$	1	0.200	[0.0987, 1]
$df_{tb}$	0.7	0.694	[0.685, 0.702]

**Table 6.2:** Parameter estimates and confidence intervals (CIs) for the neutral molecule model with IT, IV and DPI data and low measurement error ( $\sigma_{ki} = 0.01$ ). The CIs are calculated from the profile likelihoods in figure 6.1. All parameters are identifiable but the CIs for  $D$ ,  $C_s$  and  $f_{u,ELF}$  are considerably wider than the rest.

The results for  $f_{u,ELF}$  are also inconsistent and can be grouped into two cases:

1. The PL of  $f_{u,ELF}$  can have a minimum close to 0, and a flattening of the profile likelihood is observed for increasing values, see figure C.8. Technically, this case is identifiable since there is a minimum and an upper bound of 1 can always be imposed since the parameter represents a fraction.
2.  $f_{u,ELF}$  can be practically unidentifiable with no lower bound found, see figures C.6 and C.7. This can also be a consequence of the PL calculation stopping prematurely.

In both cases, calculation of CIs for the estimate of  $f_{u,ELF}$  leads to the conclusion that almost all values in its domain  $(0, 1]$  are feasible. Since the parameter had large CIs even for very precise measurements these results implies that the current setup is not suitable for estimating  $f_{u,ELF}$ .

With only IV-administration all three potentially identifiable parameters  $P, V_{u,lung}$  and  $f_{u,ELF}$  are practically unidentifiable in some runs and identifiable in other runs, see as examples figures C.9 and C.10.

The inconsistency in the results for only IT or IV administration means that this data is insufficient and therefore combined IT and IV data is considered next. To see if this gives consistent results, PL calculation is run for 1000 different error realizations with data taken as in equation (6.1) and log-normal standard deviations given by equation (6.2). The results are summarized in Table 6.3. The calculation of PLs and CIs for  $D$  and  $C_s$  was disregarded as these are not possible to estimate unless DPI administration is used. As stated above, the number of identifiable cases of  $f_{u,ELF}$  might be higher in reality as PL calculation is often stopped prematurely. However, as previously mentioned, the current setup is not suitable to obtain reliable estimates of  $f_{u,ELF}$  since even in identifiable cases the CI is  $[\delta, 1]$  with  $\delta$  close to 0 (e.g. 0.005). Moreover, some failures in the PL calculations are usually unavoidable since the individual runs are performed using the same settings. Often if the constrained optimization fails one can adjust the step sizes and the start point proposal or use another optimizer to resolve this issue. However, in this case the number of times the optimization routine returned any error or converged to a point  $\xi^2(\boldsymbol{\theta}^l)$  significantly larger than the previous  $\xi^2(\boldsymbol{\theta}^{l-1})$  are considered failures and are retracted from the results.

The results in Table 6.3 shows that  $V_{u,lung}$  is identifiable in 72% of error realizations. The lower bound  $c^-$  of the CI  $[c^-, c^+]$  for  $V_{u,lung}$  when it is identifiable can either be close to the true value of 10 or far from it, as described above. Out of all identifiable cases,  $c^-$  was less than 7 in 92% of cases and less than 2 in 26% of cases. In contrast  $c^+$  was greater than 13 in only 7% of cases. Hence, even when  $V_{u,lung}$  is identifiable a close lower bound is often not found.  $P$  and  $df_{tb}$  are identifiable in all individual runs, but it is unclear if this is still true with only IT data.

Summary over 1000 error realizations - IT/IV data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	964	72	6
$P$	975	100	6
$f_{u,ELF}$	947	33(*)	3
$df_{tb}$	990	100	4

**Table 6.3:** The PL calculations was run with settings as in Table 6.1. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) for both observables ( $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$ ) on both administration arms. Total number of data points in each run was thus  $n = 32$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.

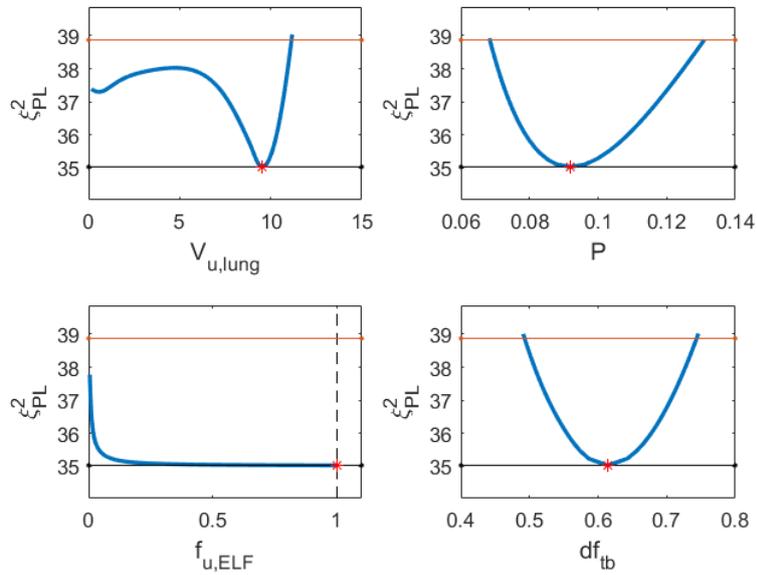
Since the analysis with low measurement error showed that IT administration alone was sufficient to obtain well-determined estimates it is of interest to see if only IT data is sufficient with a reasonable measurement error. To make the comparison fair the number of measurements are doubled, yielding a data set of only IT measurements of the same size as the IT and IV data. The measurement times were chosen as

$$\mathbf{t}^{obs} = [2/60, 5/60, 10/60, 15/60, 20/60, 25/60, 30/60, 35/60, 40/60, 45/60, 1, 2, 3, 4, 6, 24]. \quad (6.3)$$

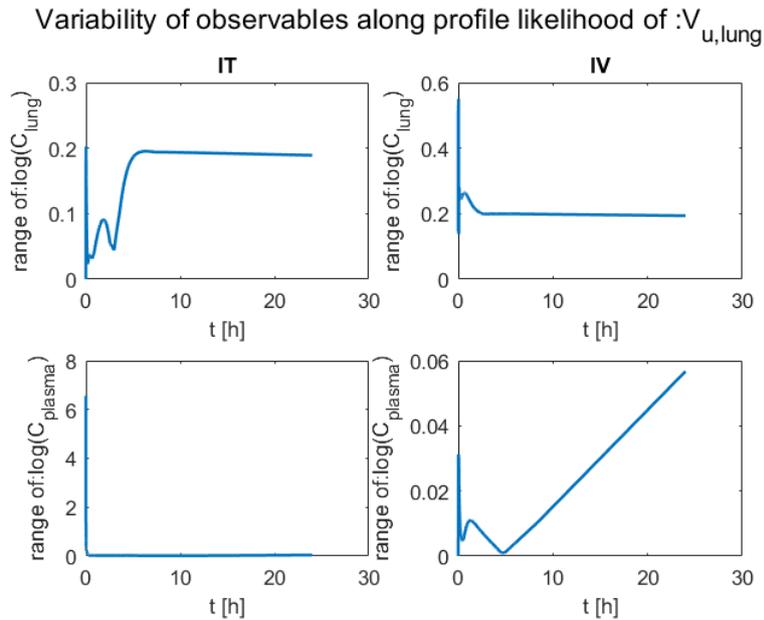
Running the profile likelihood calculation for 100 random error realizations with the same settings as before shows that  $V_{u,lung}$  is identifiable in only 21% of error realizations, see Table C.4 in appendix C.  $P$  and  $df_{tb}$  were identifiable in all individual runs. The combined IT and IV data is thus the better alternative.

Taking a case with IT and IV data for which  $V_{u,lung}$  is practically unidentifiable, see figure 6.2, and calculating the variability of trajectories along the PL of  $V_{u,lung}$  by applying equation (3.12) gives Figure 6.3. This shows that the uncertainty in the parameter greatly impacts the observable  $C_{plasma}(t)$  of the IT arm in the initial phase. The variability is highest at  $t = 0$  and rapidly decreases with time. Hence, initial measurements as soon as possible after dosing could improve our results. The same is true for  $f_{u,ELF}$ , see Figure C.11. Three additional measurements of  $C_{plasma}$  on the IT arm at 1, 2 and 3 seconds after dosing gives identifiability of  $V_{u,lung}$  and  $f_{u,ELF}$  in 100% of error realizations, see Table C.5. Moreover, the CIs for both  $V_{u,lung}$  and  $f_{u,ELF}$  are considerably smaller than in previous cases, see Figure C.12 for the PLs of an individual run.

Taking measurements in the first seconds after dosing is however unfeasible in practice, but the variability of trajectories indicate that measurements at any other time will not lead to consistent improvement since the variability of trajectories is equal



**Figure 6.2:** A case of practical unidentifiability of  $V_{u,lung}$  with IT and IV data for neutral molecules. Data was generated with standard deviations set as in equation (6.2) and measurement times as in equation (6.1).



**Figure 6.3:** Range of log of the observable trajectories calculated as in equation (3.12) with  $\theta^l$  along the PL of  $V_{u,lung}$  in figure 6.2.

to or lower than the measurement error at these times. Indeed, running the same analysis for 1000 error realizations of IT and IV data with double the amount of measurements, taken as in equation (6.3) for both arms, leads to no improvement, see Table C.6.

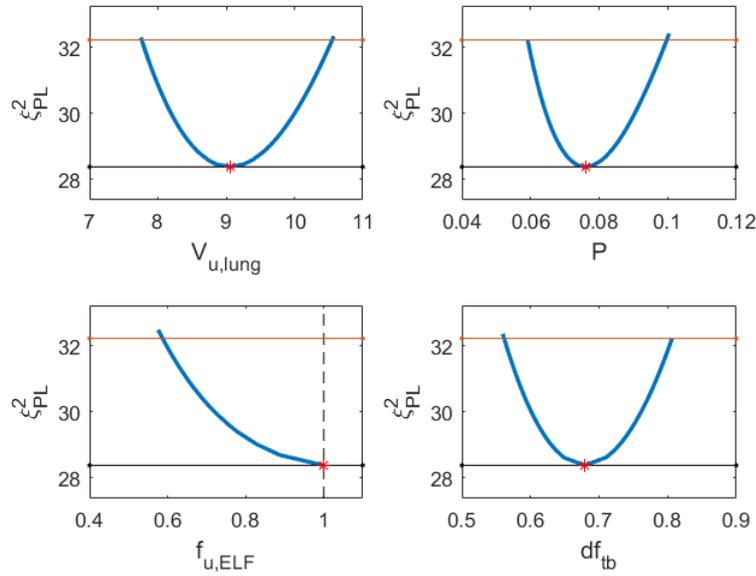
If measuring the average ELF concentration given by equation (5.33) is possible then this can pose an alternative to early measurements. With only one additional measurement of ELF on the IT arm at the final time point  $t = 24$  well-determined estimates of  $V_{u,lung}$  and  $f_{u,ELF}$  are obtained in 100% of error realizations, see Table 6.4. The CIs are again smaller than compared to a case with no measurements of ELF-concentration, see Figure 6.4.

Summary over 100 error realizations - IT/IV with additional measurement of ELF			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	100	100	7
$P$	100	100	5
$f_{u,ELF}$	98	100	3
$df_{tb}$	100	100	4

**Table 6.4:** The PL calculations was run with settings as in Table 6.1. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) for both observables ( $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$ ) on the IT and IV arms. One measurement of  $C_{ELF,avg}(t)$  as in equation (5.33) at  $t = 24$ h on the IT arm was added. The log-normal standard deviation for the ELF-measurement was set to 0.3. Total number of data points in each run was thus  $n = 33$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.

Lastly, DPI data together with IT and IV data gives information about the dissolution process. The parameters  $D$  and  $C_s$  were identifiable but with considerably larger CIs when using low measurement error (Table 6.2). Now with reasonable measurement error they are practically unidentifiable as seen in Figure 6.5. For 100 error realizations of this case, the parameters  $D$  and  $C_s$  were only identifiable in 4 cases and these were all false positives, see Table C.7.

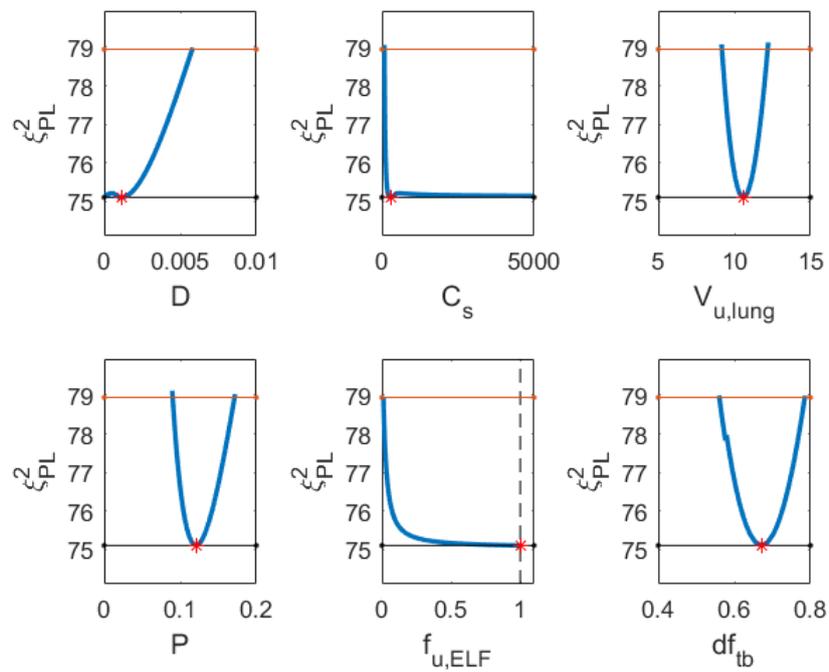
However, if Stokes-Einstein is assumed to be valid then  $D$  can be fixated [13][2]. If this assumption is made then  $C_s$  is well-determined, see Figure 6.6 for the PLs calculated with IT, IV and DPI data and where  $D$  assumed to be known. Rerunning the analysis for 100 error realizations with  $D$  fixated results in Table 6.5.  $C_s$  is identifiable in all runs. Hence, if Stokes-Einstein is assumed to be valid and  $D$  is calculated from this equation rather than by estimation, then  $C_s$  can be estimated.



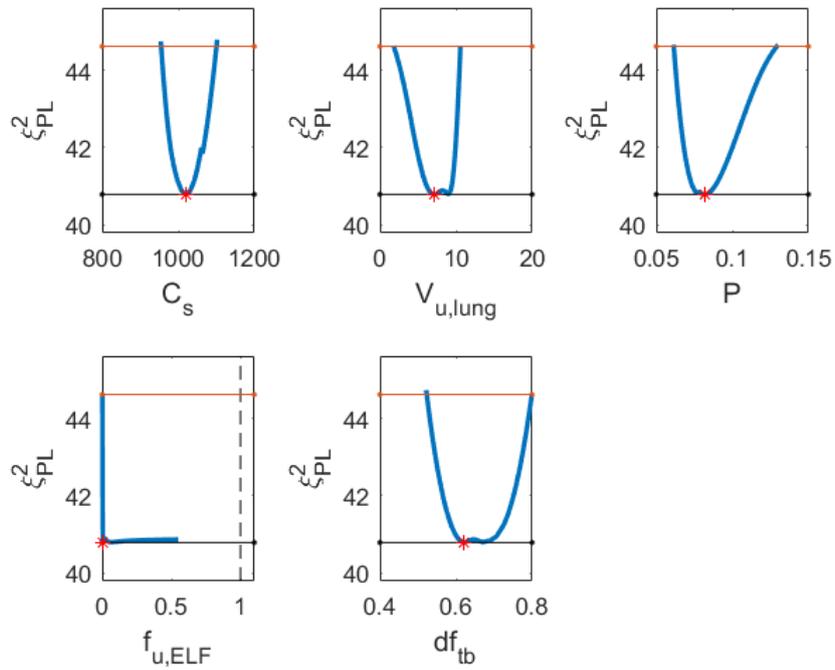
**Figure 6.4:** Profile likelihoods of parameter estimates in the neutral molecule model calculated with IT and IV data. The data was given by IT and IV administration separately where  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  were measured with  $t_i$  as in equation (6.1). A single measurement of  $C_{ELF,avg}(t)$  at the final time point  $t = 24\text{h}$  was added on the IT arm. The log-normal standard deviation was set as in equation (6.2) for measurements of  $C_{plasma}(t)$  and  $C_{lung}(t)$ . The standard deviation for the ELF-measurement was set to 0.3.

Summary over 100 error realizations - IT/IV/DPI data, $D$ -fixed			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$C_s$	100	100	4
$V_{u,lung}$	91	82	3
$P$	99	100	3
$f_{u,ELF}$	95	32(*)	2
$df_{tb}$	100	100	2

**Table 6.5:** The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2). The standard deviations for DPI data was set to the same values as IT. The measurement times was set as in equation (6.1) for both observables on all administration arms. The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.



**Figure 6.5:** Profile likelihoods of parameter estimates in the neutral molecule model calculated with IT, IV and DPI data. The data was given by IT, IV and DPI administration separately where  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  were measured with  $t_i$  as in equation (6.1). The log-normal measurement error was as in equation (6.2) at all measurement times for the IT and IV data. The standard deviations for DPI data was set to the same values as IT.



**Figure 6.6:** Profile likelihoods of parameter estimates in the neutral molecule model calculated with IT, IV and DPI data. The data was given by IT, IV and DPI administration separately where  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  were measured with  $t_i$  as in equation (6.1). The log-normal measurement error was as in equation (6.2) at all measurement times. The standard deviations for DPI data was set to the same values as IT.  $D$  was assumed to be known and calculated from Stokes-Einsteins equation.

## 6.2 Basic molecules model

For the basic molecule model the rate constants for lysosomal trapping in organs (richly and poorly) and lung (peripheral and central) can either be unique or assumed to be identical. This leads to two separate cases where the simplified case with rate constants assumed to be identical is considered first. Since the model structure for basic molecules and neutral molecules are very similar, the results are expected to be similar as well.

### 6.2.1 Simplified case

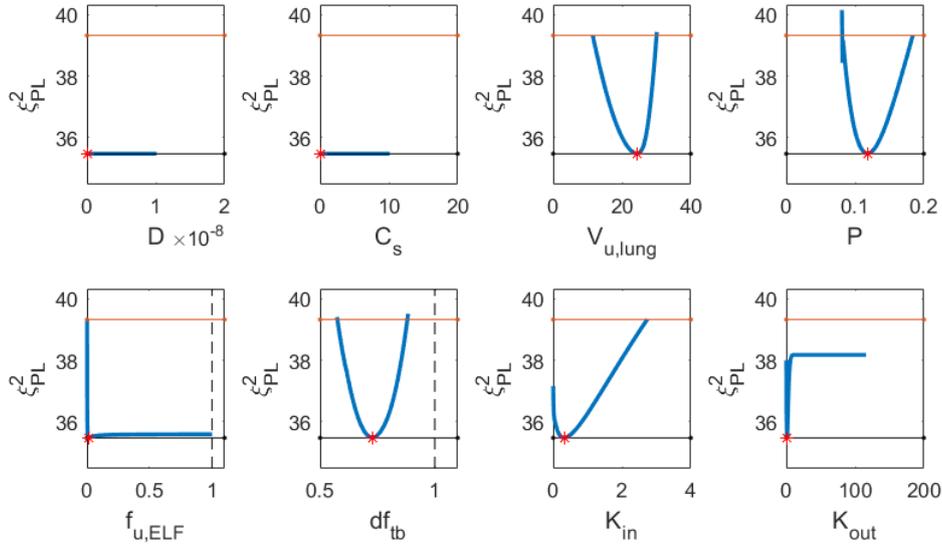
Once again a thorough analysis for data taken with precise measurements is given in appendix D. The main results are the same as for neutral molecules, only with the addition of rate constants that are also identifiable with low measurement error.

With measurement error set to reasonable values given by equation (6.2), the cases of only IT or IV administration leads to inconsistent results just as in the neutral molecule case and hence combined IT and IV data is considered immediately. The results of 1000 error realizations show that  $P$  and  $df_{tb}$  are identifiable in all runs and  $V_{u,lung}$  is identifiable in the majority of runs, see Table 6.6. As discussed in the previous section, the number of identifiable cases of  $f_{u,ELF}$  is higher in reality because of premature termination of the PL calculations, but there is high uncertainty in the estimate of  $f_{u,ELF}$  even in identifiable cases. The rate constants  $K_{in}$  and  $K_{out}$  are only identifiable in 21% and 31% of cases respectively and the number of false positives are much higher than expected. Since the rate constants were identifiable with low measurement error, this indicates that uncertainty in the rate constants impact the observables considerably less than the reasonable measurement errors.

