

Development of a population PK model using MONOLIX - Case Study II Multiple oral dose

September 9, 2019

Contents

1	Introduction	2
2	Data file and data exploration	2
3	Development of the PK model	4
3.1	Structural model: one-compartment	4
3.1.1	Initialization	6
3.1.2	Monolix tasks	7
3.1.3	Diagnostic plots	10
3.2	Structural model: two-compartments	13
3.3	Statistical model	16
4	Summary of results	18

1 Introduction

The aim of this tutorial is to develop a population PK model for a multiple dosing trial study using MonolixSuite applications of a hypothetical test drug used in the treatment of diabetes. The study was a double blind, multiple dose oral trial, in which 100 patients received 500 mg three times in a day for 5 days. The data set contains concentration measurements and the dose regimens from blood samples collected at regular time intervals.

The workflow consists of the preparation of a data set, choosing a PK model, estimating parameters and analyzing goodness of fits, building a statistical model and generating results. In this case study it takes 2 minutes in the Monolix interface to perform the whole workflow - from loading the data set to displaying plots and results - and get an output ready for the basic analysis.

The following MonolixSuite applications are used:

- [Datxplorer](#) to visualize and explore the data to give the first insights into the structural model and covariates relationships.
- [Monolix](#) to develop models, estimate parameters and generate statistical tests and diagnostic plots.
- [Sycomore](#) to manage the modeling workflow and compare models.

2 Data file and data exploration

Monolix accepts only one header line and data in the long format: each variable is in a separate column and each line corresponds to one individual and one time point. The header, the first line of the file, contains in each column the names of the variables. The choice of the columns names is free.

Figure 1 shows a sample of a csv file compatible with all applications of the MonolixSuite and Table 1 contains a complete description of all variables of the data set under study. All column types available in the MonolixSuite are described in the [data set documentation](#).

A	B	C	D	E	F	G
CID	TIME	CONC	AMT	ADDL	II	MDV
1	0	0	500	14	8	1
1	0.5	4.5799	0	0	0	0
1	1	8.5717	0	0	0	0
1	3	10.979	0	0	0	0
1	4	11.347	0	0	0	0
1	6	8.7705	0	0	0	0
1	8	7.0155	0	0	0	0
1	8.5	10.889	0	0	0	0
1	16	9.7015	0	0	0	0
1	16.5	14.115	0	0	0	0
1	24	10.809	0	0	0	0
1	24.5	14.217	0	0	0	0
1	32	11.412	0	0	0	0
1	32.5	16.441	0	0	0	0
1	40	11.038	0	0	0	0
1	40.5	16.612	0	0	0	0
1	48	10.985	0	0	0	0

Figure 1: A sample of a csv file compatible with all MonolixSuite applications.

Column name	Description	Requirement
CID	identifier of the individual tagged as ID	mandatory
TIME	time of a dose administration or observation record (blood sampling time) tagged as TIME	mandatory
CONC	observation data (plasma concentration) tagged as OBSERVATION	mandatory
AMT	dose amount at dosing time (zero for observation records) tagged as AMOUNT	mandatory
ADDL	additional identical doses given (ex. if dose administered for 5 days, 3 times per day, ADDL will be 14) tagged as ADDITIONAL DOSES	additional, if required
II	period between doses (ex. dosing every 8 hours) tagged as INTERDOSE INTERVAL	additional, if required
MDV	identifier to ignore an observation information tagged as IGNORED OBSERVATION	optional

Table 1: Description of the data set variables.

Data exploration starts with opening Datxplore, choosing a "New project" and loading the csv file. Because the columns names are free, the user must indicate to the application their meanings. In this example, Datxplore recognizes most of the columns automatically, but the remaining ones have to be tagged. Then the data set is accepted by clicking the button ACCEPT, see Figure 2 for a Datxplore interface screenshot.

