A Workflow To Detect Non-Identifiability In Parameter Estimation Using SimBiology

Florian Augustin, Sietse Braakman, Ricardo Paxson
Florian.Augustin@mathworks.com
MathWorks Natick, MA

Introduction

A central challenge working with PKPD, PBPK or QSP models is the estimation of model parameters from data. MathWorks' SimBiology toolbox provides functionalities to perform parameter estimation, however, these require the fit task to be *structural and practical identifiable*. In this poster, we discuss methods to detect and overcome identifiability problems. **Structural non-identifiability** is caused by locally **insensitive parameters or aliasing parameters** which introduce ambiguity in the estimation process. Aliasing can be observed when the same model response can be obtained by multiple combinations of parameter values. Parameter aliasing is often caused by missing model structure. We speak of **practical non-identifiability**, if a model's **responses cannot be discerned at measurement times of the data**. This problem is often caused by sparse or missing data at times of significant model dynamics.

Aim: identify local structural and practical non-identifiability and provide **guidance for data collection** to improve the quality of parameter estimations.

Methodology

Summary

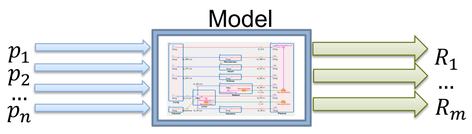
Metrics for *local parameter sensitivities*, the **sensitivity score**, and *local parameter aliasing*, the **aliasing score**, are computed for a model's responses. These metrics are used in two ways: to **inform the parameter estimation** and to **inform experimental design for data collection** to improve parameter identifiability. To aide parameter estimation, we use the metrics to partition model parameters into locally estimable and non-estimable sets. The non-estimable parameters are fixed to pre-determined values, while the estimable parameters are fitted using data. To inform data collection, we use the metrics on restricted times, e.g. data measurement times, to determine the most informative measurement times and targets.

Tools

We use the **SimBiology Modeling Desktop** for model creation. A **`SimFunction`** is used to simulate the model and compute model responses' sensitivities with respect to model parameters. **MATLAB** and `SimData`'s **`resample`** method are used to compute the local sensitivity and aliasing scores. The function **`sbiofit`** is used to perform parameter estimation. Finally, SimBiology's function **`sbioparameterci`** is used to compute profile-likelihood confidence intervals to assess the quality of the fitted parameters.

Notation

Model parameters are denoted by $(p_1, \dots, p_n)$ and have associated values $(v_1, \dots, v_n)$. Examples for model parameters in a two-compartment PKPD model are clearance, flow rate or compartment capacities. The model responses are denoted by $(R_1, \dots, R_m)$, e.g. drug concentrations in central and peripheral compartments.



Sensitivity Score

Compute sensitivities $\sigma_{i,k}(t) := \partial R_k(t)/\partial p_i$ of response $R_k(t)$ with respect to parameter p_i, for $i = 1, \dots, n$ and $k = 1, \dots, m$. Define the **response-wise local sensitivity score** $S_{i,k} := \max_{t \in T} |\sigma_{i,k}(t)|$ and the **local Sensitivity Score** $S_i := \max\{S_{i,k} : k = 1, \dots, m\}$. Here, T is a subset of the simulation time. Typical choices are $T = [t_0, t_{end}]$ or $T = \{\text{measurement times}\}$.

Aliasing Score

Normalize sensitivities by the sensitivity score, $\overline{\sigma_{i,k}}(t) := \sigma_{i,k}(t)/S_{i,k}$, if $S_{i,k} > 0$, and define the **response-wise local aliasing metric between parameters p_i and p_j** as

$$\alpha_{i,j}^k := \max_{t \in T} \left| |\overline{\sigma_{i,k}}(t)| - |\overline{\sigma_{j,k}}(t)| \right|.$$

The **local aliasing metric between parameters p_i and p_j** is defined as $\alpha_{i,j} := \min\{\alpha_{i,j}^k : k = 1, \dots, m\}$. Finally, the (response-wise) **local Aliasing Score** is defined as

$$A_{i,j}^k := 100 \cdot max\{1 - \alpha_{i,j}^k, 0\} \quad \text{and} \quad \mathrm{A}_{\mathrm{i,j}} := 100 \cdot max\{1 - \alpha_{i,j}, 0\},$$

respectively. Again, T is a subset of the simulation time
Typical choices are $T = [t_0, t_{end}]$ or $T = \{\text{measurement times}\}$.