Summary over 1000 error realizations - IT/IV data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	926	80	8
$P$	971	100	9
$f_{u,ELF}$	957	24(*)	6
$df_{tb}$	959	100	5
$K_{in}$	873	21	17
$K_{out}$	881	31	10

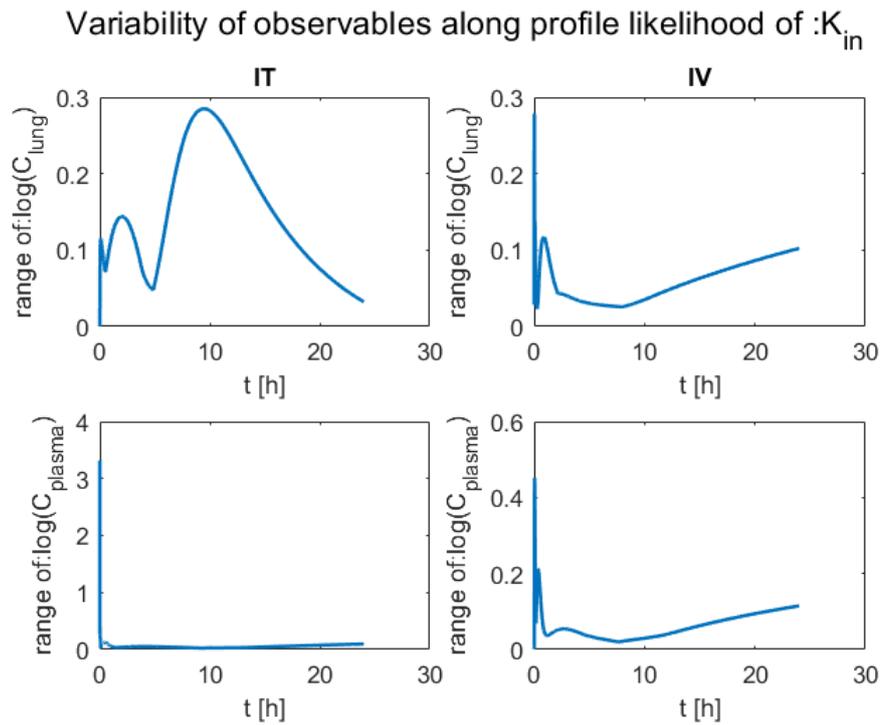
**Table 6.6:** 1000 error realizations for the simplified case of the basic molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) for both observables. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.

Taking one case of practical unidentifiability of the rate constants, see Figure 6.7, and calculating the range of trajectories as in equation (3.12) gives Figures 6.8-6.9. Excluding the initial times, the plots show that variability is highest at 8 – 16 h of the IT arm and at the final time point for the IV arm, but the range of trajectories is on the same scale as the measurement error. Nonetheless, taking additional measurements at  $t = 8, 10, 12, 14, 16$ h on the IT arm and  $t = 18, 20, 22$ h on the IV arm leads to some improvement for the rate constants, see Table 6.7. The number of false positives are still higher than expected but lower than before and the rate constants are identifiable roughly twice as often.

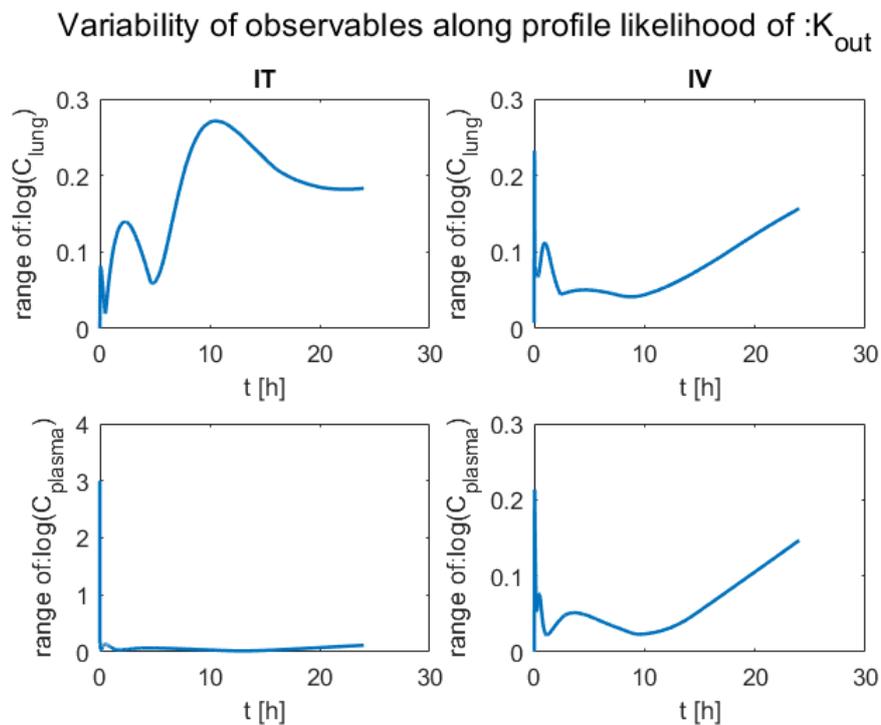


**Figure 6.7:** Profile likelihoods calculated for the basic molecule model with IT and IV data and reasonable measurement error. The log-normal standard deviations were set as in (6.2) for all measurements times. Data was taken at time points given by equation (6.1).

With additional DPI data the results for  $D$  and  $C_s$  are the same as for the neutral molecule case, i.e. practical unidentifiability of  $D$  and  $C_s$ , see Figure D.10 and identifiability of  $C_s$  is obtained with  $D$  fixed. The rate constants are still not well-determined.



**Figure 6.8:** Range of logarithm of the observables calculated over  $\theta^l$  at each simulated time point, see equation (3.12).  $\theta^l$  given by the PL calculation of parameter  $K_{in}$  in figure 6.7.



**Figure 6.9:** Range of logarithm of the observables calculated over  $\theta^l$  at each simulated time point, see equation (3.12).  $\theta^l$  given by the PL calculation of parameter  $K_{out}$  in figure 6.7.

Summary over 1000 error realizations - IT/IV data with additional measurements			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	921	94	7
$P$	978	100	7
$f_{u,ELF}$	991	22(*)	3
$df_{tb}$	915	100	6
$K_{in}$	928	59	10
$K_{out}$	961	67	7

**Table 6.7:** Settings for PL calculation as in equation (6.1). The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) + additional measurements on the IT arm at  $t = 8, 10, 12, 14, 16$ h of both observables and additional measurements on the IV arm at  $t = 18, 20, 22$ h of both observables. The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.

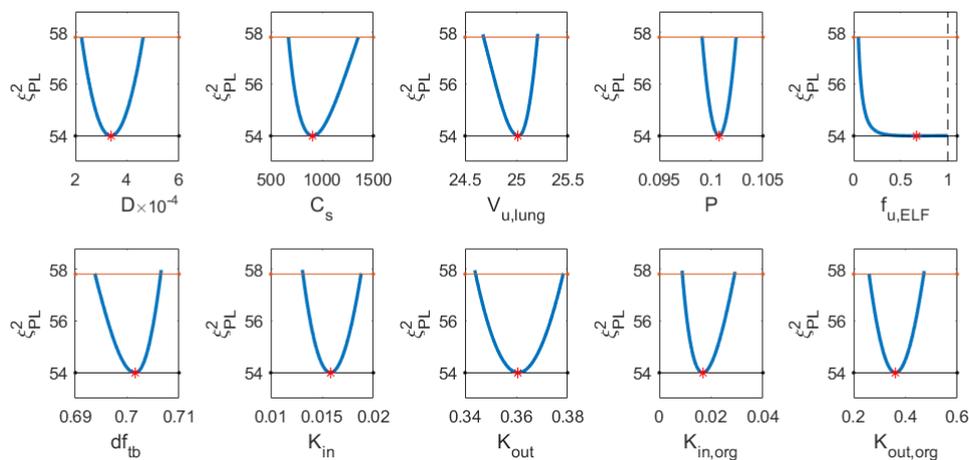
### 6.2.2 Unique case

Here the rate constants in organs and lung are assumed to be different. For only IT or IV administration cases there are problems for estimating  $K_{in,org}$  and  $K_{out,org}$  due to noise in the data, even with very precise measurements. For example see the case in Table 6.8 where the third best optimum is closest to the true values. It seems therefore that the rate constants in organs have very little impact on the observables since measurement error as low as  $\sigma = 0.01$  influences the estimation. However, combined IT and IV data seems to be sufficient to find well-determined estimates as shown in figure D.11 and with additional DPI data all parameters are identifiable see figure 6.10.

Since the simplified case of the basic molecule model showed that reliable results for the rate constants were not obtainable and since estimating  $K_{in,org}$  and  $K_{out,org}$  in the unique case of the basic molecule model is problematic even for precise measurements it is not expected that any of the rate constants should be consistently identifiable with reasonable measurement error. Indeed, the results of running the PL calculations over 1000 error realizations of IT and IV data is given in Table 6.9. The data is not sufficient to estimate any of the rate constants. Moreover, failure in the optimization routine is frequent which also suggest the lack of informative data. To resolve these issues a more in-depth analysis of the simplified case should be performed, since that case suffered from similar problems. When satisfactory results are obtained from the simplified case one can also see if the same is true in this more complex case with unique rate constants. However, this will not be considered here. Some suggestions on the other hand can be given for starting this task. For example, additional measurements at  $t = 8, 10, 12, 14, 16$ h on the IT arm and  $t = 18, 20, 22$ h on the IV arm were seen to improve the number of identifiable

Obj fcn value / Parameter	True value	Optimum 1	Optimum 2	Optimum 3
$\xi^2$	-	23.60	24.21	28.95
$V_{u,lung}$	25	24.75	22.79	24.97
$P$	0.1	0.0987	0.0901	0.0999
$f_{u,ELF}$	1	1	0.0335	1
$df_{tb}$	0.7	0.706	0.701	0.705
$K_{in}$	0.015	0.0146	0.0148	0.0156
$K_{out}$	0.36	0.346	0.325	0.366
$K_{in,org}$	0.02	0.00578	0.00633	0.0278
$K_{out,org}$	0.4	0.0854	0.0774	0.456

**Table 6.8:** Basic molecule model with unique rate constants in lung and organs, IT administration with  $\sigma_{ki} = 0.01 \forall k, i$ . The 3 best optimas found (for the same error realization) during global optimization with corresponding parameter values. Note that the third best optimum have  $K_{in,org}$  and  $K_{out,org}$  values closest to the true values. For the better two optimas the rate constants in organs are overfitted to noise in the data. The parameters  $D$  and  $C_s$  are disregarded as they are not incorporated in the model for IT administration.



**Figure 6.10:** Profile likelihoods calculated for the basic molecule model with IT, IV and DPI data and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).

cases for the rate constants in simplified case. It would therefore be of interest to see if more tightly sampled data could improve those numbers even more. Another approach is to measure  $C_{ELF,avg}(t)$  or  $C_{ri}(t)$ .

Summary over 1000 runs - IT/IV data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	739	91	10
$P$	948	100	11
$f_{u,ELF}$	913	18(*)	7
$df_{tb}$	954	100	6
$K_{in}$	764	9	16
$K_{out}$	771	29	12
$K_{in,org}$	833	12	6
$K_{out,org}$	874	17	10

**Table 6.9:** Settings for PL calculation as in equation (6.1). The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) for both observables on both administration arms. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.

### 6.3 Large molecules model

With large molecules transport into and out from tissues as well as clearance rate is slower than for small molecules. The time frame when a considerable amount of drug is still available in the system is therefore longer. For small molecules the lung and plasma concentrations dropped close to zero within 24 hours (see for example Figures C.1, C.3 and C.5), this is not the case for large molecules in general. Since the pharmacokinetics of large molecules operate on a larger time scale it is generally easier to capture the different phases of the system dynamics than for small molecules. For example, in the small molecule case the initial phase of the system when drug enters the lung/blood stream cannot be measured as this happens in a matter of seconds and such early measurements are not possible. Nonetheless, it was seen that early measurements could provide the necessary information to obtain well-determined estimates for neutral molecules. For large molecules the analogous initial phase can take several hours and so this can be easily measured compared to small molecules. However, this also means that the late phases of the system dynamics, which can be in the range of hundreds or even thousands of hours, is not captured with the measurements given by equation (6.1). Adding a measurement at a late time point is clearly possible and easy with simulated data but in practice taking measurements several hundreds of hours after dosing is not common. Therefore one measurement at  $t = 72\text{h}$  is added to the times in (6.1)

to have a starting point for identifiability analysis within a reasonable time frame. Thus the measurement times for large molecules are

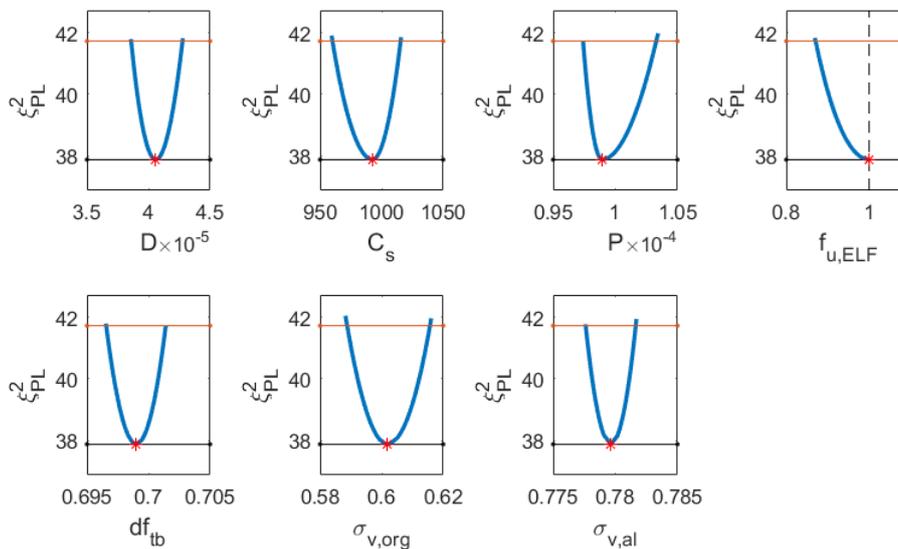
$$\mathbf{t}^{obs} = [2/60, 5/60, 15/60, 30/60, 1, 3, 6, 24, 72] \quad (6.4)$$

in hours.

### 6.3.1 Simplified case: $\sigma_{v,tb} \equiv \sigma_{v,org}$

In the simplified case of the large molecule model, the vascular reflection coefficients in central lung ( $\sigma_{v,tb}$ ) and organs ( $\sigma_{v,org}$ ) are assumed to be identical. The vascular reflection coefficient for both these regions is denoted by  $\sigma_{v,org}$ .

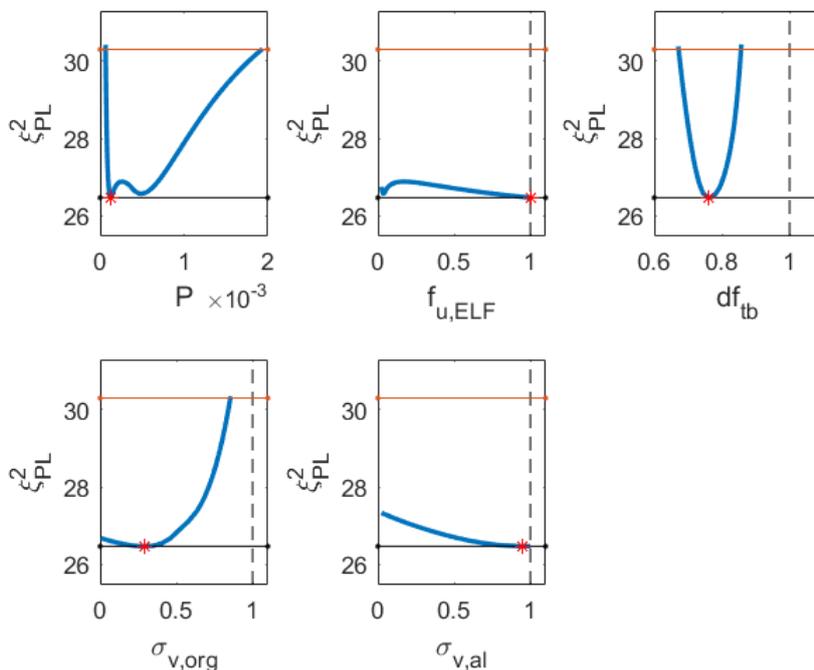
For the case with precise measurements there are no unidentifiabilities except for those parameters that are not incorporated in the model due to a certain administration route. We present here the case of IT, IV and DPI data which gives identifiability of all parameters, see Figure 6.11. The corresponding CIs are given in Table 6.10. As with the small molecule models, the CIs for the estimates of  $f_{u,ELF}$ ,  $D$  and  $C_s$  are wider than the other parameters but considerably smaller compared to the case of small molecules with low measurement error, see Table 6.2. A more thorough analysis with precise measurements, considering IT and IV data, is given in appendix E.



**Figure 6.11:** PLs of the parameters in the simplified large molecule model calculated with IT, IV and DPI data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

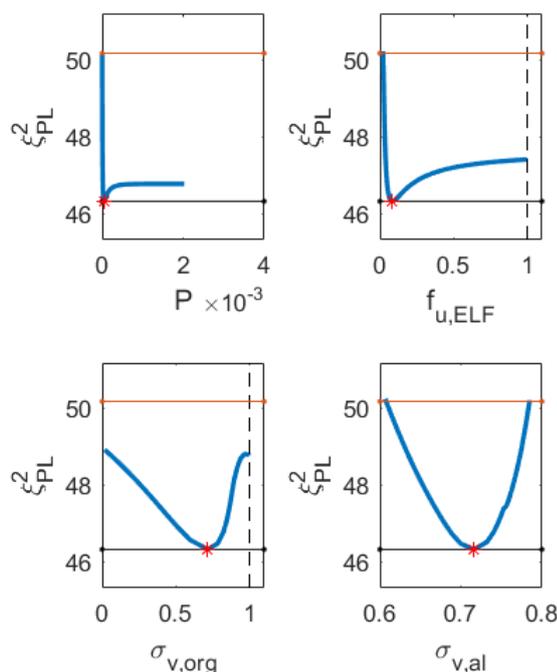
Low measurement error - IT/IV/DPI data			
Parameter	True value	Estimate	CI
$D$	0.0000394	0.0000406	[0.0000386, 0.0000428]
$C_s$	1000	992	[960, 1015]
$P$	0.0001	0.0000989	[0.0000974, 0.0001033]
$f_{u,ELF}$	1	1	[0.871, 1]
$df_{tb}$	0.7	0.699	[0.697, 0.701]
$\sigma_{v,org}$	0.6	0.602	[0.589, 0.616]
$\sigma_{v,al}$	0.78	0.780	[0.778, 0.782]

**Table 6.10:** Parameter estimates and confidence intervals (CIs) for the large molecule model with IT, IV and DPI data and low measurement error ( $\sigma_{ki} = 0.01$ ). The CIs are calculated from the profile likelihoods in figure 6.11 using spline interpolation to obtain the points of passover.



**Figure 6.12:** PLs for parameters in the simplified large molecule model with only IT administration and reasonable measurement errors. The data was simulated with log-normal error distribution and standard deviations set as in equation (6.2). Measurement times are given by equation (6.4). The parameters  $D$  and  $C_s$  were omitted as dissolution is not part of the model unless DPI administration is used.

Now with error set as in equation (6.2) and starting with only IT administration, the parameters  $P$  and  $df_{tb}$  are identifiable, see Figure 6.12. Some indication of identifiability of  $\sigma_{v,org}$  can be found but the data is not enough to obtain a well-determined estimate. For only IV administration  $\sigma_{v,al}$  is identifiable but some information on  $\sigma_{v,org}$  is found as well, although not enough to obtain a well-determined estimate, see Figure 6.13.  $P$  in this case is practically unidentifiable and  $f_{u,ELF}$  has a very large CI as before. These results show that  $\sigma_{v,org}$  is close to identifiable with only IT or IV data. Moreover, since  $P$  and  $df_{tb}$  are identifiable with IT data and  $\sigma_{v,al}$  is identifiable with IV data this suggests that the combined IT and IV data could be sufficient to obtain identifiability of all these parameters simultaneously.



**Figure 6.13:** PLs for parameters in the simplified large molecule model with only IV administration and reasonable measurement errors. The data was simulated with log-normal error distribution and standard deviations set as in equation (6.2). Measurement times are given by equation (6.4). The parameters  $D$  and  $C_s$  were omitted as dissolution is not part of the model unless DPI administration is used.  $df_{tb}$  is structurally unidentifiable as it is not part of the model with only IV administration.

The PL calculations were thus performed on 1000 different error realizations of combined IT and IV data, the summary is given in Table 6.11. Here the advantage of being able to sample the early phase of the system is clearly seen as  $f_{u,ELF}$  is identifiable in all runs, something which was only possible with measurements in the first couple of seconds for small molecules. The alveolar reflection coefficient is identifiable in all runs and  $P$  is identifiable in the majority of runs (97% of successful PL calculations). The vascular reflection coefficient in the organs (and by assumption also the central lung) is identifiable in roughly 50% of the runs. Moreover, the optimization routine only failed in 4 out of the total 5000 PL calculations (1000 runs

## 6. Results

for 5 parameters) which is significantly less compared to the small molecule cases.