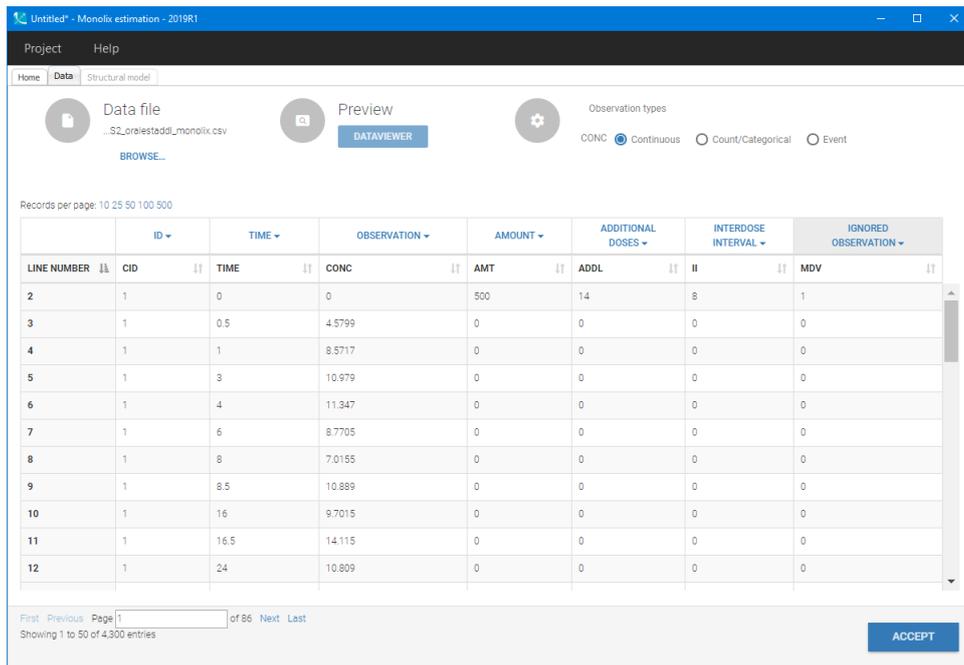


Figure 2: Preparation of the data set in Datxplore: columns are tagged according to the variable types and the data is accepted by clicking the button ACCEPT.

Datxplore allows for a fast and easy preview of the data, which helps in the identification of a first structural model. Figure 3 presents semi-log plot of the time evolution of the concentrations. Visually, the elimination slope (zoomed on the right) follows a linear relation, which corresponds to a one compartment model. However, further analysis using Monolix diagnostic plots and statistical tools, will show that a two compartment model is more accurate.

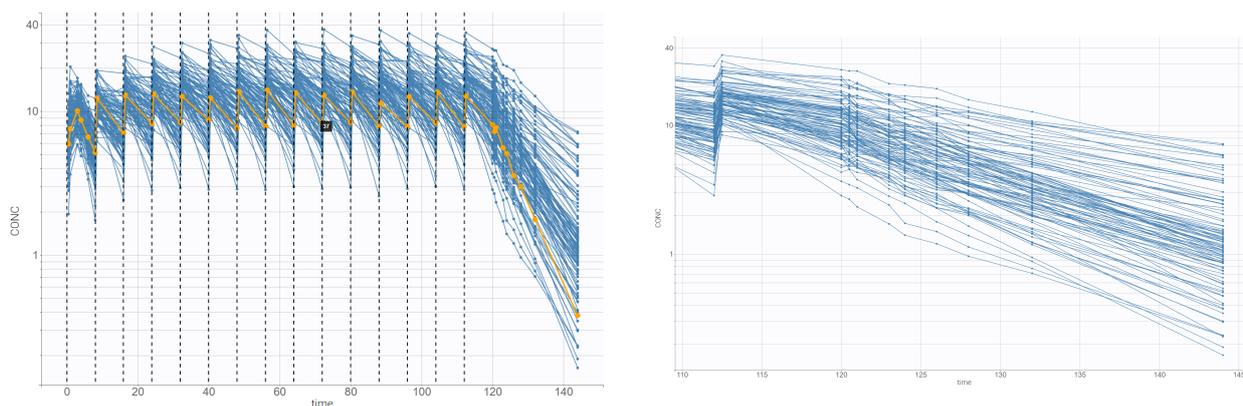


Figure 3: Datxplore viewer: time evolution of the drug concentration for the whole time of the treatment (on the left) with dosing times (dashed lines) and zoomed view on the elimination phase (on the right).

The data set visualization can be saved as a new datxplore project, and the plots can be saved as images. The data set information is then exported directly to Monolix from the menu item **Export > Export to Monolix**. The Monolix application opens with all required data information already setup.

3 Development of the PK model

This section describes how to build a PK model for this data, how to run different tasks to estimate parameters, analyze the model and interpret the results.

3.1 Structural model: one-compartment

In Monolix, the structural model is defined using the [Mlxtran language](#) as a separate text file. In this case study the model is loaded from the Monolix library, see Figure 4. Many typical models are available by clicking **LOAD FROM LIBRARY** in the "Structural model" tab: PK, PD, PKPD, PK Double Absorption, TMDD, TTE and Count.