Interpretation of Aliasing Scores

We motivate the aliasing score by means of a two-compartment PKPD model. Fig. 1 shows normalized sensitivities as well as aliasing scores for a subcutaneous dose in the peripheral compartment.

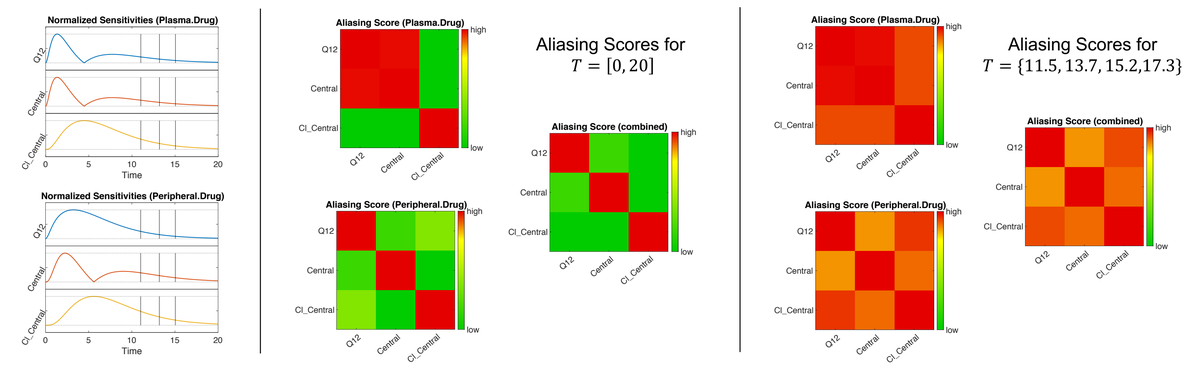
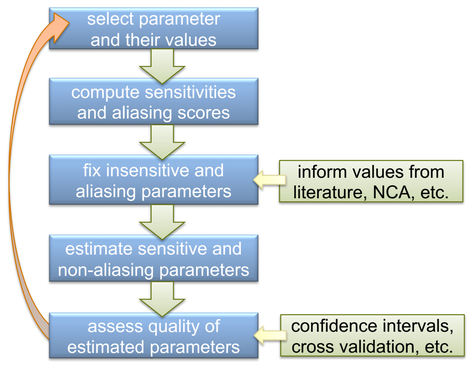


Figure 1: Normalized sensitivities (left) for the drug concentration in the central compartment (top row) and the peripheral compartment (bottom row). Vertical dotted lines indicate measurement times of available data. The middle plots show aliasing scores for $T = [0, 20]$. The right plots show aliasing scores for the measurement times $T = \{11.5, 13.7, 15.2, 17.3\}$.

Workflow Using Sensitivity and Aliasing Scores

We use the Sensitivity and Aliasing Scores at measurement times to find estimable parameters (high sensitivity and low aliasing score) and fix all remaining parameters to values from literature, NCA analyses, etc. Below we address the issue of not being able to estimate all parameters.



1) Start with initial values for parameters $(p_1, \dots, p_n) = (v_1, \dots, v_n)$.
2) Use data based **Sensitivity** and **Aliasing Scores** to fix parameters, $F = \{p_1, \dots, p_k\}$, and select estimable parameters, $V = \{p_{k+1}, \dots, p_n\}$.
3) Use **`sbiofit`** to estimate V to get new values $v_1, \dots, v_k, v_{k+1}^{est}, \dots, v_n^{est}$.
4) Assess quality of fit, e.g. using confidence intervals, or cross validation, and terminate or repeat Step 2 starting from new parameter values.

Global parameter exploration

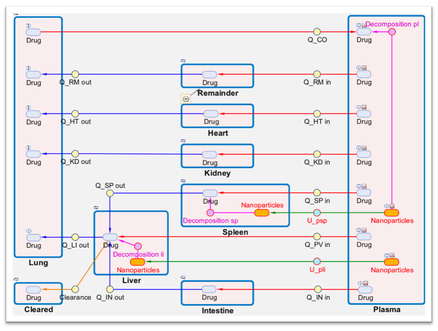
The above workflow depends on the initial guess for parameter estimates. For a more global parameter exploration, we start the above workflow from multiple initial values in Step 1. This exploration can be parallelized to minimize computational time. This approach requires specifying bounds for all parameters.

Informed data collection

The Aliasing Score can be used to inform data collection that maximizes the information gain for the parameter estimation. This can be achieved manually by shifting the scan window T and visually inspecting the respective Aliasing Score plots. Alternatively, the optimal next n measurements can be found by minimizing the maximal aliasing score over $T \in \{t_1, \dots, t_n\}$. This optimization approach will be the subject of future work.