Summary over 1000 error realizations - IT/IV data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$P$	998	97	6
$f_{u,ELF}$	1000	100	2
$df_{tb}$	1000	100	4
$\sigma_{v,org}$	998	53	5
$\sigma_{v,al}$	1000	100	5

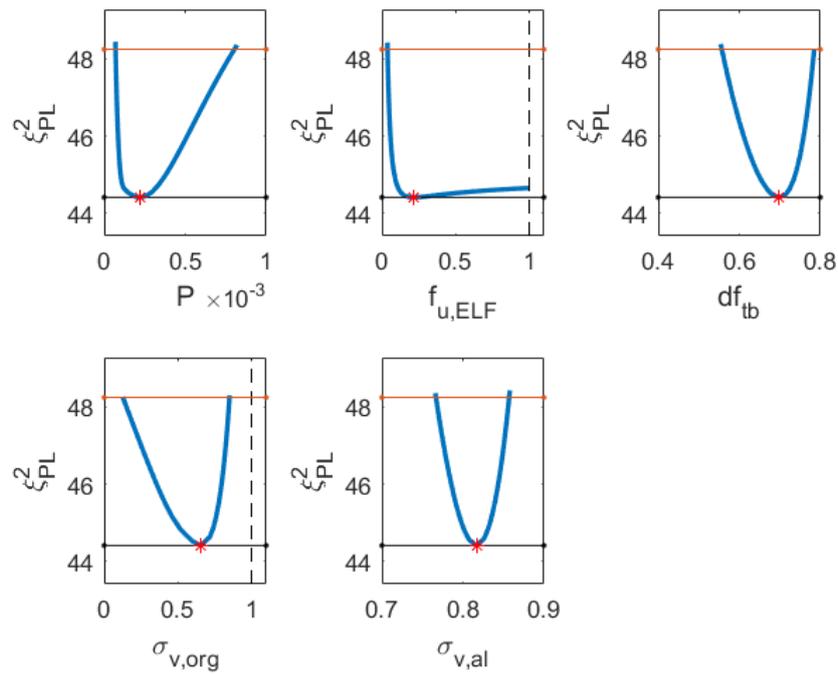
**Table 6.11:** Summary of PL calculation of 1000 different error realizations of IT and IV data for the simplified case of the large molecule model. The settings for the PL calculations were as in equation (6.1). The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.4). Total number of data points were thus  $n = 36$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.

Now to study the effect of uncertainty in the estimate of  $\sigma_{v,org}$  consider a case with identifiable  $\sigma_{v,org}$  as in Figure 6.14. The corresponding CIs are given in Table 6.12, were it should be noted that the CI for  $f_{u,ELF}$  is still very large.

Reasonable measurement error - IT/IV data			
Parameter	True value	Estimate	CI
$P$	0.0001	0.00022	[0.0000706, 0.000803]
$f_{u,ELF}$	1	0.217	[0.0383, 1]
$df_{tb}$	0.7	0.698	[0.557, 0.785]
$\sigma_{v,org}$	0.6	0.656	[0.127, 0.851]
$\sigma_{v,al}$	0.78	0.817	[0.767, 0.857]

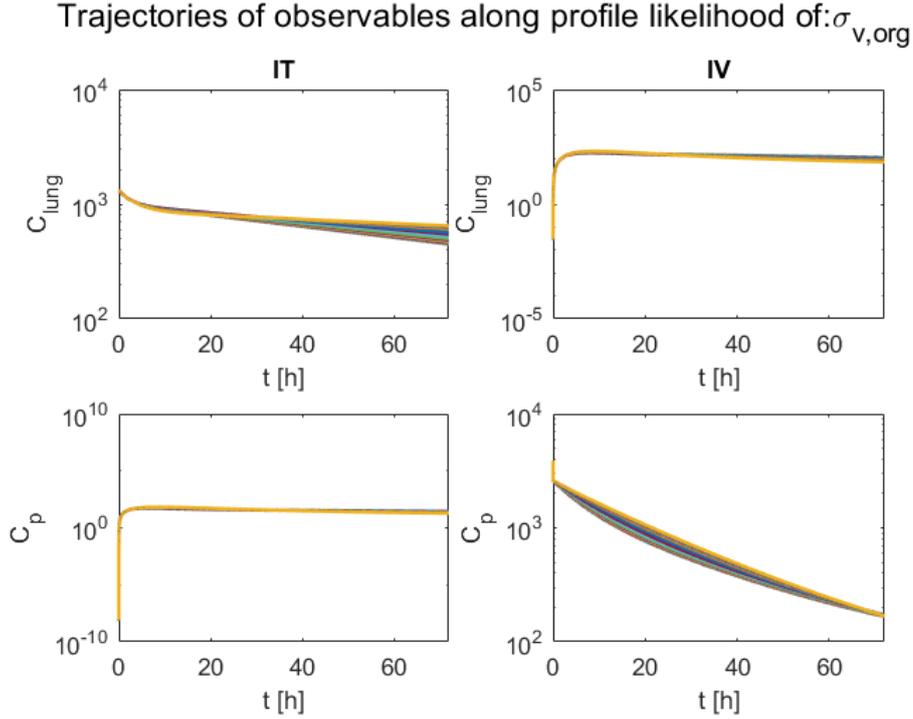
**Table 6.12:** Parameter estimates and confidence intervals (CIs) for the large molecule model with combined IT and IV data and reasonable measurement error. The CIs are calculated from the profile likelihoods in figure 6.14 using spline interpolation to obtain the points of passover. The parameters  $D$  and  $C_s$  were omitted as dissolution is not part of the model unless DPI administration is used

One can calculate the trajectories of observables  $C_{lung}(t; \theta^l)$  and  $C_{plasma}(t; \theta^l)$  for  $\theta^l$  along the PL of  $\sigma_{v,org}$ . Since  $\sigma_{v,org}$  is identifiable, these trajectories reflect how the uncertainty in the estimate of  $\sigma_{v,org}$  manifests in predictions of the model. As can be seen in Figure 6.15, there are some variability of the trajectories but the predictions are consistent. However, if one were to extrapolate this to longer time spans, it can be seen that the variability increases dramatically for  $C_{lung}(t)$  on the IT arm, see upper left panel in Figure 6.16. Moreover, there are two distinct bands of trajectories. A warning is therefore in place, that just because a parameter is identifiable does not justify extrapolating the results to unknown (i.e. unmeasured)



**Figure 6.14:** PLs for parameters in the large molecule model with IT and IV administration and reasonable measurement errors. The data was simulated with log-normal error distribution and standard deviations set as in equation (6.2). Measurement times are given by equation (6.4). The parameters  $D$  and  $C_s$  were omitted as dissolution is not part of the model unless DPI administration is used. For this particular error realization  $\sigma_{v,org}$  is identifiable.

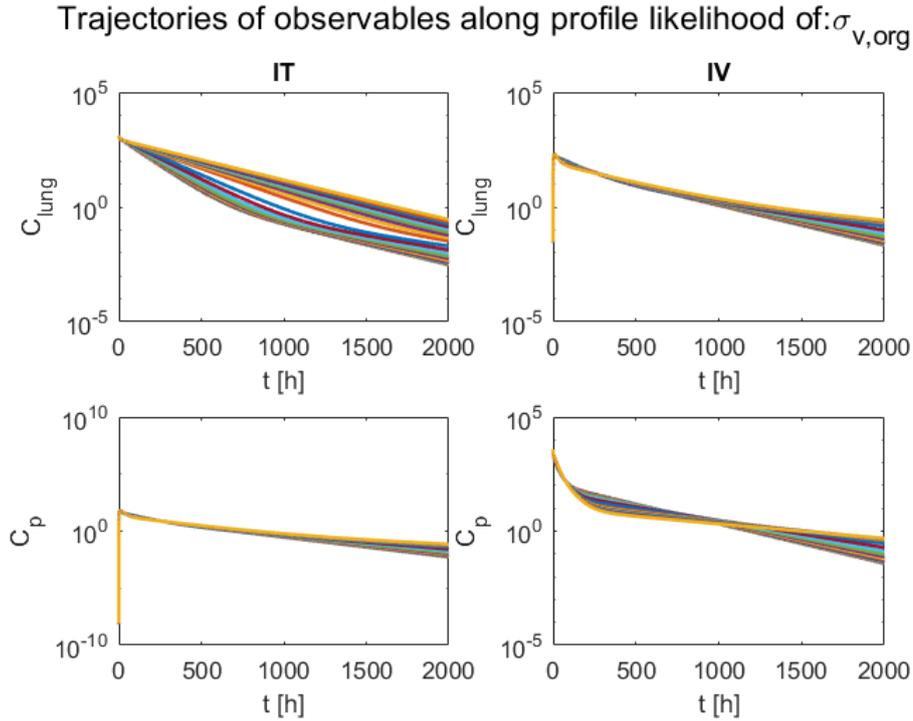
domains. Nonetheless, this analysis shows that additional measurements at later time points can improve the results for  $\sigma_{v,org}$ . Taking additional measurements at  $t = 200, 500$  and  $1000$ h on the IT arm gives identifiability of  $\sigma_{v,org}$  in 97% of error realizations, see Table E.1, as compared to only 53% without, see Table 6.11. Moreover, when it is identifiable the CI is also smaller than previously, see Figure E.7 and Table E.2.



**Figure 6.15:** Trajectories  $C_{plasma}(t; \theta^l)$  and  $C_{lung}(t; \theta^l)$  for  $\theta^l$  along the PL of  $\sigma_{v,org}$  in figure 6.14. Trajectories simulated over the same time span as the data was taken (72h).

An alternative approach to late measurements, is to add measurements of a new observable. First, one measurement at  $t = 72$ h on the IT arm of the total concentration in a richly perfused organ, given by  $C_{ri,tot}(t)$  as in equation (5.34), was tried. For a single error realization this lead to practical unidentifiability of  $\sigma_{v,org}$ , see Figure E.8. Similarly, taking one additional measurement of  $C_{ELF,avg}(t)$  at  $t = 72$ h did not improve the results, see Figure E.9. Thus the same dramatic improvement found in the neutral molecule case with only adding one measurement of a new observable is not seen here.

Lastly, to estimate the dissolution parameters, DPI data is added to the IT and IV data (without late measurements) to estimate  $D$  and  $C_s$ . 100 error realizations of combined IT, IV and DPI data shows that  $C_s$  is identifiable without need to fixate  $D$ , see Table 6.13.  $D$ , however, is only identifiable in ca 60% of the error realizations.



**Figure 6.16:** Trajectories  $C_{plasma}(t; \theta^l)$  and  $C_{lung}(t; \theta^l)$  for  $\theta^l$  along the PL of  $\sigma_{v,org}$  in figure 6.14. Trajectories simulated for 2000h.

Summary over 100 error realizations - IT/IV/DPI data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$D$	99	60	1
$C_s$	98	100	3
$P$	97	100	5
$f_{u,ELF}$	99	100	4
$df_{tb}$	97	100	4
$\sigma_{v,org}$	98	51	3
$\sigma_{v,al}$	98	100	8

**Table 6.13:** Summary of PL calculation of 100 different error realizations of IT, IV and DPI data for the simplified case of the large molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the DPI data was generated with log-normal errors and standard deviations same as for IT. The measurement times was set as in equation (6.4). Total number of data points were thus  $n = 54$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.

### 6.3.2 Unique case

Here it is assumed that the vascular reflection coefficients in central lung and organs are different. Thus the parameter  $\sigma_{v,tb}$  for the reflection coefficient in the central lung is added.

The analysis with low measurement error shows similar results as the simplified case with the main difference being the additional parameter  $\sigma_{v,tb}$  which is identifiable in all cases (only IT or IV, IT/IV, IT/IV and DPI). However, it was seen that  $P$  was practically unidentifiable for only IV administration.

Again, with reasonable measurement error the results are similar to the simplified case. With IT administration  $P$  and  $df_{tb}$  are identifiable. For the reflection coefficients there is a clear optimum for  $\sigma_{v,tb}$  but more data is needed to obtain a well-determined estimate, see Figure E.14. For the remaining reflection coefficients,  $\sigma_{v,org}$  and  $\sigma_{v,al}$ , the information after IT administration seems to be insufficient to obtain identifiability. With IV administration only  $\sigma_{v,al}$  is identifiable, see Figure E.15. As with the simplified case, the reflection coefficient in the alveolar part of the lung is identifiable with IV administration and the reflection coefficient in the trancheobronchular part of the lung is almost identifiable with IT administration. However, the reflection coefficient in the organs show no sign of identifiability for either administration route in the this case.

Running the analysis for 1000 error realizations of data from both IT and IV arms leads to identifiability of  $\sigma_{v,al}$  in all runs, see Table 6.14. The trancheobroncheolar reflection coefficient is identifiable slightly more often compared to  $\sigma_{v,org}$  in the simplified case. In only 122 cases was  $\sigma_{v,org}$  identifiable. The addition of DPI data

Summary over 1000 error realizations - IT/IV data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$P$	1000	98	6
$f_{u,ELF}$	1000	100	2
$df_{tb}$	1000	100	5
$\sigma_{v,org}$	1000	12	2
$\sigma_{v,al}$	1000	100	5
$\sigma_{v,tb}$	997	65	5

**Table 6.14:** Summary of PL calculation of 1000 different error realizations of IT and IV data for the complex case of the large molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.4). Total number of data points were thus  $n = 36$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.

gives identifiability of  $C_s$  in 100% of runs and identifiability of  $D$  in 64% of runs, see Table E.4. Since additional measurements on the IT arm at late time points

were seen to improve the number of identifiable cases of the central lung reflection coefficient in the simplified case, the same should apply for  $\sigma_{v,tb}$  here. Moreover, to see if this also helps in estimating the diffusion coefficient, three late measurements are added to the combined IT,IV and DPI data. The new measurements are the same as in the simplified case, i.e. at  $t = 200, 500$  and  $1000$ h on the IT arm. The results of running 100 error realizations is presented in Table 6.15. Here  $\sigma_{v,tb}$  is identifiable in all successful runs, as expected. However, there is not much improvement for  $D$  (73 out of 100 identifiable cases contra 64 without the additional measurements). The proportion of identifiable cases for  $\sigma_{v,org}$  increased slightly compared to the values in Table 6.14.

Summary over 100 runs - IT/IV/DPI with additional data for IT			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$D$	100	73	5
$C_s$	99	100	4
$P$	93	100	9
$f_{u,ELF}$	99	94	2
$df_{tb}$	100	100	6
$\sigma_{v,org}$	98	37	3
$\sigma_{v,al}$	100	100	1
$\sigma_{v,tb}$	90	100	8

**Table 6.15:** Summary of PL calculation of 100 different error realizations of IT, IV and DPI data for the complex case of the large molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the DPI data was generated with log-normal errors and standard deviations same as for IT. The measurement times was set as in equation (6.4) + at  $t = 200, 500$  and  $1000$  on the IT arm. Total number of data points were thus  $n = 60$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.



# 7

## Discussion and Conclusion

We have covered three PBPK models, two of which have a simplified and a complex version of parametrization and 4 different combinations of administration routes. Despite there being so many cases considered there are still a few overall conclusions that can be made. First, reliable estimates of both the in-vivo pulmonary permeability  $P$  and the deposition fraction  $df_{tb}$  can be found in all cases. Second, the standard setup (see beginning of chapter 6 and section 6.3) used for the analyses was not suitable to obtain reliable estimates of  $f_{u,ELF}$  since it resulted in large CIs even in identifiable cases. The remedy seems to be to measure the ELF-concentration, which also greatly improved the identifiability of  $V_{u,lung}$  in the small molecule case. Third, with reasonable measurement error, it was seen that combined IT and IV data were necessary or considerably better than only IT or IV data. However, a fair comparison was only done in the neutral molecule case where the number of data points were also equal. Lastly, with low measurement error identifiability of all parameters were possible, which rules out any problems of structural unidentifiability.

With neutral molecules it was seen that a well-determined estimate of  $V_{u,lung}$  could only be obtained in ca 70% of cases, depending on the error realization. This was for combined IT and IV data and did not improve with increasing the amount of data. In the case that  $V_{u,lung}$  is identifiable, a close lower bound for the CI was often not found. Initial measurement on the IT arm was considered as a hypothetical test, and served to show the methodology of using the PL approach to resolve unidentifiabilities. However, since the measurements needed to be in the first seconds (where variability was high) this approach would not be feasible in practice. Instead, if the ELF-concentration could be measured, this would lead to consistent results of identifiable parameter estimates. It was also seen that only one measurement of ELF-concentration would be needed in that case.

For basic molecules the main conclusion is that reliable estimates of the lysosomal input/output rate constants cannot be inferred with the current setup. Besides the rate constants, the results are mostly the same as for neutral molecules, which is expected since the model structure is identical with the exception of the lysosomal compartments. The rate constants could be made identifiable with a low measurement error. Thus, a future study of the effect of increasing the amount of data could be made to assess if and when the rate constants are consistently identifiable with

reasonable measurement error. Moreover, additional measurements of the drug concentrations in organs and/or ELF for basic molecules were only discussed briefly and a separate, thorough study of this would be informative. In particular, additional measurement of an organ would be of interest when considering the simplified case, since rate constants in the organs are assumed to be identical to the ones in the lung. Thus only a few measurements of the drug concentration in the organs could provide the necessary information to obtain reliable estimates. The question is then, how many measurements are needed and when should they be taken? Following such a study would be the more complex case, where rate constants in lung and organs are assumed to be unique. A more realistic model perhaps, but in this case measuring the organ is not likely to improve the identifiability of rate constants in the lung which we saw could not be obtained consistently with reasonable measurement error.

The DPI data is only added as a last step to see if  $D$  and  $C_s$  can be estimated, and for small molecules both parameters cannot be estimated simultaneously. However, since  $D$  can be calculated from Stokes-Einstein we may fixate this one and estimate  $C_s$  which in that case is consistently identifiable.

With large molecules we get identifiability of  $f_{u,ELF}$  but the estimate is still unreliable as the CI is large. It is also clear that a well-determined estimate of the alveolar reflection coefficient can be obtained using data with IV administration. In the more complex case, where reflection coefficients in central lung and organs are assumed to be different, we could obtain identifiability of the tracheobronchular reflection coefficient but not the reflection coefficient in the organs. In the simplified case, the two reflection coefficients are assumed to be identical and the similarities in the results indicates that the ability of  $\sigma_{v,org}$  to be estimated in the simplified case comes mostly from the central lung. It should also be noted that to estimate  $\sigma_{v,tb}$  one needs to observe the system at the late phase, since this is where the parameter impacts the observable the most. In our case with simulated data this is not an issue but in practice taking measurements several hundred hours after dosing is not common. Instead, with the assumptions in the simplified case, one could try to estimate this parameter by observing the drug-concentration in an organ (since  $\sigma_{v,tb}$  is assumed to be identical to  $\sigma_{v,org}$ ). We saw however, that only one measurement of the organ-concentration was not enough and a more thorough study of this particular case is needed. Similarly the addition of only one ELF-measurement did not seem to lead to consistent improvement.

The DPI data was more informative for large molecules than small molecules, since  $C_s$  could be determined without fixating  $D$ . The diffusion coefficient was also identifiable in some 60% of error realizations using the standard setup for the data. For the more complex model (with unique reflection coefficients) we also considered if observing the system's late phase on the IT arm could benefit the estimation of the diffusion coefficient. Although  $\sigma_{v,tb}$  was identifiable more often, just as the same analysis for the simplified model showed, we did not see a big improvement for the number of identifiable cases for the diffusion coefficient.

Lastly, it should be noted that in all cases where identifiability is obtained we did not consider optimizing the number of data points. This is no trivial task as there are countless ways to decide on the measurement times of each observable. The results of this project could nevertheless give a reasonable starting point for further analysis of this sort.

The profile likelihood method was seen to work well with these models overall and the implementation used here is flexible and can be reused in future projects.