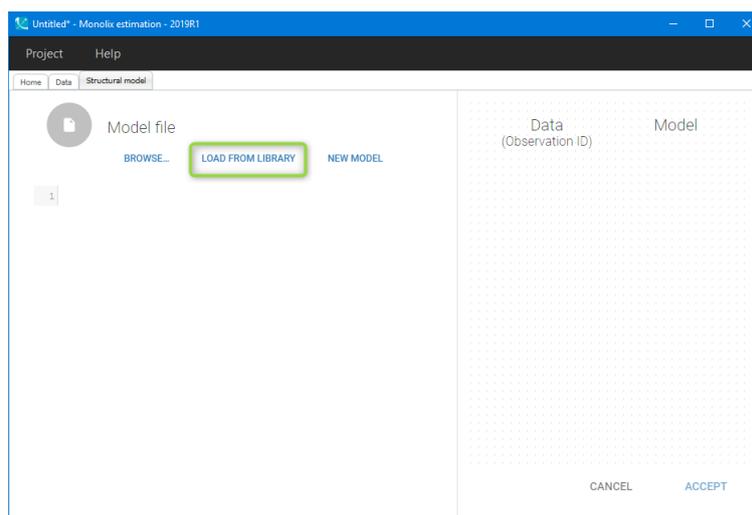


Figure 4: Options in Monolix to define a structural model. **BROWSE** - load a txt file with model written in the Mlxtran language; **LOAD FROM LIBRARY** - load a model from the Monolix library; **NEW MODEL** - write a new, customized model using Mlxtran language (MlxEditor opens automatically).

PK library is selected and the top panel allows to filter the list of models, see Figure 5. After choosing oral/extravascular as administration method, 1st-order absorption without delay, one compartment

distribution - according to the data visualization, the drug seems to follow a one-compartment model behaviour - and a linear elimination, two parametrizations are available: (ka, V, Cl) and (ka, V, k) . File `oral1_1cpt_kaVcl.txt` is selected

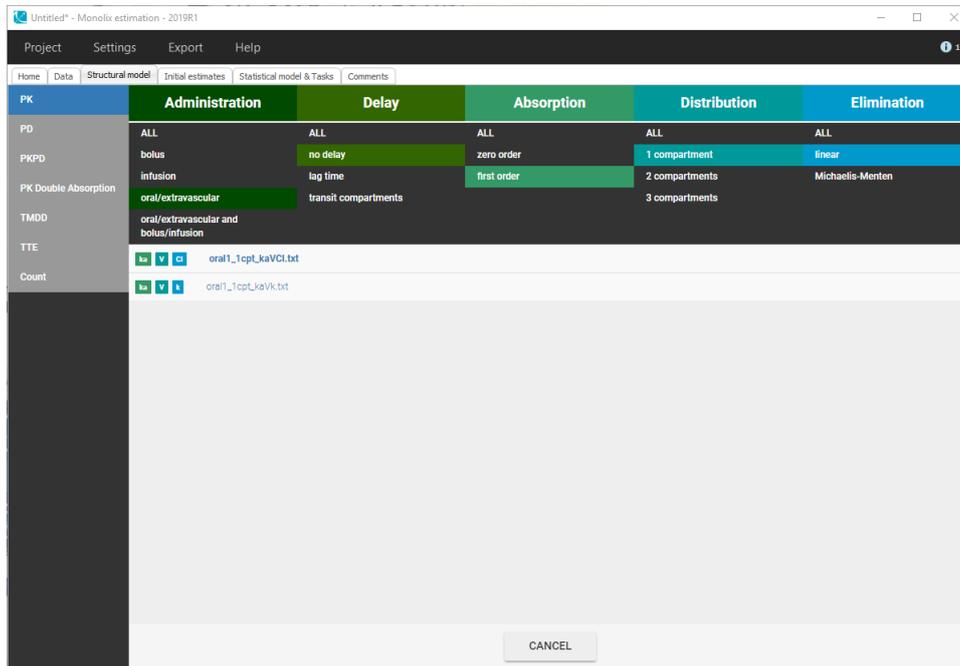


Figure 5: Structural model definition using Monolix PK library.

In this case study, the bioavailability of the drug is reported to be 50%. There are two equivalent approaches to take into account this information:

1. Use the original model from the library and then scale the estimated parameters by the factor $F=0.5$ as a post-processing of the estimates outside of Monolix:

$$Cl_{est} = \frac{Cl}{F} \quad \text{and} \quad V_{est} = \frac{V}{F}.$$

2. Modify the existing model from the library to add the bioavailability to the structural model.

The second approach is chosen. The txt file containing the model can be opened and modified in the MlxEditor by clicking "OPEN IN EDITOR" in the "Structural model" tab. To add bioavailability it is sufficient to modify the PK model definition by adding "p=0.5" as follows:

$$Cc = \text{pkmodel}(ka, V, Cl, p = 0.5).$$

Parameter p is a recognized keyword in the `pkmodel()` definition and automatically scales the volume and the clearance parameters.

A txt file with the new model definition has to be saved in the folder where the current project is located. Then, it is important to compile the file in the MlxEditor. It is a verification step to check if the code is written correctly and the file can be executed in Monolix. If no errors are reported, the structural model is loaded to Monolix by clicking "BROWSE" and choosing the txt file, see Figure 6.