Mechanistic Nanoparticle PBPK model

In [1], Dong et al. introduce a PBPK model describing nanocrystal and drug concentrations across several tissues and compartments. We implemented the PBPK model using the SimBiology Modeling Desktop and analyzed the model structure using the Aliasing Score for all parameters estimated from data as reported in [1]: partition coefficients, nanocrystal uptake and release rates as well as clearance. No insensitivities were detected. We use synthetic data and highlight how the Aliasing Score can be used to discover most valuable measurement times.



Data For Parameter Estimation

We performed a global parameter exploration using the workflow described above. Here we showcase one scenario to highlight how Aliasing Scores are used to explore parameter non-identifiability as well as inform experimental design for making new measurements. Below, two parameter estimations with synthetic data are performed. The two data sets used contain measurements *covering the whole time from 0 to 20 hours* (dataset 1), and only *measurements at times not later than 5 hours* (dataset 2), respectively.

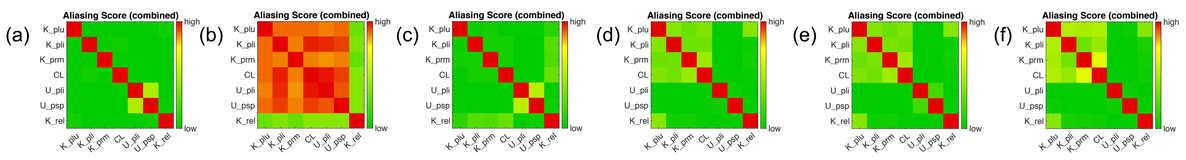


Figure 2: Aliasing Scores around initial estimates for $T = [0, 5]$ (a), $T = [5, 20]$ (b), as well as measurement times $T = \{0.05, 0.11, 0.29, 1.2, 2.1, 4.0, 5.25, 10.0, 15.3, 20.0\}$ (dataset 1) in plot (c). Plot (d) shows the Aliasing Score for the fit with full dataset 1 around the initial estimates. Plots (e,f) show Aliasing Scores for dataset 1 and dataset 2, $T = \{0.05, 0.11, 0.29, 1.2, 2.1, 4.0\}$, around the respective estimated values.

Identifiability Analysis Using Aliasing Scores

Fig. 2(a-c) show the combined (Plasma and Tissue) Aliasing Scores for $T = [0, 5]$, $T = [5, 20]$, and measurement times of dataset 1, $T = \{0.05, 0.11, 0.29, 1.2, 2.1, 4.0, 5.25, 10.0, 15.3, 20.0\}$, around initial estimates. We see that, around the initial parameter values, early measurements are paramount over later measurements. However, sensitivity plots (not shown) indicate significant dynamics at later times, which justifies the inclusion of measurements at times past 5 hours in dataset 1.

Fit Results

Both fits yield similar estimated values (Fig. 3), however, the Aliasing Scores shown in Fig. 2(d-f) indicate an identifiability problem when omitting measurements at later times. This is confirmed in the profile likelihood confidence intervals in Fig. 3. While the confidence intervals could be computed for all parameters for dataset 1, there are estimation problems for the reduced dataset 2.

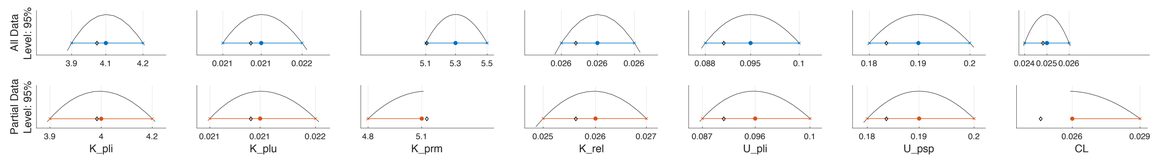


Figure 3: Profile likelihood confidence intervals for the parameter estimation with full data (top row) and reduced data (bottom row). The profile likelihood confidence interval for the reduce data are indicated as not estimable (red).

Conclusion

The presented Sensitivity and Aliasing Scores are a helpful tool for detecting parameter non-identifiability. Additionally, the scores can be used to inform future experiments about measurement times and targets to maximize the information gain for fitting model parameters to data.

[1] Dong D, Wang X, Wang H, Zhang X, Wang X, Wu B; International Journal of Nanomedicine 2015;10: 2521–2535

MathWorks®
Accelerating the pace of engineering and science