# Bibliography

- [1] Amici online documentation. <http://icb-dcm.github.io/AMICI/index.html>. Accessed: 2019-04-24.
- [2] Einstein relation (kinetic theory). [https://en.wikipedia.org/wiki/Einstein\\_relation\\_\(kinetic\\_theory\)](https://en.wikipedia.org/wiki/Einstein_relation_(kinetic_theory)). Accessed: 2019-05-16.
- [3] Matlab online documentation. <https://se.mathworks.com/help/matlab/>. Accessed: 2019-04-24.
- [4] M. Schilling A. Raue, J. Bachmann, A. Matteson, and M. Schelker. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8(9), 2013.
- [5] E. Boger, N Evans, M Chappell, A Lundqvist, P Ewing, A Wigeborg, and M Fridén. Systems pharmacology approach for prediction of pulmonary and systemic pharmacokinetics and receptor occupancy of inhaled drugs. *CPT Pharmacometrics Syst. Pharmacol.*, 00(00), 2016.
- [6] Per Bäckman, Sumit Arora, William Couet, Ben Forbes, Wilbur de Keuij, and Amrit Paudel. Advances in experimental and mechanistic computational models to understand pulmonary exposure to inhaled drugs. *European Journal of Pharmaceutical Sciences*, 113:41–52, 2018.
- [7] S. Y. Amy Cheung, James W. T. Yates, and Leon Aaron. The design and analysis of parallel experiments to produce structurally identifiable models. *J Pharmacokinet Pharmacodyn*, 40:93–100, 2013.
- [8] Fabian Frölich, Fabian J. Theis, Joachim O.Rädler, and Jan Hasenauer. Parameter estimation for dynamical systems with discrete events and logical operations. *Bioinformatics*, 33(7):1049–1056, 2016.
- [9] Ramon Hendrickx, Eva Lamm Bergström, David L. I. Janzén, Markus Fridén, Ulf Eriksson, Ken Grime, and Douglas Ferguson. Translational model to predict pulmonary pharmacokinetics and efficacy in man for inhaled bronchodilators. *CPT Pharmacometrics Syst Pharmacol*, 7(3):147–157, 2018.
- [10] Clemens Kreutz, A. Raue, and Jens Timmer. Likelihood base observability

- analysis and confidence intervals for predictions of dynamical models. *BMC Systems Biology*, 6(120), 2012.
- [11] Clemens Kreutz, Andreas Raue, Daniel Kaschek, and Jens Timmer. Profile likelihood in systems biology. *FEBS Journal*, 280:2564–2571, 2013.
- [12] R Boiger nad J Hasenauer, S Hro, and B Kaltenbacher. Integration based profile likelihood calculation for pde constrained parameter estimation problems. *Inverse Problems*, 32(12), 2016.
- [13] Ali Nokhodchi and Gary P. Martin. *Pulmonary Drug Delivery: Advances and Challenges*. John Wiley Sons, s, 2015.
- [14] A. Raue, C. Kreutz, T. Maiwalda, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.
- [15] A. Raue, T. Maiwald, C. Kreutz, U. Klingmüller, and J. Timmer. Addressing parameter identifiability by model-based experimentation. *IET Systems Biology*, 5(2):120–130, 2011.
- [16] R.Bellman and K.J.Åström. On structural identifiability. *Mathematical Biosciences*, 7(3-4):329–339, 1970.
- [17] Maria Pia Saccomani. Structural vs practical identifiability in system biology. *Granada*, 2013.

# A

## Test problems

To evaluate the performance of the profile-likelihood program it is tested on three problems where analytical solutions are known. One of these is given in section 4.4 and the other two are presented here.

### A.1 Problem 1: A structurally unidentifiable model

The model is given by a single ODE as

$$\frac{dx(t)}{dt} = -(k_1 + k_2)x(t), \quad x(0) = x_0 \quad (\text{A.1})$$

and the measurements are

$$y(t_i) = k_3x(t_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i), \quad i = 1, \dots, N$$

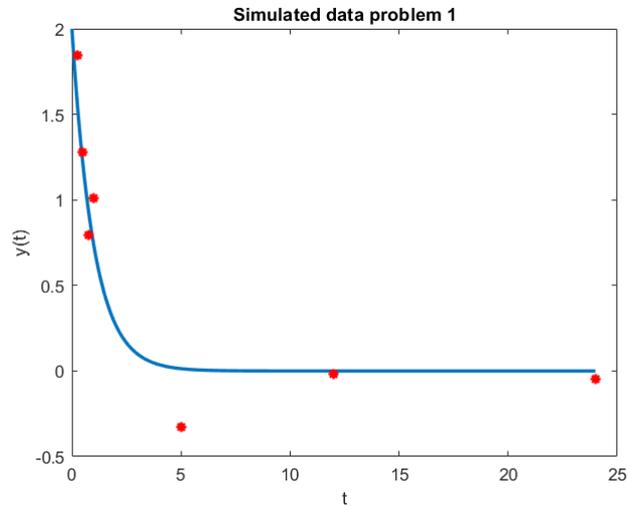
Assuming that  $x_0$  is known the parameter  $k_3$  is identifiable but  $k_1$  and  $k_2$  are structurally unidentifiable as only the sum  $k_1 + k_2$  is well-determined by the model, i.e. we have a functional relationship between the parameters  $k_1$  and  $k_2$  as in  $k_1 + k_2 = \text{const}$ .

The true parameter values were set as in Table A.1 and 7 observed time points are chosen as

$$\mathbf{t}_{obs} = [0.25, 0.5, 0.75, 1, 5, 12, 24]. \quad (\text{A.2})$$

Simulation of measurements taken with additive Gaussian error distribution and  $\sigma_i = 0.2, \forall i = 1, \dots, 7$  is presented in figure A.1. The global estimation procedure was performed using *lsqnonlin* with the *trust-region-reflective* algorithm and 21 different start points. The parameter ranges were set to  $[-5, 5]$  for all three parameters  $k_1, k_2$  and  $k_3$ . For this setting the global optimum was found as in Table A.1.

Calculation of the profile likelihood for  $k_1, k_2$  and  $k_3$  results in figure A.2 where it can clearly be seen that  $k_1$  and  $k_2$  are structurally unidentifiable. This analysis shows that just performing a global estimation is not enough as the estimate of  $k_2$  is close to its true value but in reality the parameter is not identifiable. Furthermore we can plot the parameter values of  $k_1$  and  $k_3$  along the profile likelihood of  $k_2$  to recover the functional relationship, see figure A.3. A clear negative relationship between  $k_1$

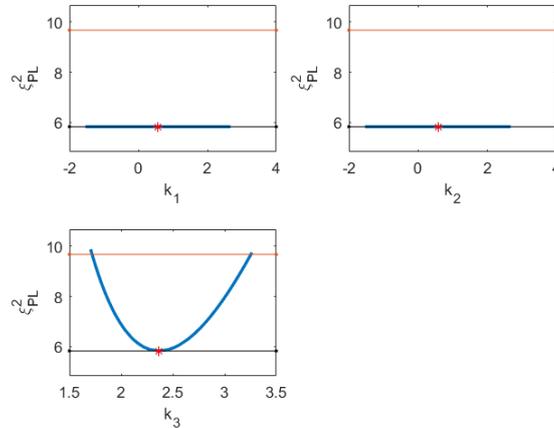


**Figure A.1:** Simulated model trajectory in blue and observed data in red. The observable was assumed to be normally distributed with standard deviation set to 0.2 for all measurements.

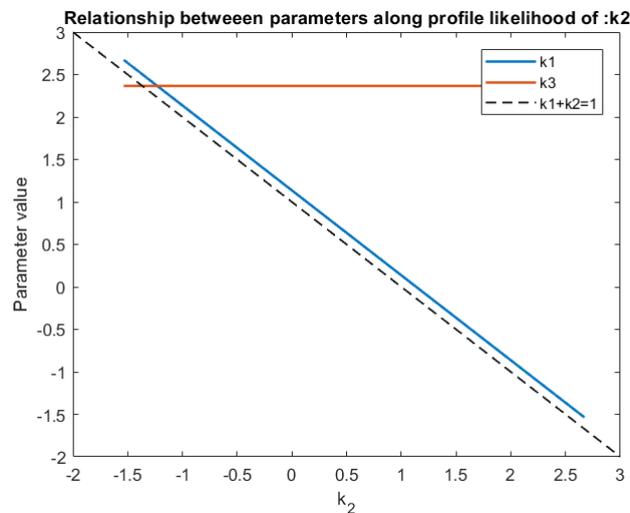
Parameter	True value	Estimate
$k_1$	0.333	0.56825
$k_2$	0.667	0.56825
$k_3$	2	2.3666

**Table A.1:** Arbitrarily set true parameter values and their global estimates for problem 1. The estimation was performed using *lsqnonlin* with the *trust-region-reflective* algorithm.

and  $k_2$  is seen which is expected as the functional relationship between  $k_1$  and  $k_2$  is  $k_1 + k_2 = \text{const}$ . In this particular case it is  $k_1 + k_2 = 1$  but due to measurement error we obtain  $\hat{k}_1 + \hat{k}_2 = 1.13$ .



**Figure A.2:** Profile likelihoods for parameters in problem 1:  $k_1$  and  $k_2$  are structurally unidentifiable,  $k_3$  is identifiable.  $\Delta_\alpha = 95\%$ -quantile of  $\chi_1^2$ . Note that the true value  $k_3 = 2$  is included in the confidence interval.



**Figure A.3:** The values of parameters  $k_1$  and  $k_3$  along the profile likelihood of  $k_2$  plotted against  $k_2$ . Note that this plot can only detect functional relationships of one parameter (in this case  $k_2$ ) as only one parameter is fixed at a time during the calculation of profile likelihood.

## A.2 Problem 2

The model is given by the ODE

$$\frac{dx(t)}{dt} = -k_2x(t), \quad x(0) = k_1x_0 \quad (\text{A.3})$$

and the observable is

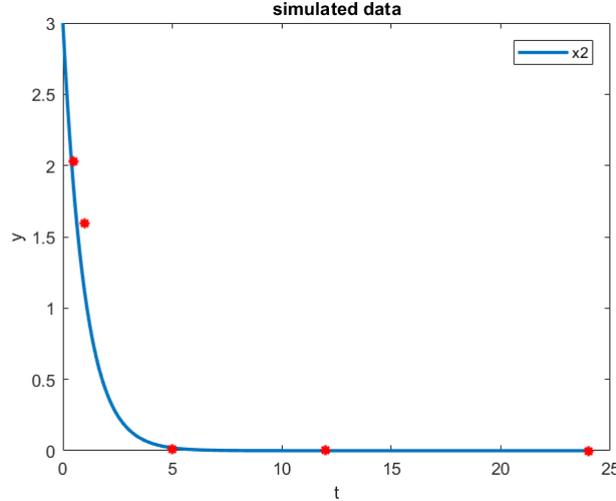
$$y(t_i) = k_3x(t_i) \cdot e_i, \quad e_i \sim \log\mathcal{N}(0, \sigma_i).$$

Assuming  $x_0$  is known we have the parameters  $\theta = (k_1, k_2, k_3)$  but only  $k_2$  and the product  $k_1 \cdot k_3$  are well-determined [7]. Here the parameters  $k_i$ ,  $i = 1, 2, 3$  are assumed to be positive.

Measurement time points for the observable are chosen as

$$t_{obs} = [0.5, 1, 5, 12, 24].$$

and the log-normal standard deviations for the measurement error are set to  $\sigma_i = 0.2$ ,  $\forall i$ . The simulated data is shown in figure A.4 and the true parameter values along with the estimates obtained after global optimization using *trust-region-reflective* is presented in Table A.2.

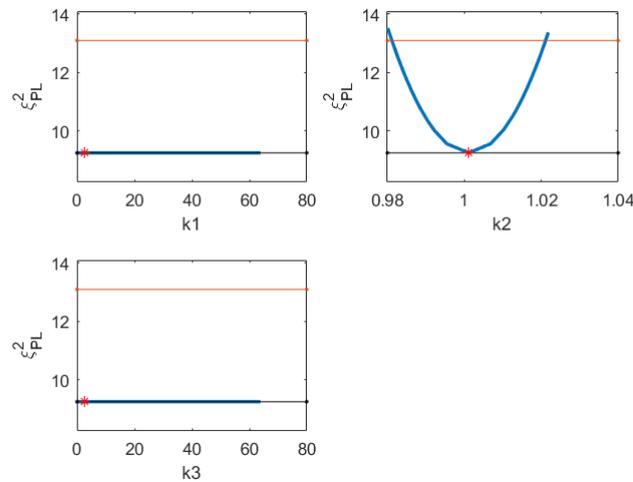


**Figure A.4:** Simulated model trajectory in blue and observed data in red. The observable was assumed to be log-normal with standard error set to 0.2 for all measurements.

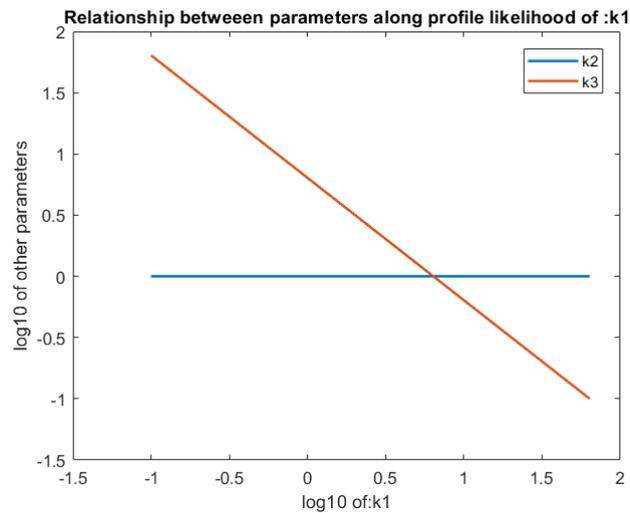
The profile likelihoods for the parameters are presented in figure A.5. As expected  $k_2$  is the only identifiable parameter and looking at the plot for  $k_2$  and  $k_3$  along the profile likelihood of  $k_1$  (figure A.6) we see a negative relationship between the logarithm of  $k_1$  and  $k_3$ , which means that  $\log(k_1 \cdot k_3) = const. \iff k_1 \cdot k_3 = const..$

Parameter	True Value	Estimate
$k_1$	2	2.34
$k_2$	1	1.045
$k_3$	3	2.34

**Table A.2:** True parameter values and global estimates found by optimization for problem 3.



**Figure A.5:** Profile likelihoods for parameters in problem 3.



**Figure A.6:** The parameter values of  $k_2$  and  $k_3$  along the profile likelihood of  $k_1$  plotted against  $k_1$ . Note the use of log10-parametrization. We see a negative relationship between  $k_1$  and  $k_3$  while  $k_2$  is unaffected.

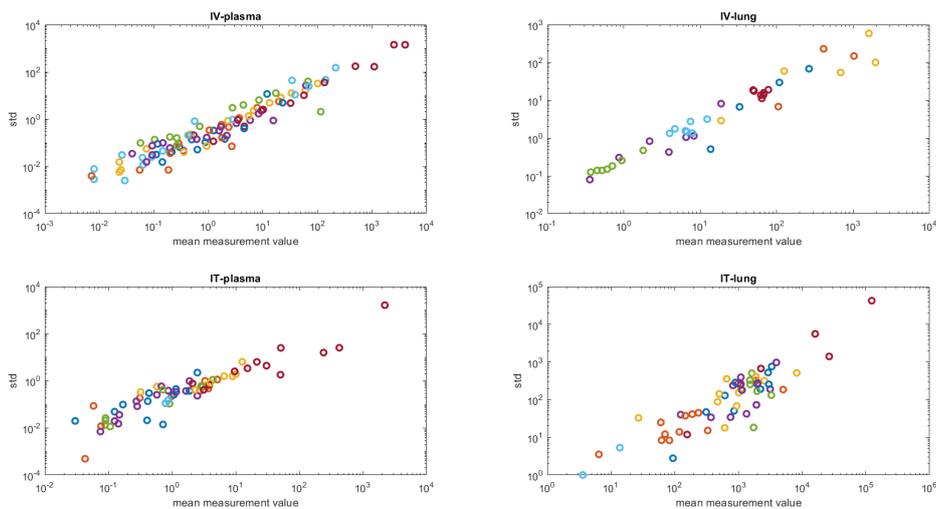


# B

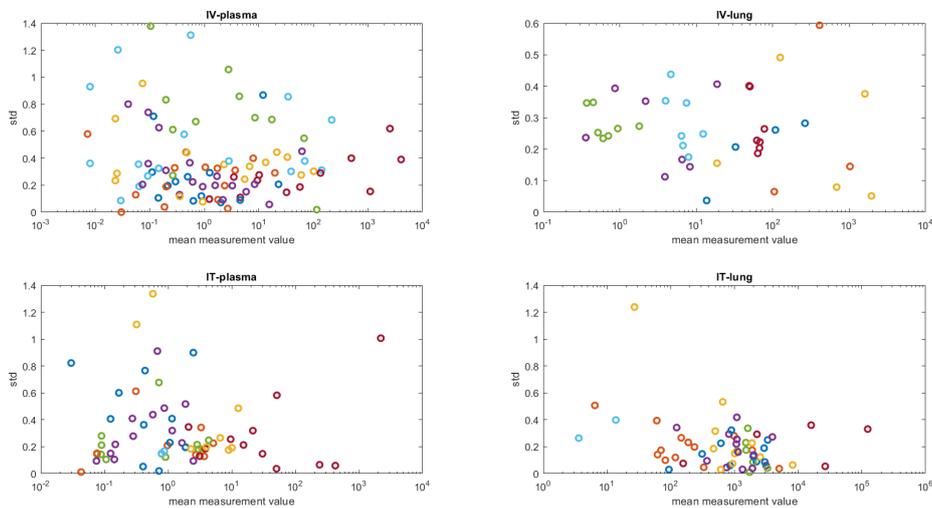
## Calculation of reasonable measurement error

The measurement error for simulated data used in this project was estimated using data from Hendrickx et.al [9]. The data provided is for 12 different compounds in the small molecule category with experimental measurements for  $C_{lung}$  and  $C_{plasma}$  taken with IT and IV administration. The measurement error was estimated by taking an average of the sample standard deviation calculated for all measurement times with more than 1 measurements for each of the cases IT- $C_{lung}$ , IT- $C_{plasma}$ , IV- $C_{plasma}$  and IV- $C_{lung}$ . This was done in two steps. First the data was split into the 4 groups: IT- $C_{lung}$ , IT- $C_{plasma}$ , IV- $C_{plasma}$  and IV- $C_{lung}$  for all compounds. Then, for each group, the sample standard deviation was calculated for measurements at each measurement time when the number of measurements was more than 1. Plotting the standard deviation against the mean of measured values shows a clear increasing trend, see figure B.1, indicating that the error is log-normal rather than normally distributed. Hence, the standard deviations of the log of measurements (at each measurement time) is calculated to obtain the standard deviations assuming log-normal error distribution. Plotting again against the mean of measurements in figure B.2 shows no clear groupings or trends, indicating that the measurement error doesn't have any strong dependency on when the measurement is taken. Therefore it will be assumed that the error is log-normally distributed with constant standard deviation over time, i.e. no time-dependency in the measurement error. To obtain a value for the logarithmic standard deviation for each group the average of the standard errors over all measurement times and compounds is taken to give the values as in Table B.1.

## B. Calculation of reasonable measurement error



**Figure B.1:** Standard deviations vs mean of measurement of data from [9], plotted on log-log scale. For each group (IT -  $C_{plasma}$ , IT -  $C_{lung}$ , IV -  $C_{plasma}$ , IV -  $C_{lung}$ ) the standard deviations were estimated from the experimental measurements whenever more than one measurements were taken at the same time. The x-axis is the corresponding mean of measurement values at that time. The different colors represent the 12 different compounds considered in [9]. Note that there is no grouping of colors.



**Figure B.2:** Logarithmic standard deviations vs mean of measurement of data from [9], plotted with logarithmic x-axis. For each group (IT -  $C_{plasma}$ , IT -  $C_{lung}$ , IV -  $C_{plasma}$ , IV -  $C_{lung}$ ) the logarithmic standard deviations were estimated from the experimental measurements whenever more than one measurements were taken at the same time. The x-axis is the corresponding mean of measurement values at that time. The different colors represent the 12 different compounds considered in [9].

Group	std
IT - $C_{plasma}$	0.3159
IT - $C_{lung}$	0.2056
IV - $C_{plasma}$	0.3752
IV - $C_{lung}$	0.2609

**Table B.1:** Measurement errors estimated from data.

Since no DPI data was provided it is assumed that the standard deviations for DPI administration is the same as for IT administration since these are most similar. For measurements of additional observables such as  $C_{ELF}$  or  $C_{ri}$  the standard deviation is to 0.3. These values will be used also for large molecules since no other data was available.

# C

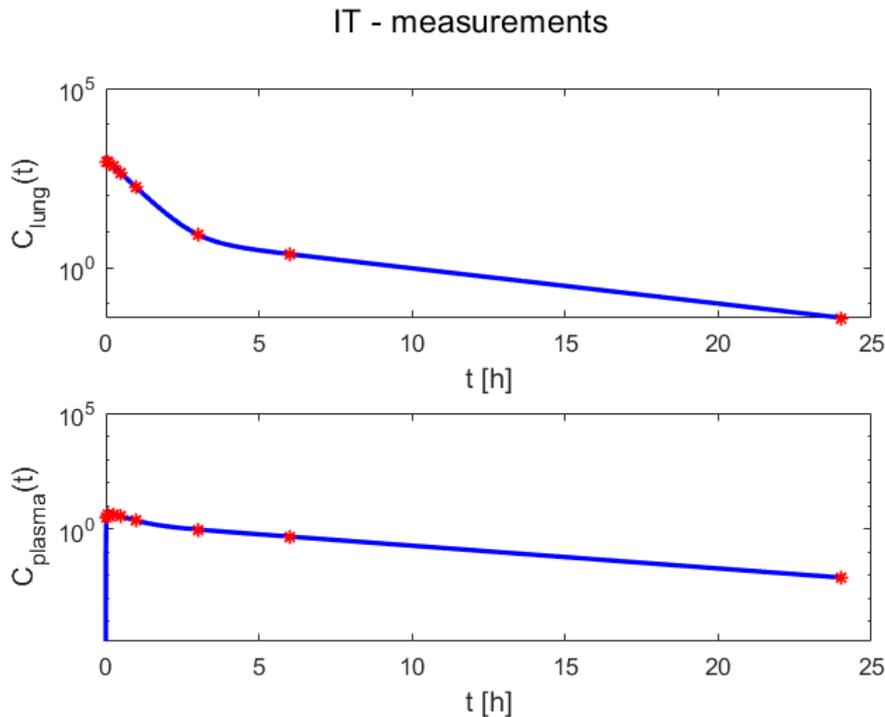
## Neutral molecules

The analysis with low measurement error and additional Figures and Tables from section 6.1 are given here.