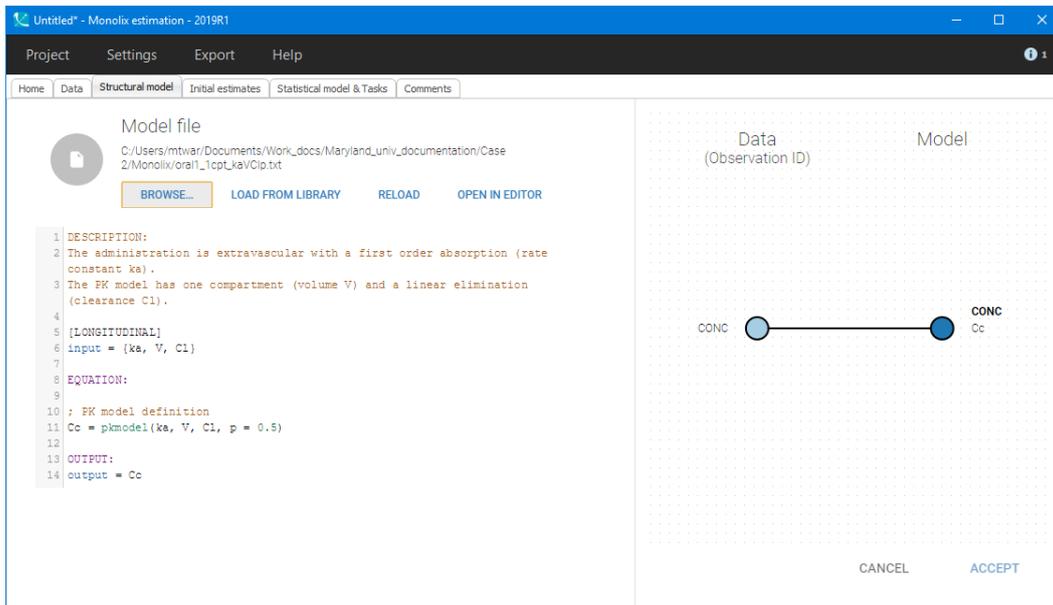


Figure 6: Structural model definition written in the Mlxtran language. The PK model from the library is modified to take into account a fixed bioavailability.

3.1.1 Initialization

At the next step the "Check initial estimates" tab is displayed, where initial parameter values have to be chosen. Although Monolix is very robust with respect to the initial parameter choice, it is a good practice to set reasonable values to speed up the convergence. By default, all values are set to one. They can be modified manually or by selecting "AUTO-INIT" button - a built-in algorithm finds the best fit for the pooled data.

Figure 7 shows concentration data points for several individuals (in blue) and the model prediction (in red). The auto-init option is unavailable, because the PK library model was modified. Thus, the parameters should be modified manually to find a good initial prediction. Setting a reference model prediction by clicking "REFERENCE" next to the green button "SET AS INITIAL VALUES" can help. On the figure, the reference prediction (dashed lines) corresponds to population parameters equal to one. To increase the accuracy of prediction curves, the time discretization grid is set to 1000 in the right panel at the bottom.

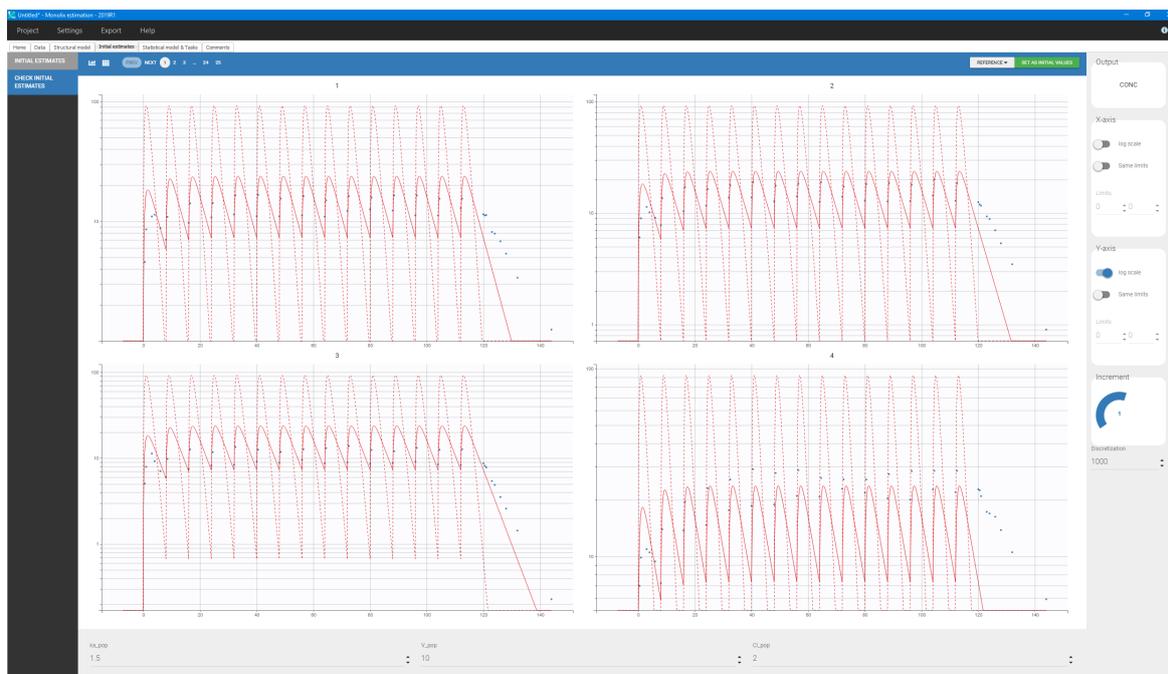


Figure 7: Graphical guide to choose initial parameters: concentration data points for several individuals (in blue), model prediction (red solid line), reference model prediction, which in this case corresponds to the population parameters equal to one (red dashed line). The time discretization grid is set to 1000.

Initial values are accepted by clicking the green box "SET AS INITIAL VALUES" and this brings the user to the next tab with the statistical part of the model. The base model uses:

- **Residual error model** - default combined1 model with an additive and a proportional term.
- **Parameter distributions** - default log-normal distributions for all parameters, and no correlations between random effects.

Before running any Monolix task, the project has to be saved, for example as "project_01.mlxtran" (menu: Project > Save), so that result files will be saved in a folder next to the project file.

3.1.2 Monolix tasks

Running a task in Monolix is done by clicking on the corresponding button in the "Tasks" section of the "Statistical model and tasks" tab. It is also possible to chose (tick) several tasks and then the green button RUN. For each task, settings are available via the wrench icon. They are for advanced use only; running typical PKPD models does not require to adjust them.

The following estimation tasks can be run:

- **Estimation of population parameters**
The goal of this task is to calculate the population parameters that characterize the population: typical parameters in the population (called fixed effects), inter-individual variability (standard deviations of the random effects) and residual error parameters. In this case, the algorithm shows a good convergence: after an initial exploratory phase, the parameter estimates stabilize, see Figure 8. Once the task has finished, the estimated population parameters are displayed in a table in the "Result" tab, presented in Figure 9.

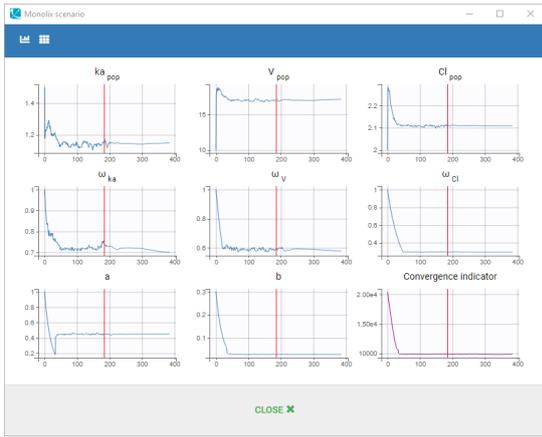


Figure 8: Estimation of population parameters using the Stochastic Approximation Expectation Maximization (SAEM) algorithm.

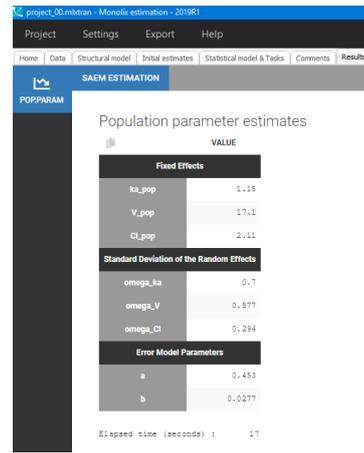


Figure 9: Results: population parameter estimates obtained by the Stochastic Approximation Expectation (SAEM) algorithm.

- **Estimation of the EBEs.**

This task calculates the most probable parameter values for each individual (also called EBEs for empirical bayes estimates). Figure 10 presents the results: the estimates of individual parameters (on the left) and a summary (on the right), outputted in the "Results" tab. The EBEs are useful to generate the individual predictions used in plots such as the "Individual fits".