### C.1 Analysis for Low measurement error

All data simulated with log-normal errors with standard deviation 0.01, i.e.  $e_{ki} \sim \log\mathcal{N}(0, 0.01)$ ,  $\forall k, i$  at time points given by 6.1.

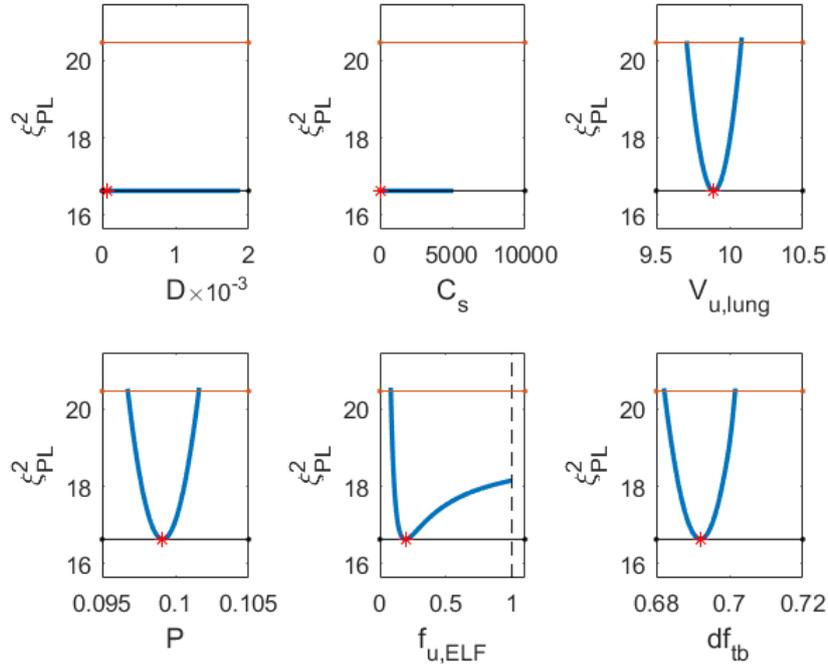
With only IT administration the simulated data is seen in Figure C.1. Global optimization to find the estimates, followed by PL calculation gives Table C.1 and Figure C.2. As expected,  $D$  and  $C_s$  are structurally unidentifiable. All parameters except  $D$  and  $C_s$  are identifiable since both upper and lower bounds for the CIs are found. Looking at the profile for  $f_{u,ELF}$  in Figure C.2 it is reasonable to suggest that it is practically unidentifiable if the upper bound of  $f_{u,ELF} \leq 1$  is removed, but this is not considered here since  $f_{u,ELF}$  is per definition a fraction. However, the CI for  $f_{u,ELF}$  is much larger than the others (Table C.1) indicating large uncertainty for this parameter even with precise measurements.



**Figure C.1:** Simulated trajectories of the observables  $C_{plasma}(t)$  and  $C_{lung}(t)$  for the neutral molecules model with IT administration. The red stars are simulated data taken with low measurement error ( $\sigma_{ki} = 0.01$ ) at times given by equation (6.1).

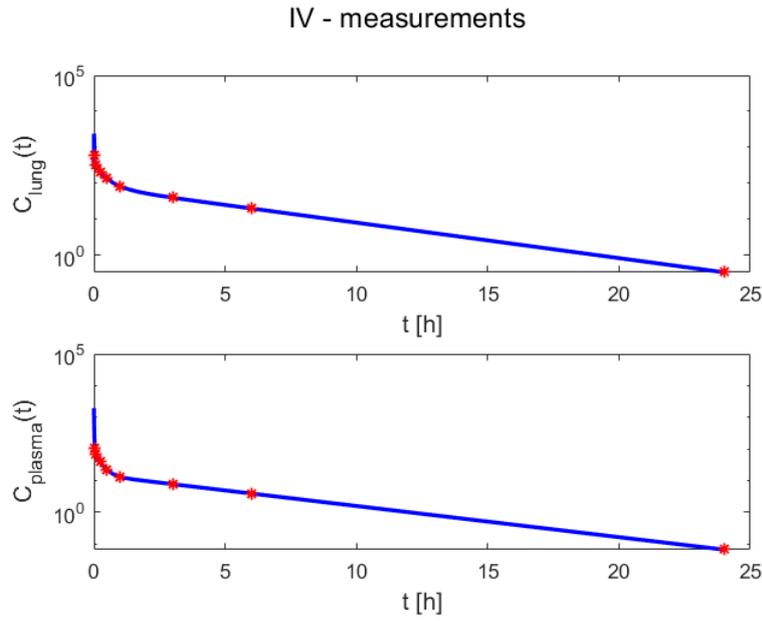
Low measurement error - IT data only, $n = 16$			
Parameter	True value	Estimate	CI
$D$	0.000305	0.000063	$(0, \infty)$
$C_s$	1000	50.1	$(0, \infty)$
$V_{u,lung}$	10	9.89	[9.708, 10.08]
$P$	0.1	0.099	[0.0968, 0.103]
$f_{u,ELF}$	1	0.19	[0.083, 1]
$df_{tb}$	0.7	0.69	[0.682, 0.702]

**Table C.1:** Parameter estimates and confidence intervals (CIs) for the neutral molecule model with only IT administration and low measurement error ( $\sigma_{ki} = 0.01$ ). The total number of measurements were  $n = 16$  points taken as in figure C.1. The CIs are calculated from the profile likelihoods in figure C.2 using spline interpolation to obtain the points of passover.  $D$  and  $C_s$  are structurally unidentifiable.  $f_{u,ELF}$  has a large CI. The parameters  $V_{u,lung}$ ,  $P$  and  $df_{tb}$  have narrow CIs.

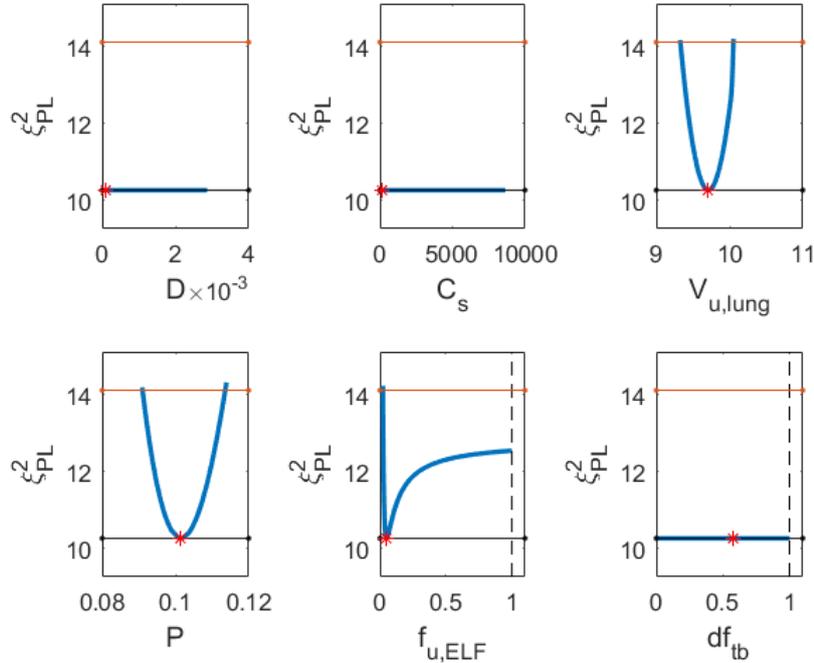


**Figure C.2:** Profile likelihoods of parameter estimates in the neutral molecule model with only IT administration and  $n = 16$  data points taken as in figure C.1. The log-normal standard deviations for the measurement errors were set to  $\sigma_{k,i} = 0.01 \forall k, i$ .  $D$  and  $C_s$  are not incorporated in the model unless drug is administered by DPI. Note that the profile of  $f_{u,ELF}$  suggest it is practically unidentifiable (no upper bound) if it is allowed to take on values greater than 1. The parameters  $V_{u,lung}$ ,  $P$  and  $df_{tb}$  are clearly identifiable.

In the case of only IV-administration the parameter  $df_{tb}$  will be unidentifiable since it is not incorporated in the model structure. With data as in Figure C.3 the estimates and CIs in Table C.2 are obtained. Compared to IT-administration with the same measurement error and measurement times the CIs are wider in the case of IV-administration.



**Figure C.3:** Simulated trajectories of the observables  $C_{plasma}$  and  $C_{lung}$  for the neutral molecules model with IV administration. The red stars are simulated data taken with low measurement error ( $\sigma_{k_i} = 0.01$ ) at times given by equation (6.1).



**Figure C.4:** Profile likelihoods of parameter estimates in the neutral molecule model with only IV administration and  $n = 16$  data points taken as in figure C.1. The log-normal standard deviations for the measurement errors were set to  $\sigma_{k,i} = 0.01 \forall k, i$ . The structurally unidentifiable parameters  $D, C_s$  and  $df_{tb}$  are not incorporated in the model structure unless drug is administrated by DPI.

Low measurement error - IV data only			
Parameter	True value	Estimate	CI
$D$	0.000305	0.0001	$(0, \infty)$
$C_s$	1000	100	$(0, \infty)$
$V_{u,lung}$	10	9.71	[9.33, 10.05]
$P$	0.1	0.10146	[0.091, 0.114]
$f_{u,ELF}$	1	0.050973	[0.0232, 1]
$df_{tb}$	0.7	0.57735	$(0, 1)$

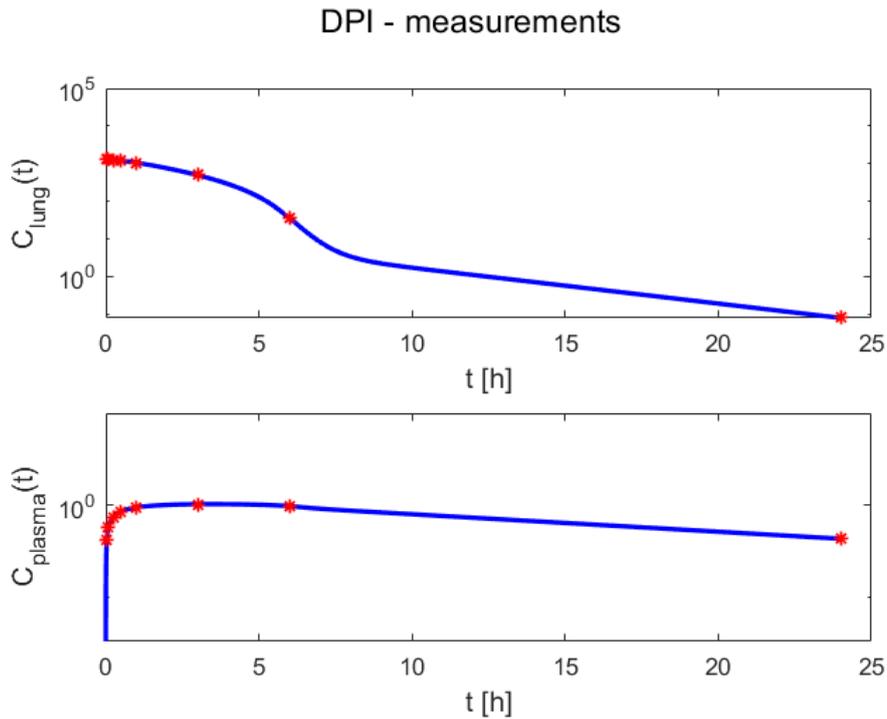
**Table C.2:** Parameter estimates and confidence intervals (CIs) for the neutral molecule model with only IV administration and low measurement error ( $\sigma_{ki} = 0.01$ ). The CIs are calculated from the profile likelihoods in figure C.4 using spline interpolation to obtain the points of passover.  $D$ ,  $C_s$  and  $df_{tb}$  are structurally unidentifiable. The CIs for the other parameters are slightly wider compared to the case of only IT-administration in Table C.1.

When estimating over the combined data set of IT administration and IV administration the results are similar to the case of only IT administration but with slightly narrower CIs (see Table C.3), implying that IV-data is not necessary to obtain well-determined estimates if precise measurements are possible.

Low measurement error - IT/IV data			
Parameter	True value	Estimate	CI
$D$	0.000305	0.000159	$(0, \infty)$
$C_s$	1000	200	$(0, \infty)$
$V_{u,lung}$	10	10.02	[9.91, 10.11]
$P$	0.1	0.1	[0.099, 0.101]
$f_{u,ELF}$	1	0.430	[0.118, 1]
$df_{tb}$	0.7	0.695	[0.686, 0.702]

**Table C.3:** Parameter estimates and confidence intervals (CIs) for the neutral molecule model with IT and IV data and low measurement error ( $\sigma_{ki} = 0.01$ ). The CIs are calculated from the profile likelihoods in figure C.4 using spline interpolation to obtain the points of passover.  $D$  and  $C_s$  are structurally unidentifiable.

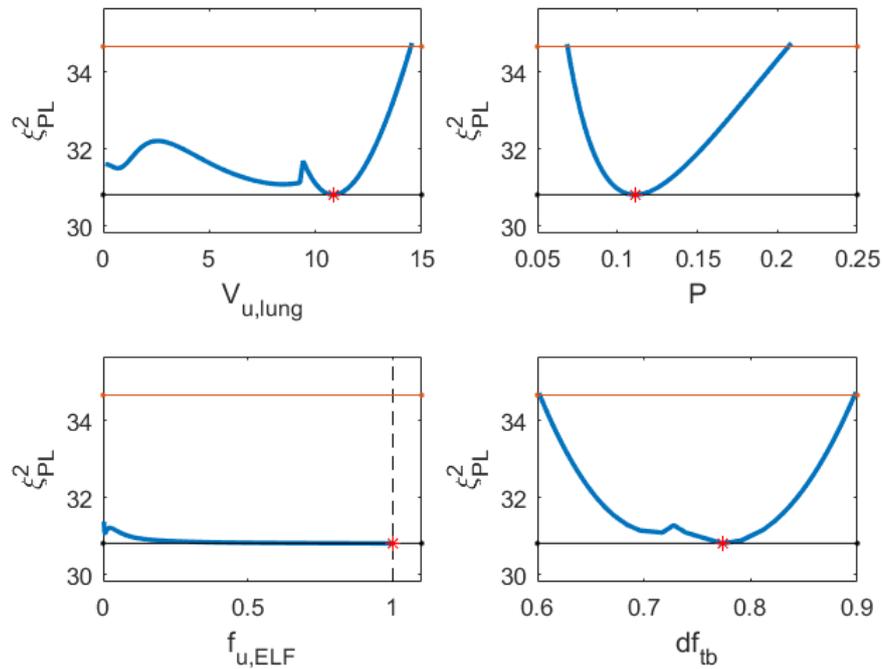
Lastly, with additional DPI-data to the IT and IV data, information about the dissolution process of the drug compound is also incorporated. Hence the goal here is to be able to estimate  $D$  and  $C_s$  together with the other parameters. The additional DPI data is shown in Figure C.5. Calculation of the profile likelihoods with the combined IT, IV and DPI data yields Figure 6.1 and Table 6.2 in the main text. Even though all parameters are identifiable it is clear that the CIs of  $D$ ,  $C_s$  and  $f_{u,ELF}$  are much larger than the other parameters.



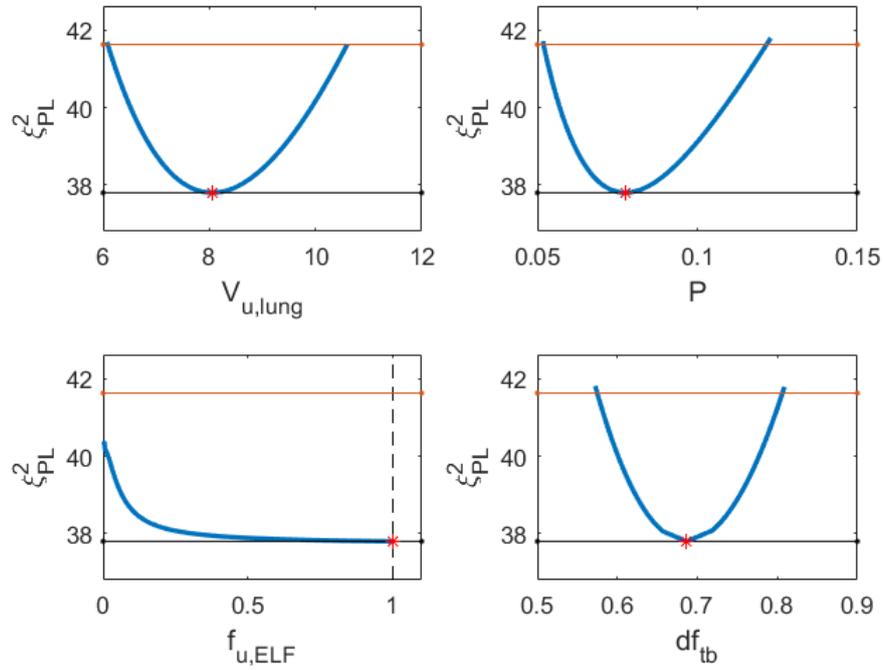
**Figure C.5:** Simulated trajectories of the observables  $C_{plasma}$  and  $C_{lung}$  for the neutral molecules model with DPI administration. The red stars are simulated data taken with low measurement error ( $\sigma_{ki} = 0.01$ ) at times given by equation (6.1). max steps was set to 1000 in this case.

## C.2 Additional Figures and Tables - Reasonable measurement error

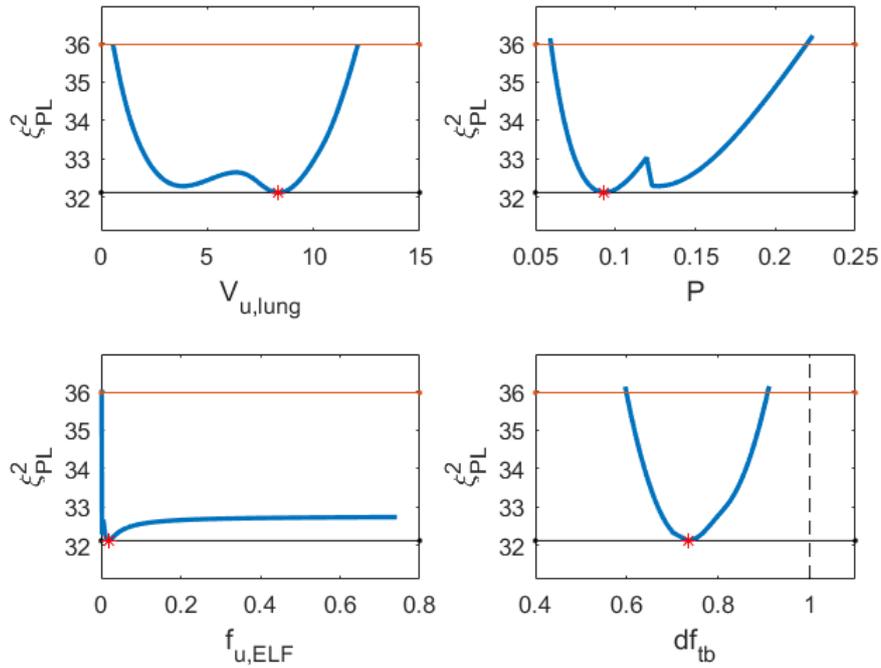
With only IT or IV administration, the data is insufficient to obtain consistent results. Here are some particular cases of PLs calculated for different error realizations.



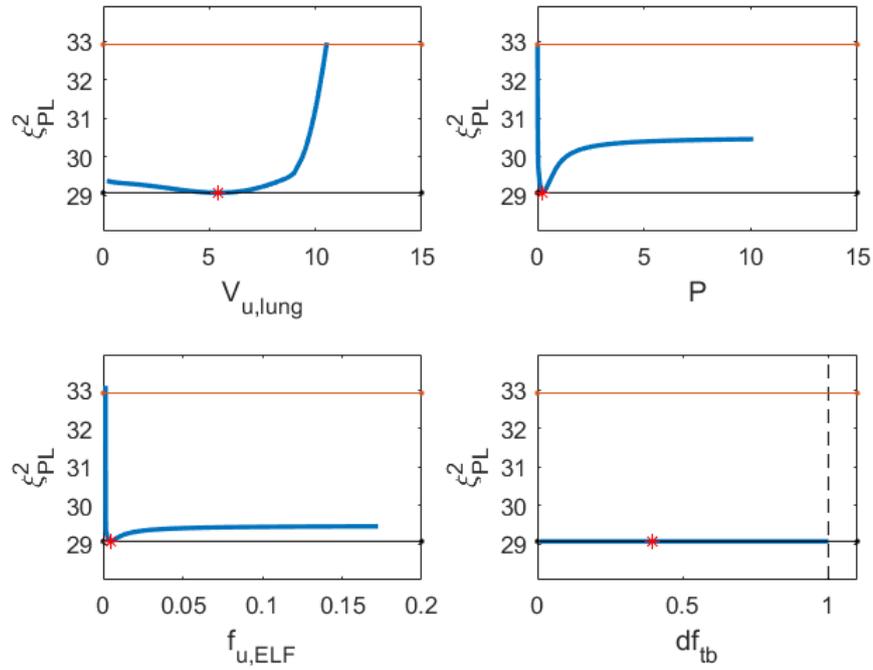
**Figure C.6:** A case when the random error realization leads to  $V_{u,lung}$  and  $f_{u,ELF}$  being practically unidentifiable. Profile likelihoods of the parameters calculated for the neutral molecule model with only IT administration. The parameters  $D$  and  $C_s$  were excluded as they are not incorporated in the model with only IT administration. The data was given by  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  for the neutral model taken at time points given by equation (6.1) and log-normal errors with standard deviations set as in equation (6.2). max steps was set to 1000 in this case.



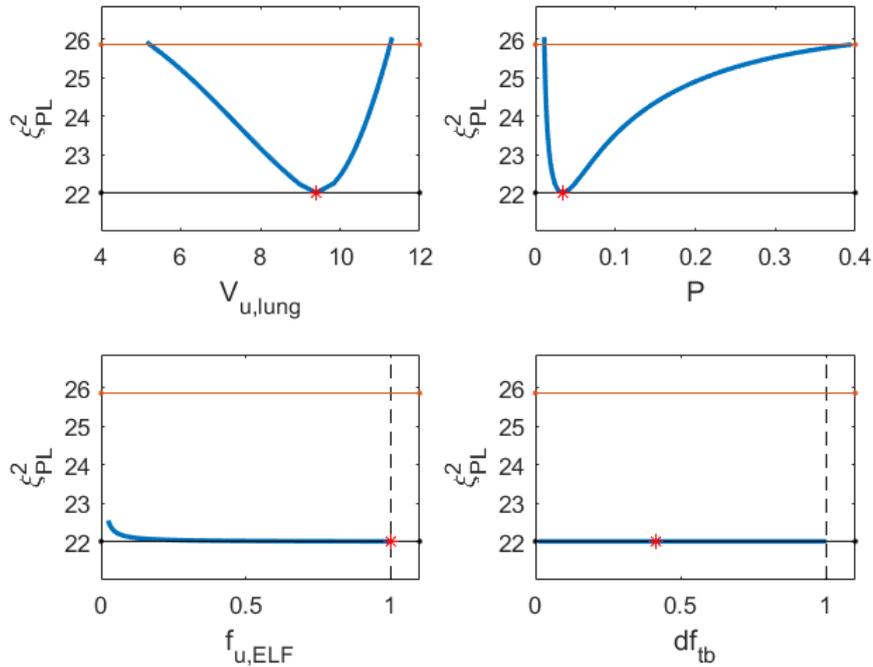
**Figure C.7:** A case when the random error realization leads to  $V_{u,lung}$  being identifiable with a narrow CI. Profile likelihoods of the parameters calculated for the neutral molecule model with only IT administration. The parameters  $D$  and  $C_s$  were excluded as they are not incorporated in the model with only IT administration. The data was given by  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  for the neutral model taken at time points given by equation (6.1) and log-normal errors with standard deviations set as in equation (6.2).



**Figure C.8:** A case when the random error realization leads to identifiability of  $V_{u,lung}$  but with a wide CI.  $f_{u,ELF}$  is considered identifiable in this case even though the profile likelihood shows practical unidentifiability, this is because an upper bound of 1 is imposed since the parameter represents a fraction. The calculation of the PL of  $f_{u,ELF}$  was stopped before it reached 1 in this case. Profile likelihoods of the parameters calculated for the neutral molecule model with only IT administration. The parameters  $D$  and  $C_s$  were excluded as they are not incorporated in the model with only IT administration. The data was given by  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  for the neutral model taken at time points given by equation (6.1) and log-normal errors with standard deviations set as in equation (6.2).



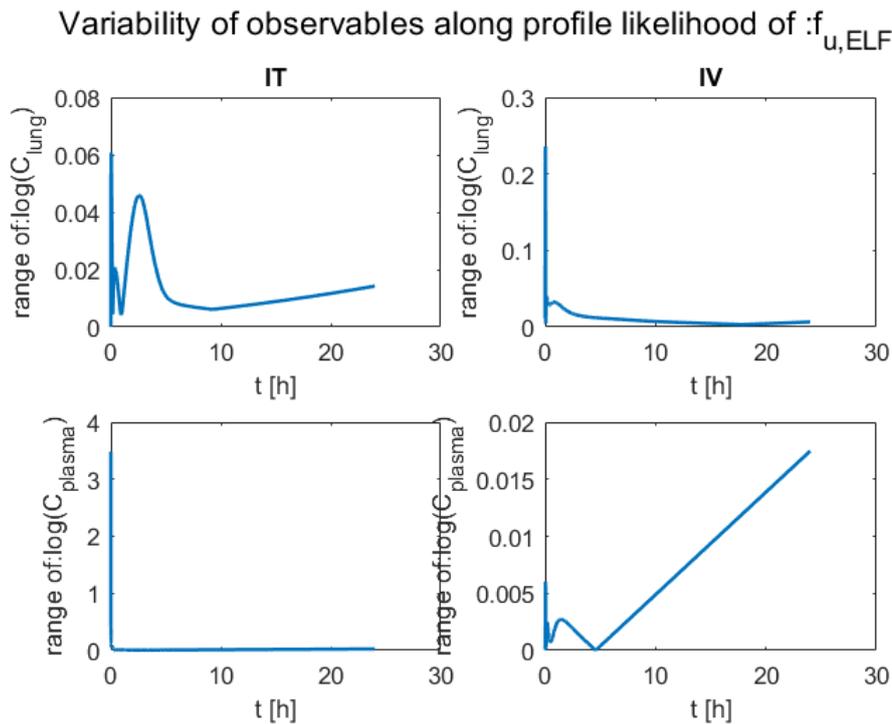
**Figure C.9:** A case when the random error realization for data with IV administration leads to practical unidentifiability of  $V_{u,lung}$  and  $P$ . The parameters  $D$  and  $C_s$  were excluded. The data was given by  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  for the neutral model taken at time points given by equation (6.1) and log-normal errors with standard deviations set as in equation (6.2).



**Figure C.10:** A case when the random error realization for data with IV administration leads to identifiability of  $V_{u,lung}$  and  $P$ . The parameters  $D$  and  $C_s$  were excluded. The data was given by  $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$  for the neutral model taken at time points given by equation (6.1) and log-normal errors with standard deviations set as in equation (6.2).

Summary over 100 error realizations - IT data with double the measurement times			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	87	21	5
$P$	99	99	13
$f_{u,ELF}$	89	21(*)	8
$df_{tb}$	97	97	7

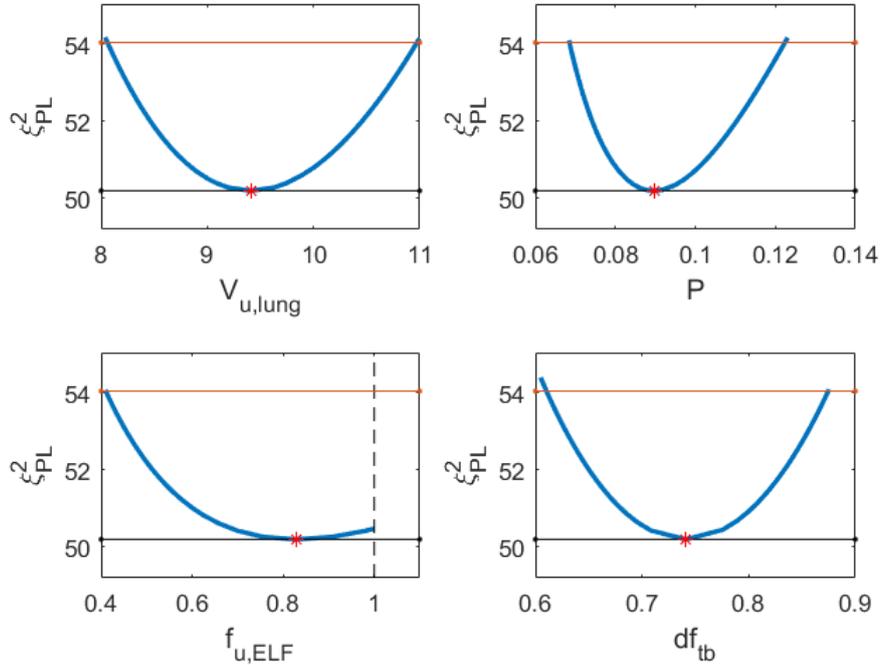
**Table C.4:** The PL calculations was run with settings as in Table 6.1. The IT data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.3) for both observables ( $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$ ). Total number of data points in each run was thus  $n = 32$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.



**Figure C.11:** Range of log of the observable trajectories calculated as in equation (3.12) along the PL of  $f_{u,ELF}$  in figure 6.2.

Summary over 100 error realizations - IT/IV with additional IT data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	100	100	6
$P$	100	100	5
$f_{u,ELF}$	98	98	2
$df_{tb}$	100	100	7

**Table C.5:** The PL calculations was run with settings as in Table 6.1. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.1) + at  $t = 1, 2$  and 3 seconds on the IT arm for both observables ( $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$ ). Total number of data points in each run was thus  $n = 38$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.



**Figure C.12:** One example of profile likelihoods for when additional measurements of  $C_{plasma}$  in the IT arm at 1,2 and 3 seconds is considered.

Summary over 1000 error realizations - IT/IV with double the amount of data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$V_{u,lung}$	987	76	8
$P$	978	100	5
$f_{u,ELF}$	924	30(*)	4
$df_{tb}$	992	100	3

**Table C.6:** The PL calculations was run with settings as in Table 6.1. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.3). Total number of data points in each run was thus  $n = 64$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.

Summary over 100 error realizations - IT/IV/DPI data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$D$	100	4	4
$C_s$	100	4	4
$V_{u,lung}$	92	90	9
$P$	100	100	6
$f_{u,ELF}$	93	39	3
$df_{tb}$	98	100	2

**Table C.7:** The PL calculations was run with settings as in Table 6.1. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2). The log-normal standard deviations for the DPI data was set to the same values as for IT. The measurement times was set as in equation (6.1) for both observables ( $C_{lung}(t_i)$  and  $C_{plasma}(t_i)$ ) on all administration arms. Total number of data points in each run was thus  $n = 48$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.

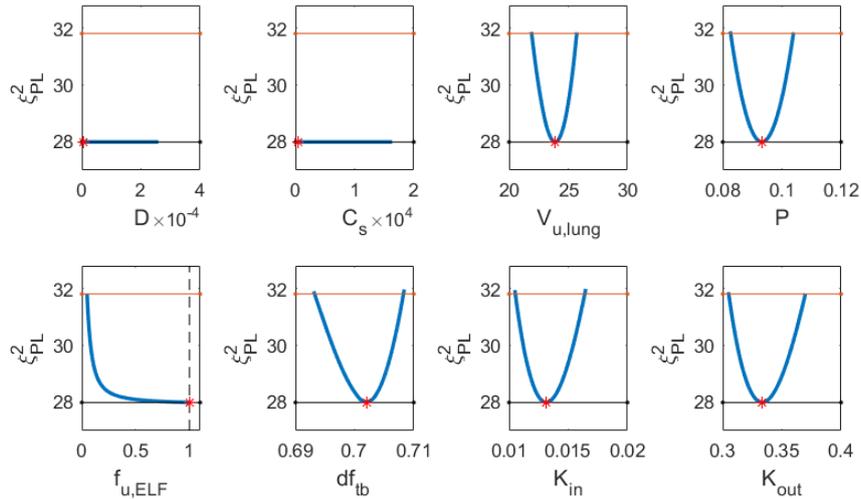
# D

## Basic molecules

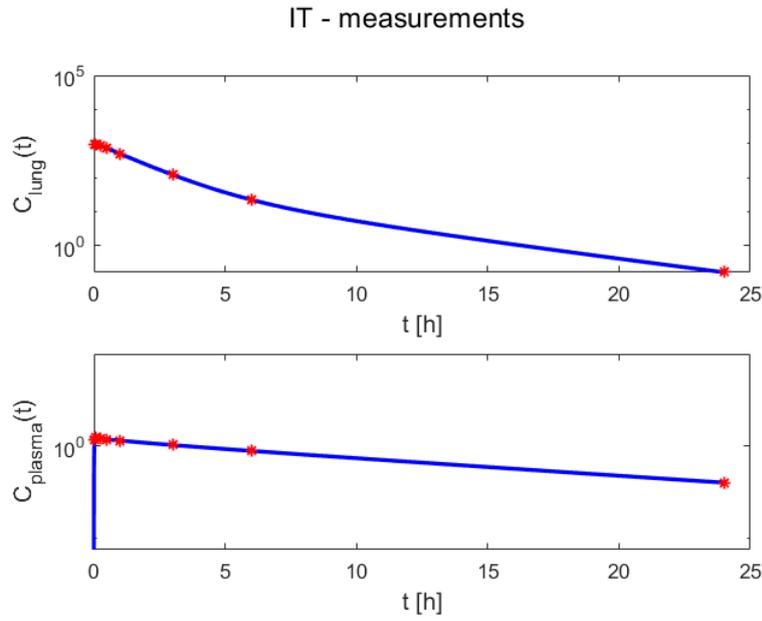
The analysis with low measurement error for the simplified case and additional Figures and Tables for section 6.2 results are given here.

### D.1 Simplified case - Low measurement error

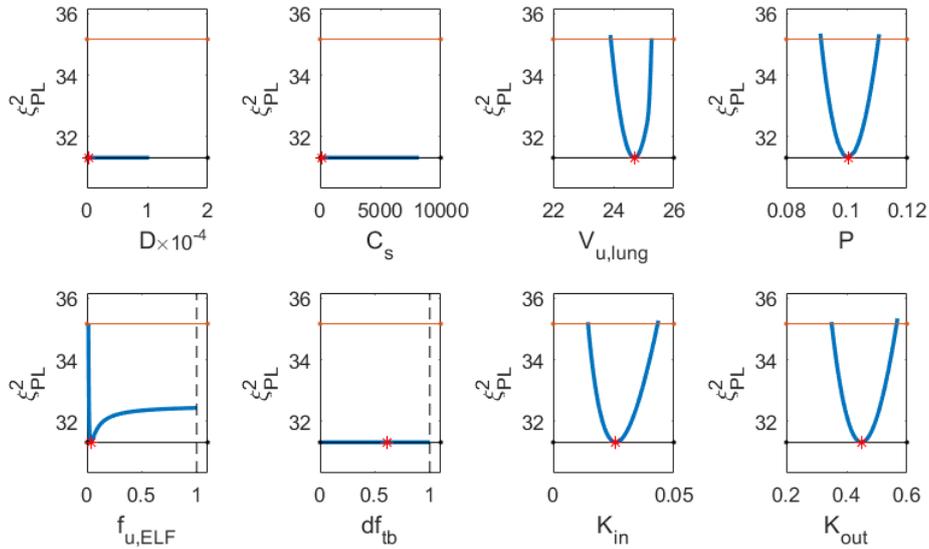
With precise measurements, i.e. data generated with measurement error set to 0.01 for all measurements, and only IT administration the profile likelihoods in Figure D.1 are obtained. Simulated data in this case is seen in Figure D.2. The results of PL calculation show that all parameters except  $D$  and  $C_s$  are identifiable, but the CI for  $f_{u,ELF}$  is large. Changing the administration route to IV leads to similar results, see Figure D.3 and corresponding data in figure D.4.



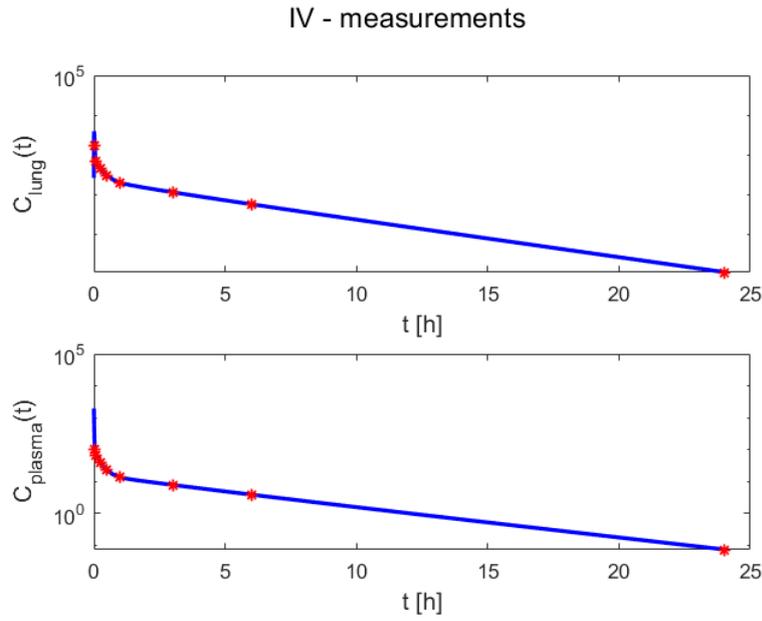
**Figure D.1:** Profile likelihoods calculated for the simplified case of the basic molecule model with IT administration and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).  $D$  and  $C_s$  are not identifiable with only IT administration.



**Figure D.2:** Simulated trajectories of the observables  $C_{plasma}$  and  $C_{lung}$  for the simplified case of the basic molecules model with IT administration. The red stars are simulated data taken with low measurement error ( $\sigma_{ki} = 0.01$ ) at times given by equation (6.1).

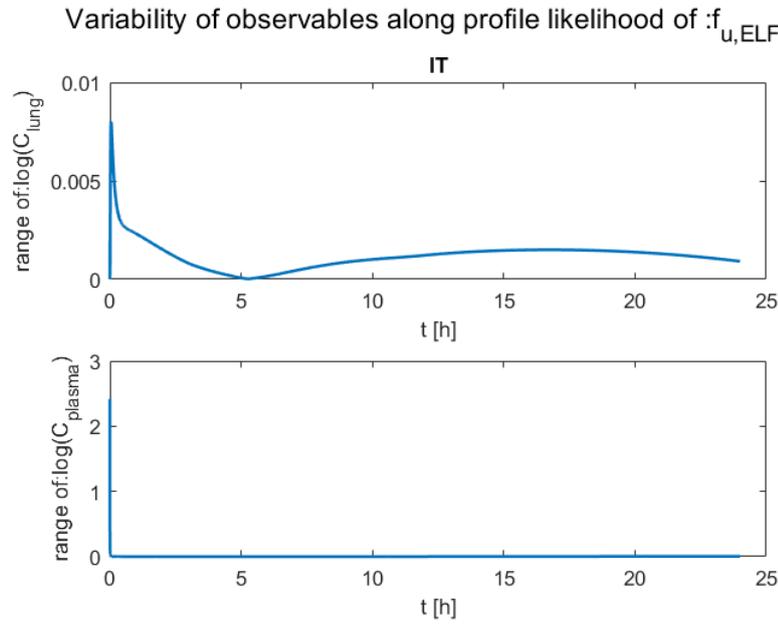


**Figure D.3:** Profile likelihoods calculated for the simplified case of the basic molecule model with IV administration and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).  $D$ ,  $C_s$  and  $df_{tb}$  are not identifiable with only IV administration.



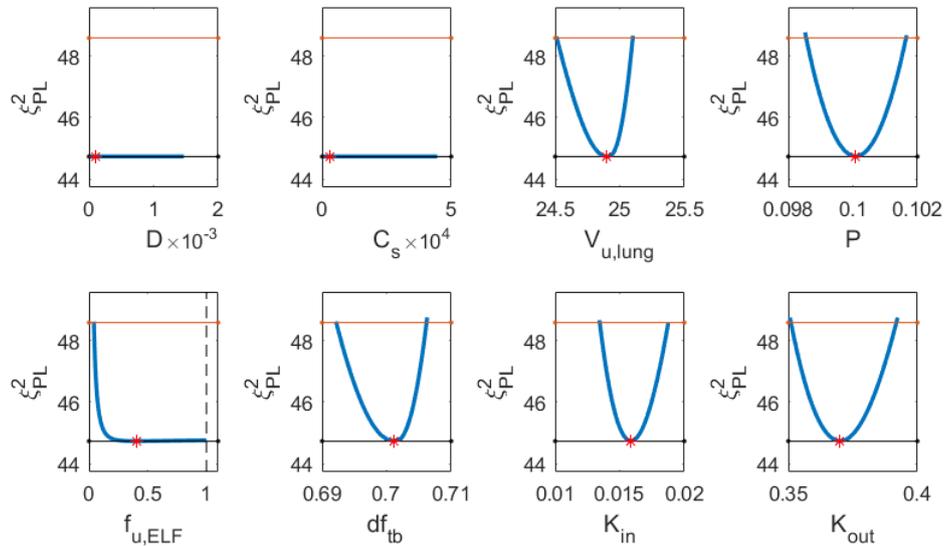
**Figure D.4:** Simulated trajectories of the observables  $C_{plasma}$  and  $C_{lung}$  for the simplified case of the basic molecules model with IV administration. The red stars are simulated data taken with low measurement error ( $\sigma_{ki} = 0.01$ ) at times given by equation (6.1).