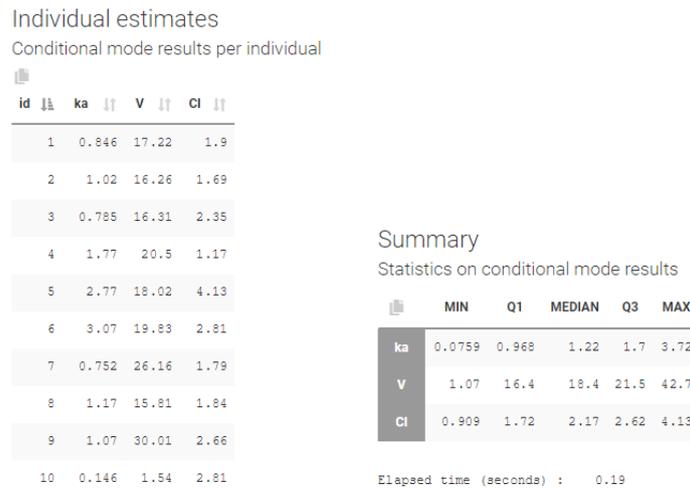


Figure 10: Results: individual parameters estimated (on the left) and a summary (on the right).

- **Estimation of the conditional distribution.**

This task permits to sample individual parameter values from the conditional distribution of each individual using a MCMC procedure. The conditional distribution represents the uncertainty of the individual parameters. These values improve the diagnostic power of the plots. In the "Results" tab, the conditional mean (i.e average from the sampled values) is displayed for each individual.

- **Estimation of the standard errors via the Fisher Information Matrix (FIM).**

 two-compartment model at hand runs fast, so the stochastic approximation option is chosen by deselecting the toggle "Use linearization method". The results are:

- The estimated standard errors and relative standard errors are displayed in the "POP.PARAM" sub-tab, see Figure 11 on the left. They represent the uncertainty of the estimated popu-

lation parameters. In this run, all parameters are estimated with a good confidence (RSE < 10%).

- In the "STD.ERRORS" tab, the correlation matrix of the estimates is shown, see Figure 11 on the right. Values close to 1 or -1 on off-diagonal terms indicate that the corresponding population parameter values are highly correlated. Strong correlations usually suggest a poorly identifiable model. In this example, the error model parameters "a" and "b" are negatively correlated, which means that an almost equally good likelihood could be obtained by increasing a and decreasing b, or doing the opposite. As none of the value for "a" and "b" is small, a "combined2" error model, which has a different formula, could be tried.
- The ration of maximal and minimal eigenvalues is called the condition number. A low value (i.e < 100) suggests that the model is not over-parameterized.

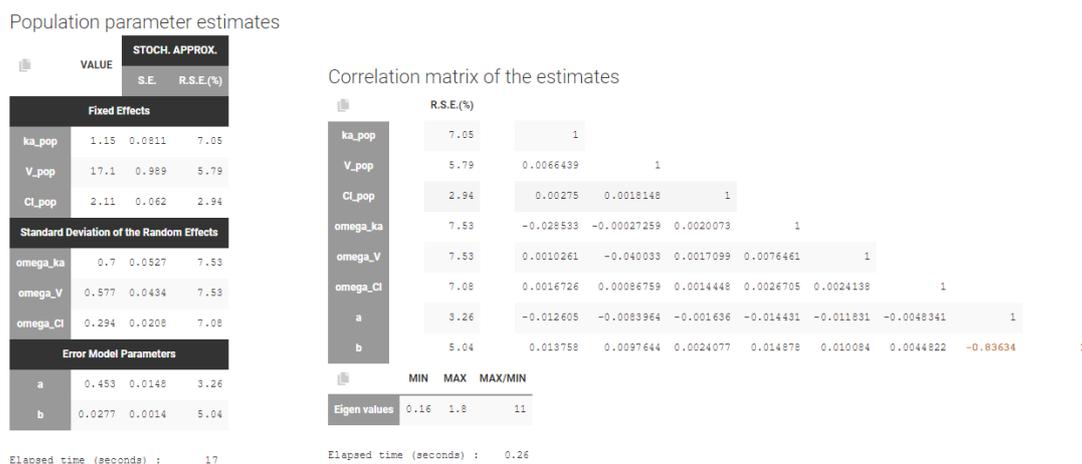


Figure 11: Results: population parameters estimated with standard errors (on the left) and correlation matrix of the estimates (on the right).

- **Estimation of the log-likelihood**

The log-likelihood is the objective function optimized during the population parameter estimation. The importance sampling gives a -2log-likelihood, AIC and BIC criteria, which are displayed in the "LIKELIHOOD" sub-tab. The AIC, BIC and BICc criteria take into account the number of estimated parameters and the number of observations, and can be used to compare models with each other.

- **Generation of the diagnostic plots**

This task generates automatically selected built-in graphics. By default, a small set of rapid-to-generate plots are selected. Clicking on the "list" icon of the right of the "PLOTS" button shows the full list (including VPC, etc), see Figure 12

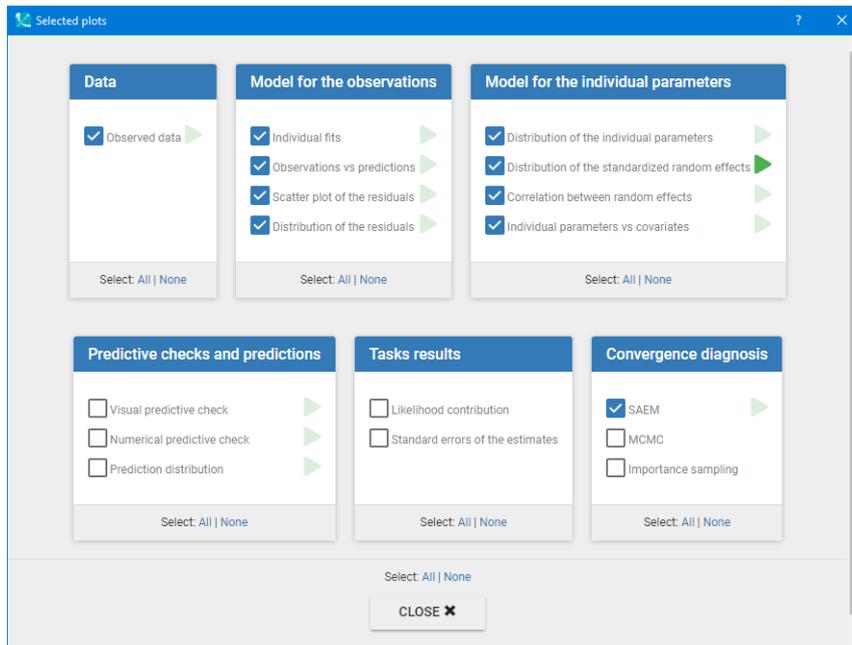


Figure 12: List of built-in diagnostic plots.

3.1.3 Diagnostic plots

The primary analysis of the goodness of the model is done through the diagnostic plots. The panel on the left of the "Plots" tab is divided into sections that correspond to model aspects that can be assessed by particular plots. The panel on the right allows to set many options for all plots: what to display (tab "Settings"), the layout such as colors, font size, etc (tab "Preferences") and stratification via splitting/coloring/filtering based on covariates (tab "Stratify").

At first the assessment of the structural model using the "Model for the observations" section is performed.

- **Individual fits** compare model predictions using the individual parameters with the observed data. Results, presented in Figure 13 for several individuals, show that the later phase of the elimination is underpredicted. Pages for all individuals can be browsed with the buttons at the top.