The variability of trajectories along the PL of  $f_{u,ELF}$  indicate the need for additional measurement of  $C_{plasma}$  in the initial seconds after dosing, see figure D.5. As mentioned in the neutral molecule case however, this is not feasible in practice and will therefore not be considered again.

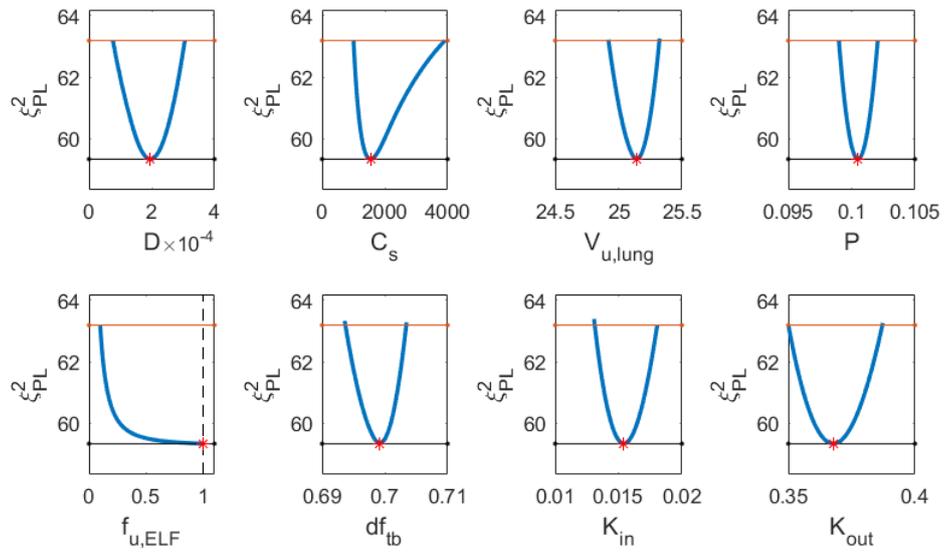


**Figure D.5:** Range of log of the observable trajectories calculated as in equation (3.12) with  $\theta^l$  along the PL of  $f_{u,ELF}$  in figure D.1.

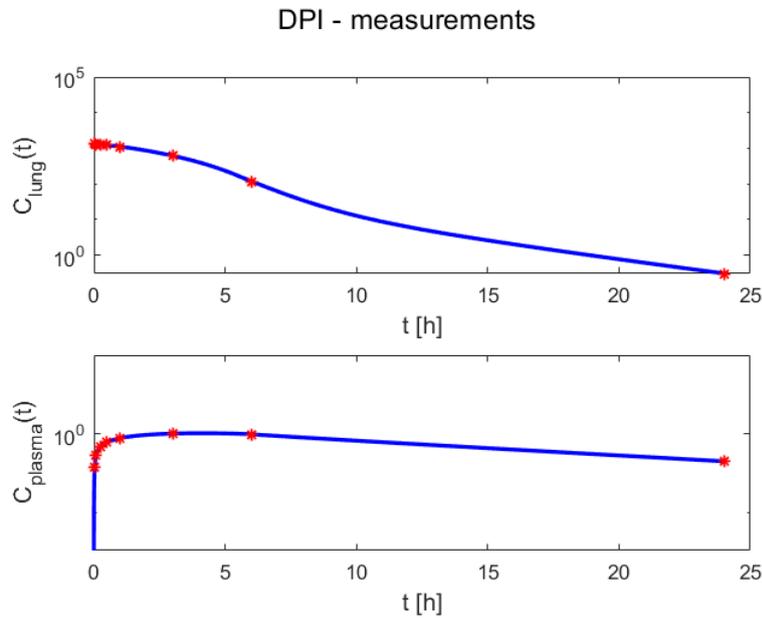
With both IT and IV data the results are similar to the case of only IT administration, see Figure D.6 and with the additional DPI data identifiability of  $D$  and  $C_s$  is obtained as well, see Figure D.7. The DPI data is shown in Figure D.8.



**Figure D.6:** Profile likelihoods calculated for the simplified case of the basic molecule model with IT and IV data and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).  $D$  and  $C_s$  are not identifiable with no DPI administration.



**Figure D.7:** Profile likelihoods calculated for the simplified case of the basic molecule model with IT, IV and DPI data and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).

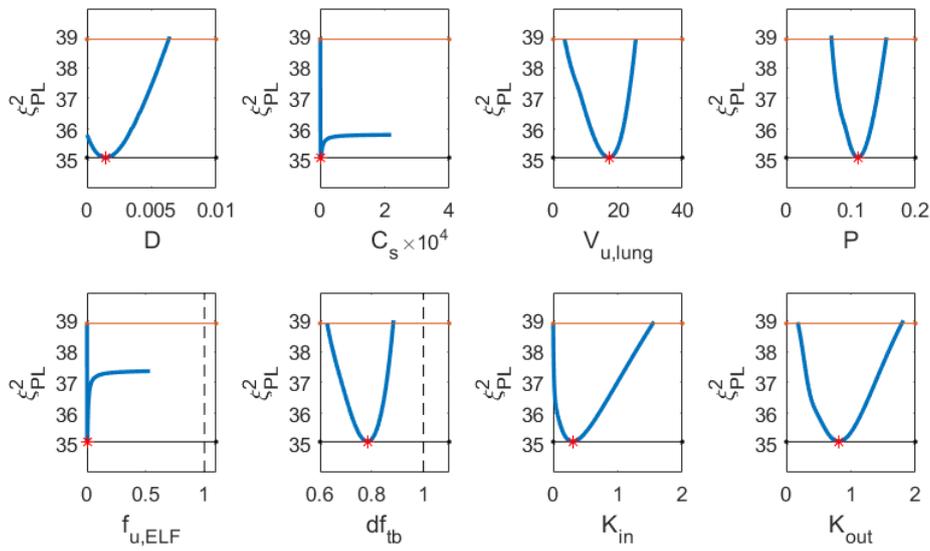


**Figure D.8:** Simulated trajectories of the observables  $C_{plasma}$  and  $C_{lung}$  for the simplified case of the basic molecules model with DPI administration. The red stars are simulated data taken with low measurement error ( $\sigma_{ki} = 0.01$ ) at times given by equation (6.1).

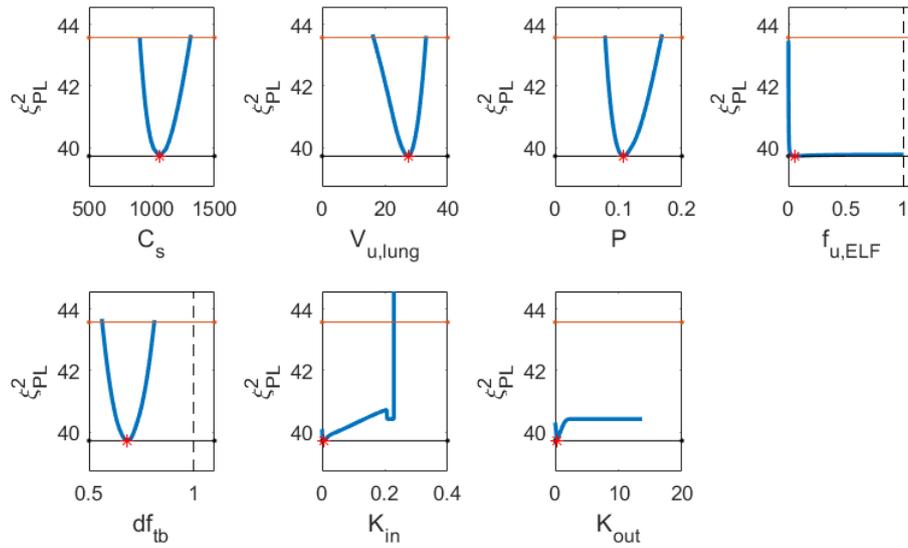
Thus the results with low measurement error for basic molecules are the same as for neutral molecules, i.e. only IT seems to be sufficient to obtain well-determined

estimates of all parameters except  $D$  and  $C_s$ . With IT, IV and DPI data the calculations show that  $D$  and  $C_s$  are also well-determined but with larger CIs compared to all other parameters except  $f_{u,ELF}$ . The uncertainty in the estimate of  $f_{u,ELF}$  is high regardless of administration route.

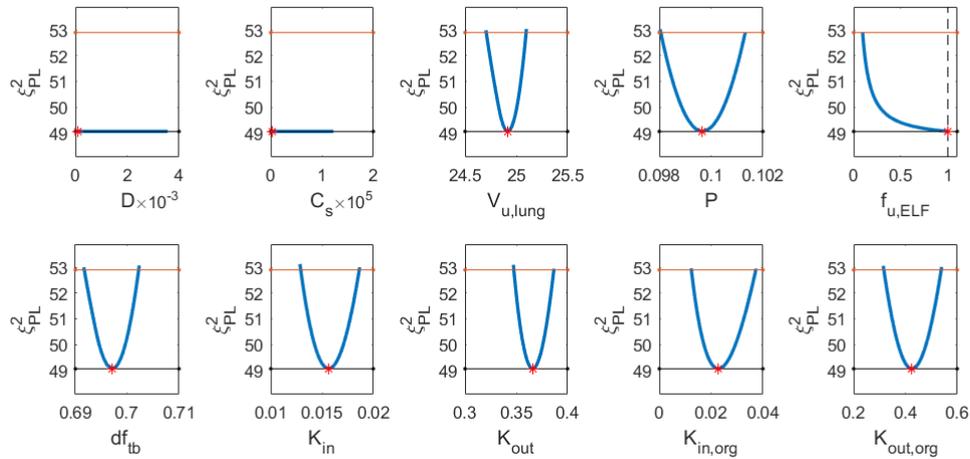
## D.2 Additional Figures and Tables - Reasonable measurement error



**Figure D.9:** Profile likelihoods calculated for the simplified case of the basic molecule model with IT, IV and DPI data and reasonable measurement error. The log-normal standard deviations were set as in equation (6.2) for all measurement times. Data was taken at time points given by equation (6.1). In this particular case both rate constants are identifiable.



**Figure D.10:** Profile likelihoods calculated for the simplified case of the basic molecule model with IT, IV and DPI data and reasonable measurement error. The log-normal standard deviations were set as in equation (6.2) for all measurement times. Data was taken at time points given by equation (6.1). Stokes-Einsteins equation was assumed to be valid for calculating  $D$ . In this particular case  $K_{out}$  is practically unidentifiable and  $K_{in}$  gives an example of optimization failure.



**Figure D.11:** Profile likelihoods calculated for the unique case of the basic molecule model with IT and IV data and low measurement error. The log-normal standard deviations were set to 0.01 for all measurements. Data was taken at time points given by equation (6.1).  $D$  and  $C_s$  are not identifiable unless DPI administration is used.

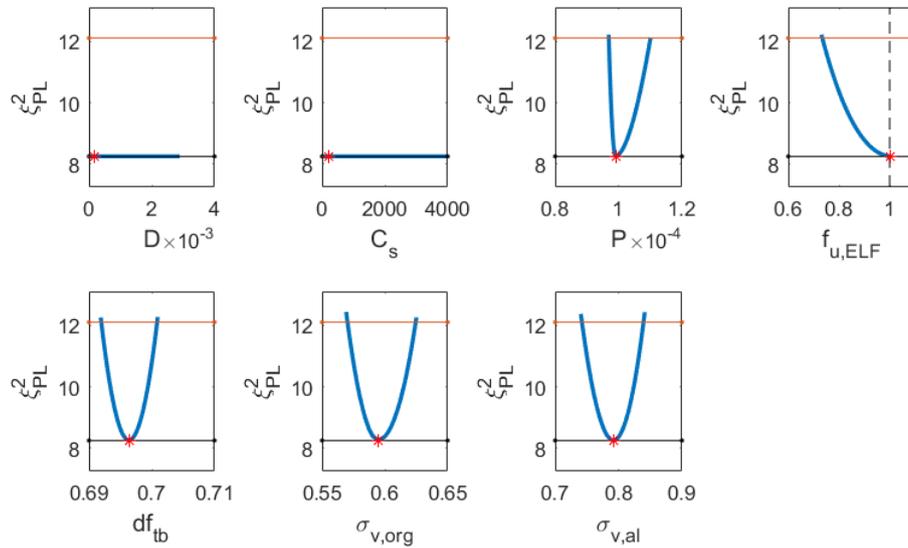
# E

## Large molecules

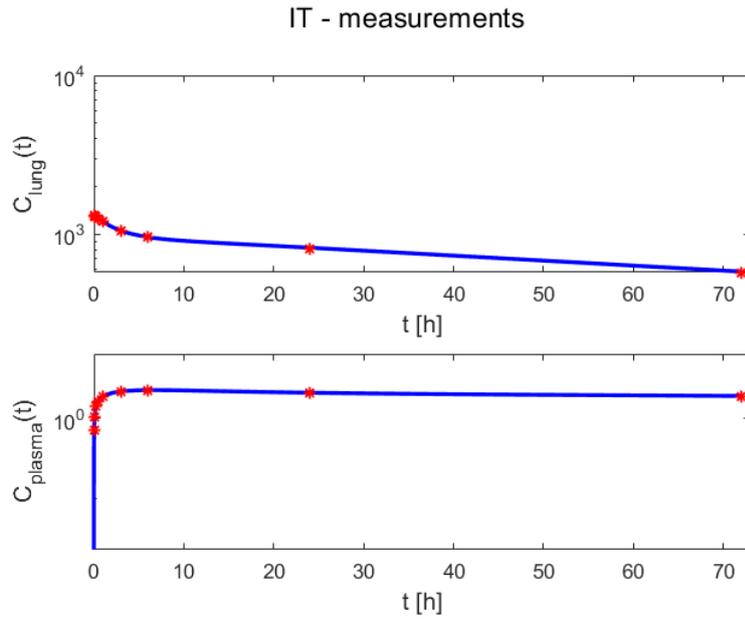
The analyses with low measurement error for both simplified and unique case as well as additional Figures and tables for section 6.3 are given here.

### E.1 Simplified case - Low measurement error

With a low measurement error ( $\sigma_{ki} = 0.01 \forall k, i$ ) and measurement times as in equation (6.4), the PL calculation for only IT gives PLs as in Figure E.1. Corresponding data is given in Figure E.2. All parameters except  $D$  and  $C_s$  are identifiable.

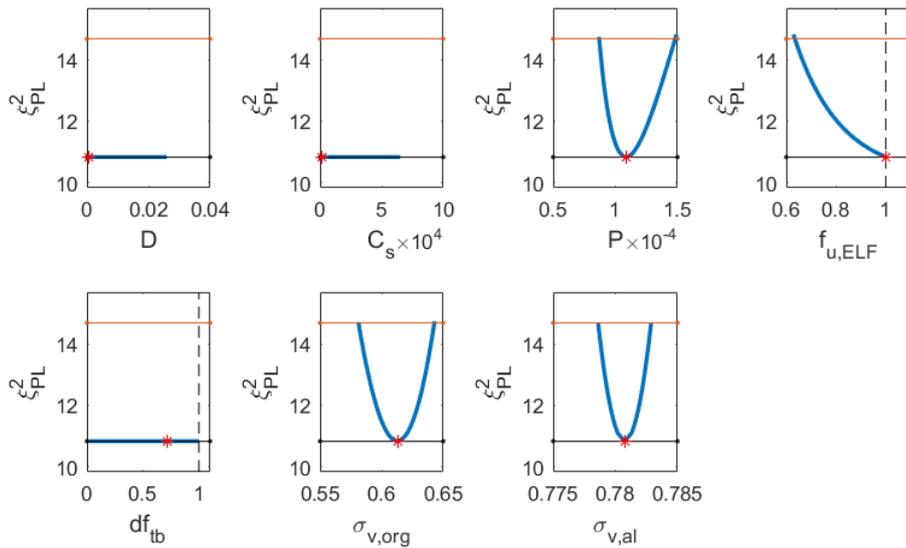


**Figure E.1:** PLs of the parameters in the simplified case of the large molecule model calculated with only IT data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

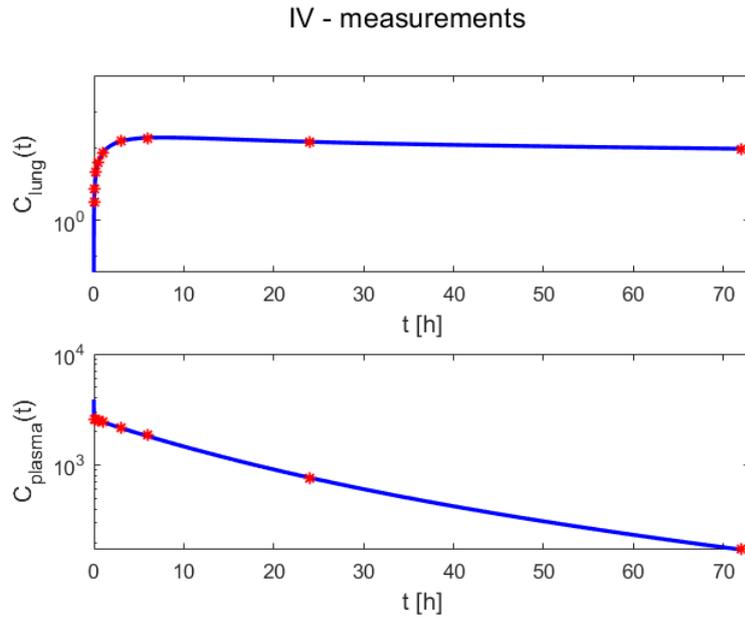


**Figure E.2:** Simulated trajectories of  $C_{lung}(t)$  and  $C_{plasma}(t)$  for the simplified case of the large molecule model with IT administration. The red stars are simulated data with log-normal error and standard deviation set to 0.01, taken at time points given by equation (6.4).

Similarly, only IV data gives identifiability of all parameters except  $D, C_s$  and  $df_{tb}$ , see Figure E.3 for the PLs and Figure E.4 for the corresponding data.

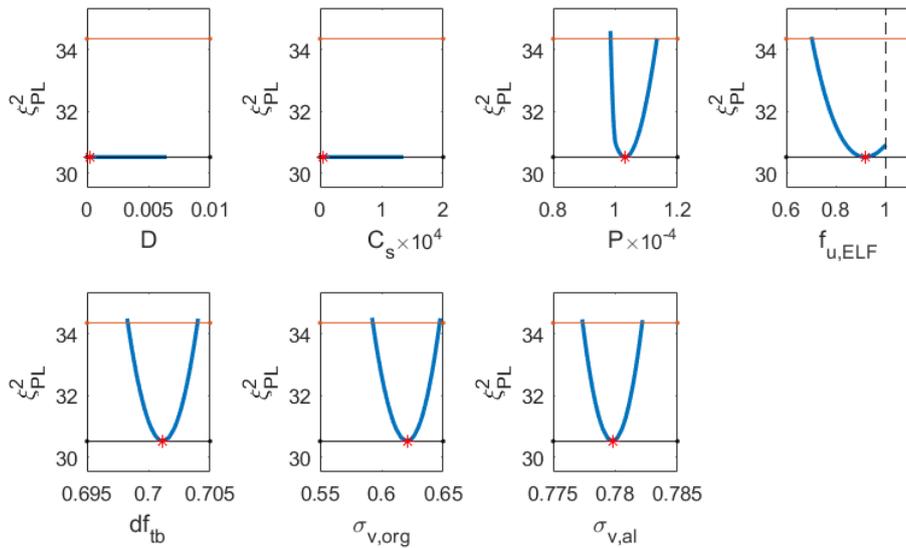


**Figure E.3:** PLs of the parameters in the simplified case of the large molecule model calculated with only IV data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.



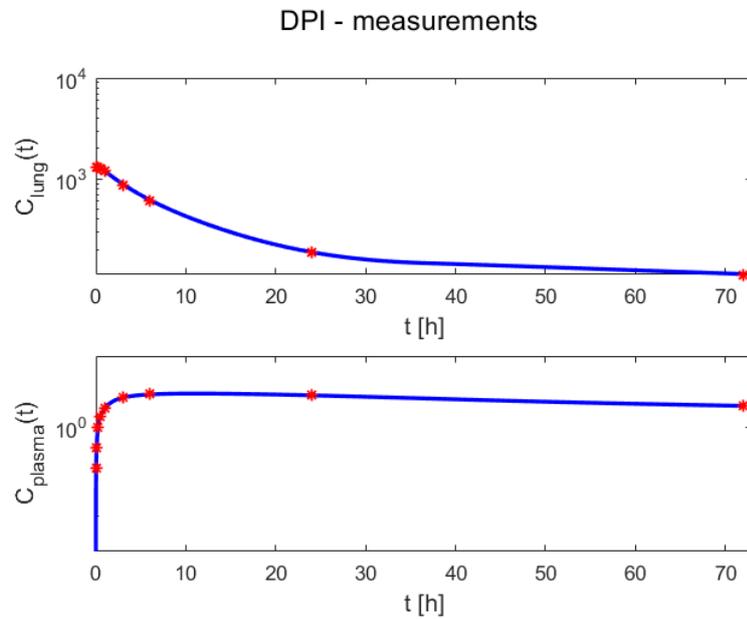
**Figure E.4:** Simulated trajectories of  $C_{lung}(t)$  and  $C_{plasma}(t)$  for the simplified case of the large molecule model with IV administration. The red stars are simulated data with log-normal error and standard deviation set to 0.01, taken at time points given by equation (6.4).

The combined IT and IV data should then obviously give identifiability of all parameters except  $D$  and  $C_s$  and this is also what is obtained, see figure E.5.



**Figure E.5:** PLs of the parameters in the simplified case of the large molecule model calculated with IT and IV data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

The addition of DPI data, seen in Figure E.6, gives identifiability of all parameters as seen in Figure 6.11 of the main text.

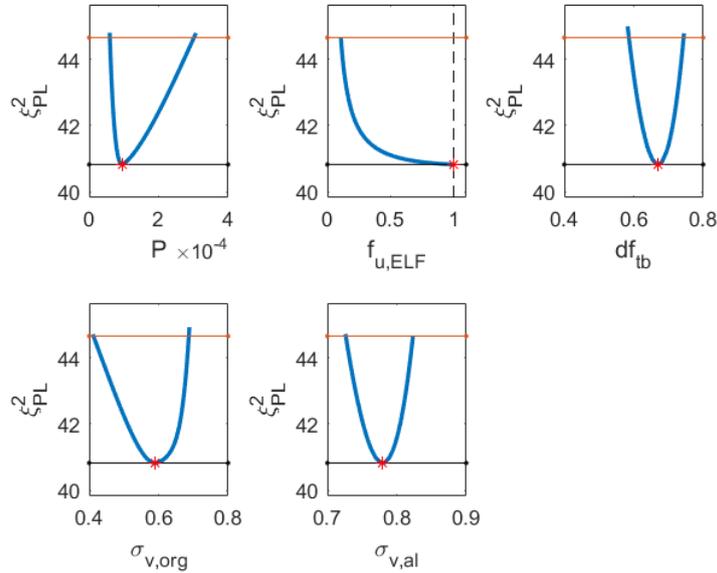


**Figure E.6:** Simulated trajectories of  $C_{lung}(t)$  and  $C_{plasma}(t)$  for the simplified case of the large molecule model with DPI administration. The red stars are simulated data with log-normal error and standard deviation set to 0.01, taken at time points given by equation (6.4).

## E.2 Additional Figures and Tables - Simplified case

Summary over 1000 error realizations - IT/IV data with additional measurements of IT			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$P$	992	94	6
$f_{u,ELF}$	999	82(*)	2
$df_{tb}$	998	100	5
$\sigma_{v,org}$	980	97	8
$\sigma_{v,al}$	997	100	6

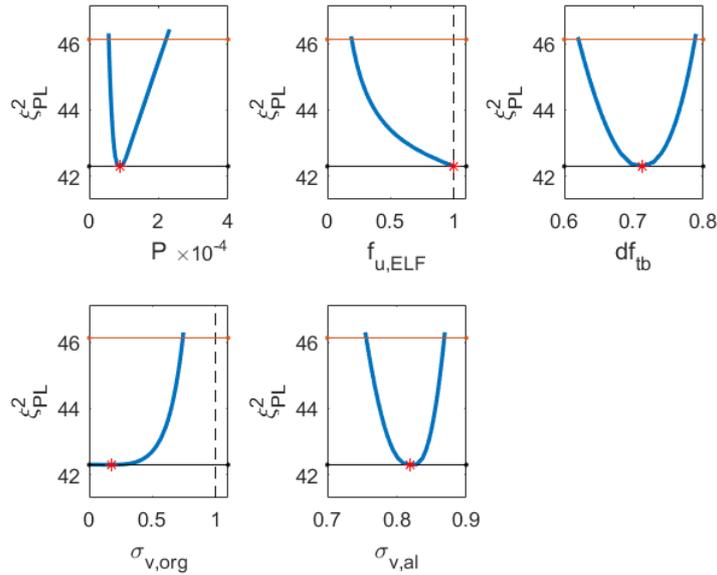
**Table E.1:** Summary of PL calculation of 1000 different error realizations of IT and IV data for the simplified case of the large molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the measurement times was set as in equation (6.4) + additional measurements of both  $C_{lung}(t)$  and  $C_{plasma}(t)$  at  $t = 200, 500$  and  $1000$ h on the IT arm. Total number of data points were thus  $n = 42$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations. (\*)The number of identifiable cases are higher in reality as the PL calculation was stopped prematurely (by reaching maximum number of steps) in some of the individual runs.



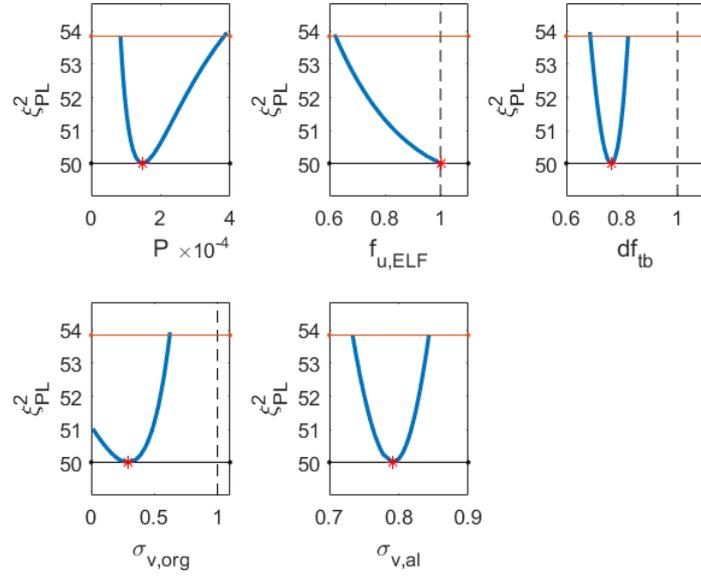
**Figure E.7:** PLs of the parameters in the simplified case of the large molecule model calculated with IT and IV data with additional measurements at late time points on the IT arm, all with reasonable measurement error. The IT and IV data was given by measurement times as in equation (6.4) +  $t = 200, 500$  and  $1000$  for IT, with log-normal standard deviations as in equation (6.2).

Reasonable measurement error - IT/IV data with additional measurements of IT			
Parameter	True value	Estimate	CI
$P$	0.0001	0.0000944	[0.0000596, 0.000301]
$f_{u,ELF}$	1	1	[0.107, 1]
$df_{tb}$	0.7	0.670	[0.586, 0.743]
$\sigma_{v,org}$	0.6	0.589	[0.412, 0.687]
$\sigma_{v,al}$	0.78	0.780	[0.727, 0.823]

**Table E.2:** Parameter estimates and confidence intervals (CIs) for the simplified case of the large molecule model with combined IT and IV data and reasonable measurement error. The measurement times were taken as in equation (6.4) +  $t = 200, 500$  and  $1000$  for IT. The CIs are calculated from the profile likelihoods in figure 6.14 using spline interpolation to obtain the points of passover. The parameters  $D$  and  $C_s$  were omitted as dissolution is not part of the model unless DPI administration is used



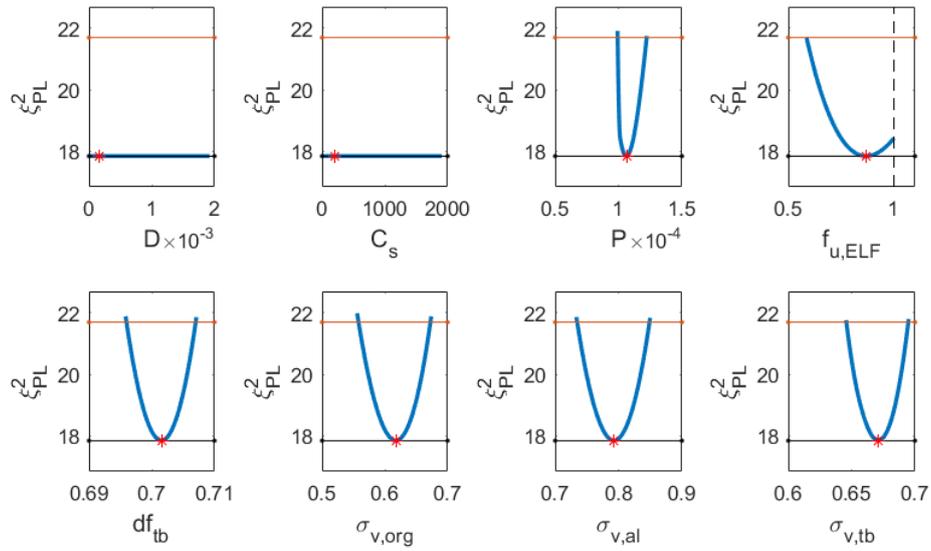
**Figure E.8:** PLs of the parameters in the simplified case of the large molecule model calculated with IT and IV data + one additional measurement of  $C_{ri,tot}$ , all with reasonable measurement error. The IT and IV data was given by measurement times as in equation (6.4) with log-normal standard deviations as in equation (6.2). The  $C_{ri,tot}$  - data was taken at  $t = 72h$  on the IT arm with log-normal standard deviation set to 0.3.



**Figure E.9:** PLs of the parameters in the simplified case of the large molecule model calculated with IT and IV data + one additional measurement of  $C_{ELF,avg}$ , all with reasonable measurement error. The IT and IV data was given by measurement times as in equation (6.4) with log-normal standard deviations as in equation (6.2). The  $C_{ELF,avg}$  - data was taken at  $t = 72\text{h}$  on the IT arm with log-normal standard deviation set to 0.3.

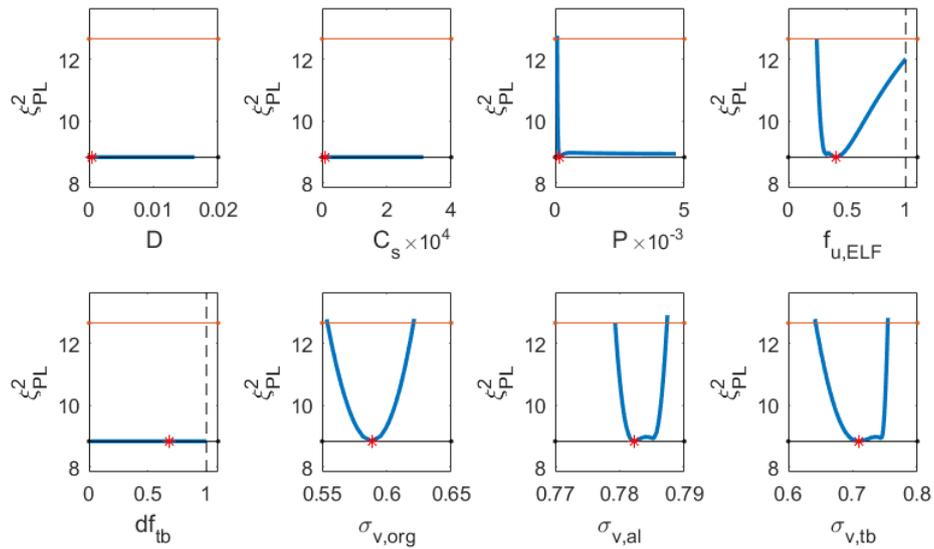
### E.3 Unique case - Low measurement error

With low measurement error and only IT data similar results as in the simplified case is obtained, with the only difference being the addition of  $\sigma_{v,tb}$  which is identifiable, see Figure E.10.



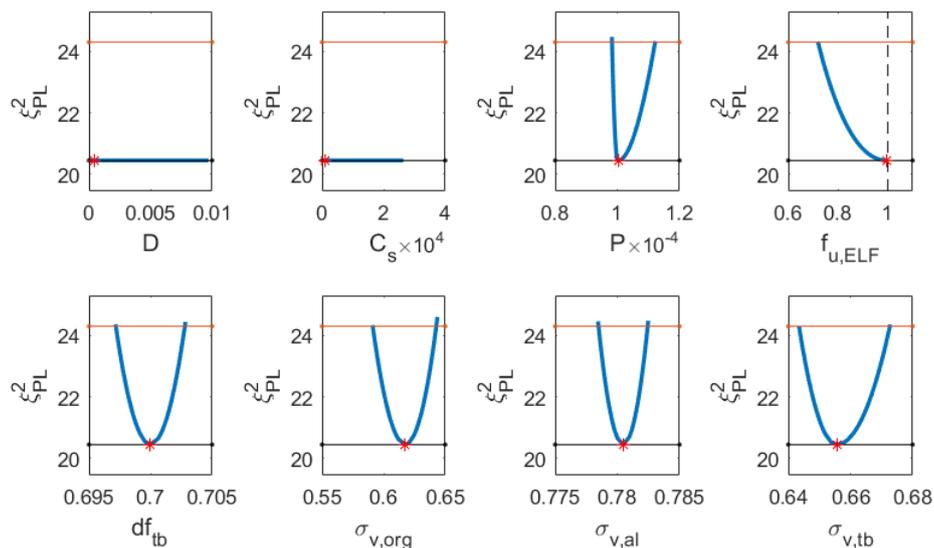
**Figure E.10:** PLs of the parameters in the unique case of the large molecule model calculated with only IT data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

However, with only IV administration it is seen that  $P$  is practically unidentifiable, see figure E.11.



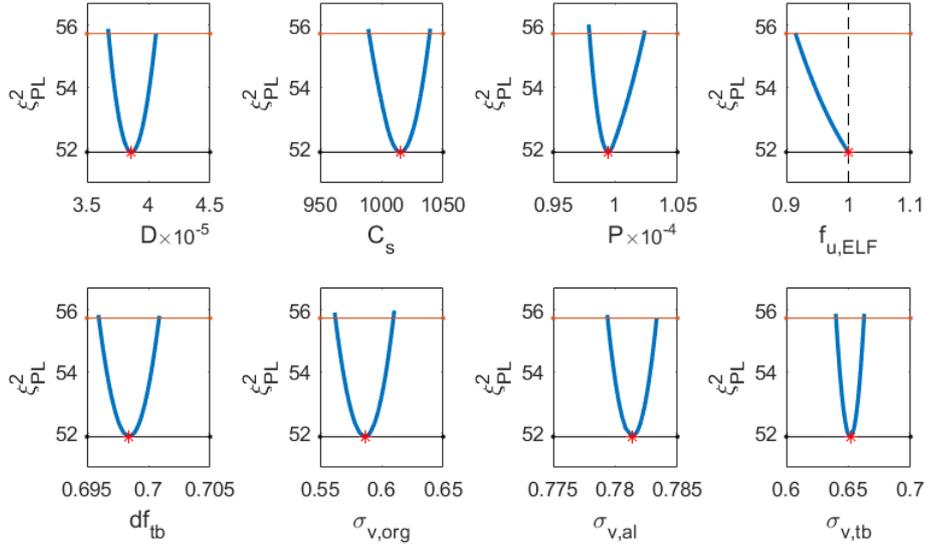
**Figure E.11:** PLs of the parameters in the unique case of the large molecule model calculated with only IV data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

With both IT and IV data the PLs as in figure E.12 were obtained. All parameters except  $D$  and  $C_s$  are identifiable.



**Figure E.12:** PLs of the parameters in the unique case of the large molecule model calculated with IT and IV data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

Lastly, the addition of DPI data gives the PLs as in figure E.13 and corresponding CIs in Table E.3. All parameters are identifiable.

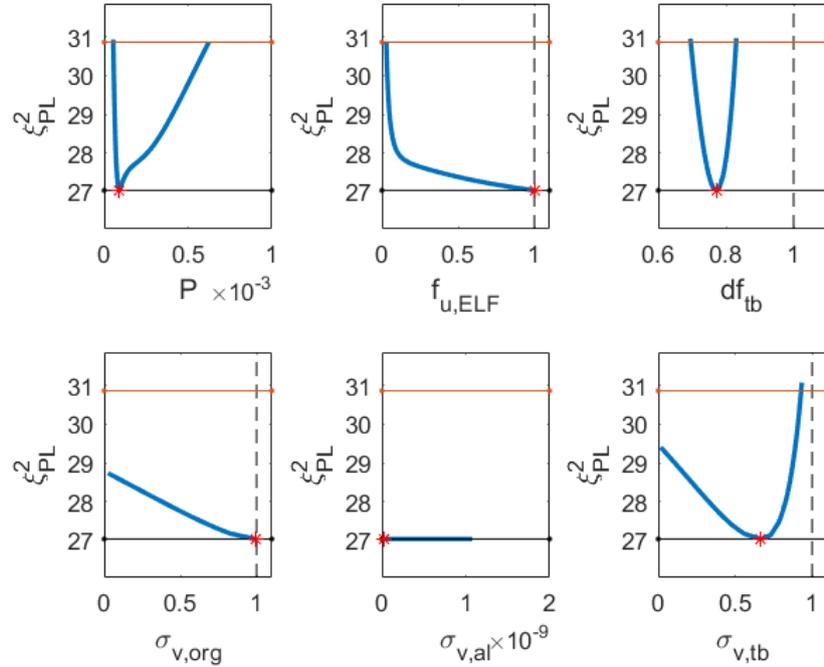


**Figure E.13:** PLs of the parameters in the unique case of the large molecule model calculated with IT, IV and DPI data and low measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set to 0.01 for all simulated measurements.

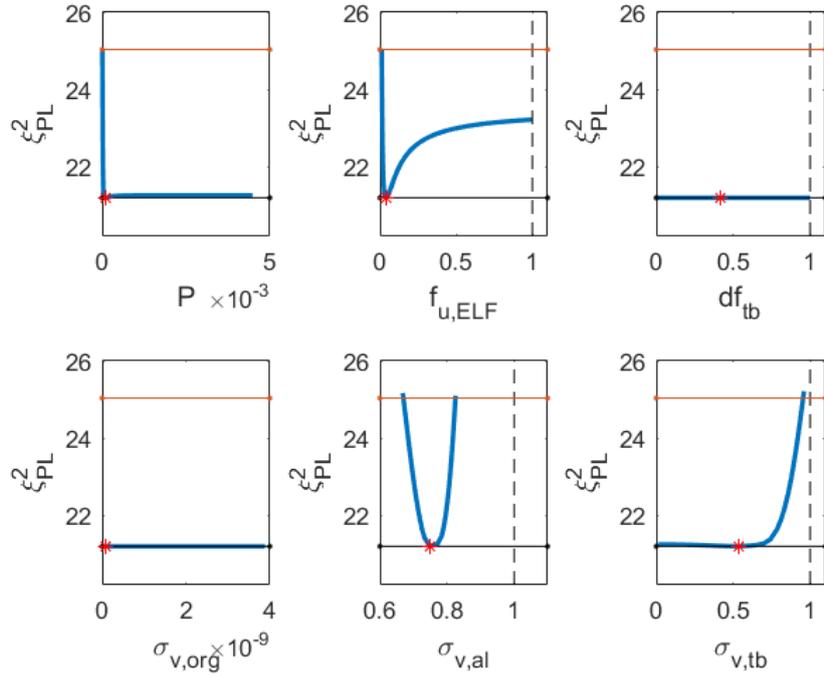
Low measurement error - IT/IV/DPI data			
Parameter	True value	Estimate	CI
$D$	0.0000394	0.0000386	[0.0000368, 0.0000406]
$C_s$	1000	1015	[990, 1039]
$P$	0.0001	0.0000994	[0.0000979, 0.0001023]
$f_{u,ELF}$	1	1	[0.915, 1]
$df_{tb}$	0.7	0.698	[0.696, 0.701]
$\sigma_{v,org}$	0.6	0.586	[0.562, 0.610]
$\sigma_{v,al}$	0.78	0.781	[0.779, 0.783]
$\sigma_{v,tb}$	0.66	0.651	[0.640, 0.662]

**Table E.3:** Parameter estimates and confidence intervals (CIs) for the unique case of the large molecule model with IT, IV and DPI data and low measurement error ( $\sigma_{ki} = 0.01$ ). The CIs are calculated from the profile likelihoods in figure E.13 using spline interpolation to obtain the points of passover.

## E.4 Additional Figures and Tables - Unique case



**Figure E.14:** PLs of the parameters in the unique case of the large molecule model calculated with only IT data and reasonable measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set as in equation (6.2).



**Figure E.15:** PLs of the parameters in the unique case of the large molecule model calculated with only IV data and reasonable measurement error. The data was given by measurement times as in equation (6.4). The log-normal standard deviations were set as in equation (6.2).

Summary over 100 error realizations - IT/IV/DPI data			
Parameter	Successful PL calculations	Identifiable cases [%]	False positives [%]
$D$	100	64	2
$C_s$	100	100	3
$P$	96	100	6
$f_{u,ELF}$	100	100	2
$df_{tb}$	100	100	1
$\sigma_{v,org}$	98	19	1
$\sigma_{v,al}$	100	100	7
$\sigma_{v,tb}$	100	75	5

**Table E.4:** Summary of PL calculation of 100 different error realizations of IT, IV and DPI data for the unique case of the large molecule model. The IT and IV data was generated with log-normal error distribution with standard deviations set as in equation (6.2), the DPI data was generated with log-normal errors and standard deviations same as for IT. The measurement times was set as in equation (6.4). Total number of data points were thus  $n = 54$ . The constrained optimization was considered to have failed if the optimizer returned an error or a non-global optimum and these cases are retracted to give the number of successful PL calculations.