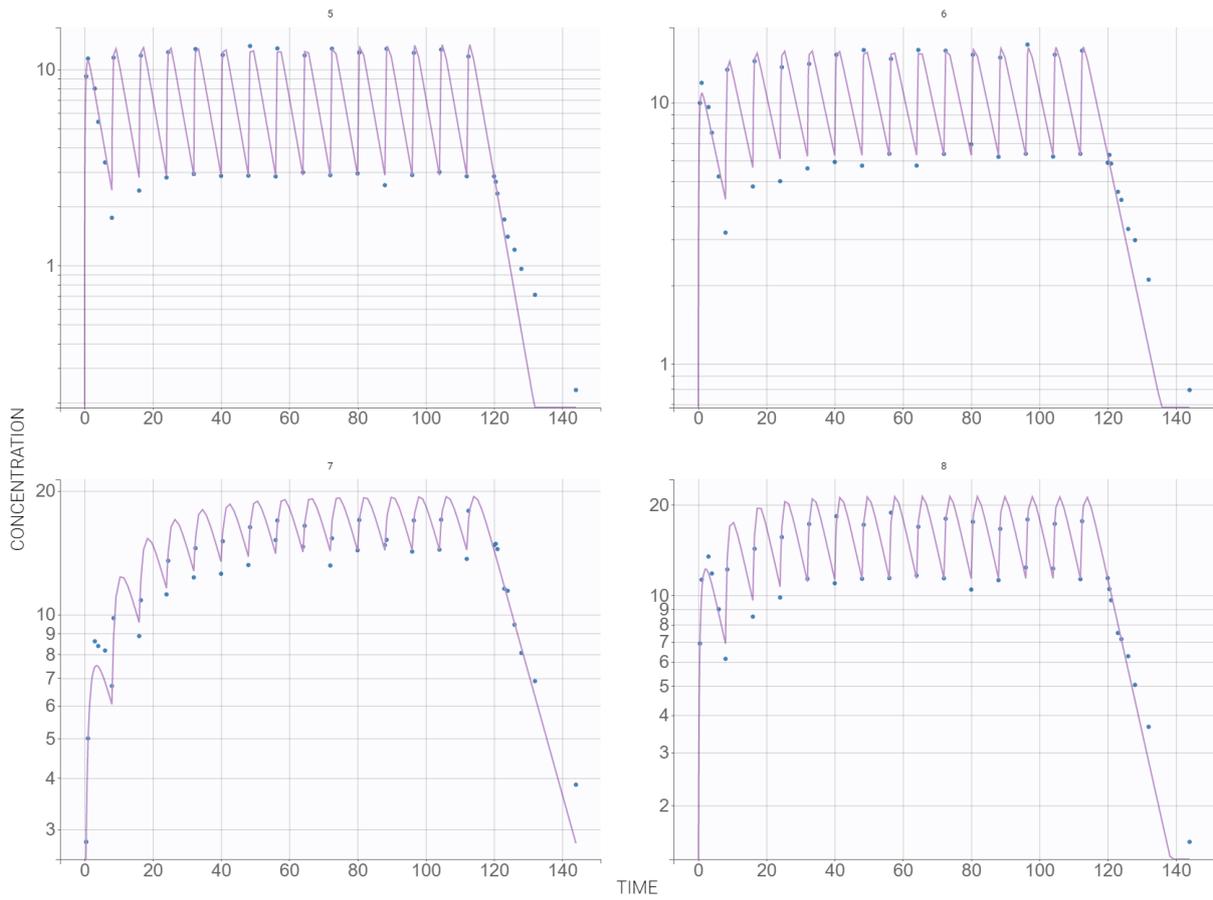


Figure 13: Diagnostic plot: individual fits for the one-compartment model with 1st-order absorption without delay and linear elimination. Log-scale for all individuals have been chosen in the tab "Settings".

- **Observations vs. predictions** plot is used to analyze possible mis-specifications of the structural model. In Figure 14 individual estimates sampled from the conditional distribution are used to generate the individual predictions. Points should be well balanced on each side of the $y = x$ line, but in this case they diverge for low values. It confirms the under-prediction observed previously at individual fits plots.

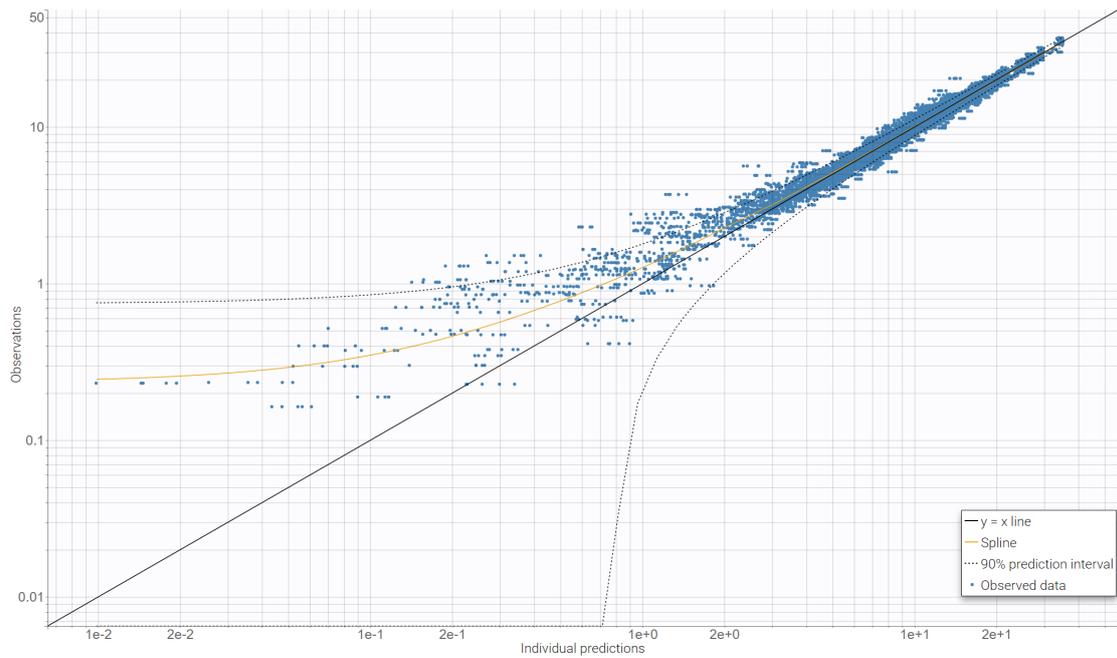


Figure 14: Diagnostic plot: observations vs. predictions for the one-compartment model with 1st-order absorption without delay and linear elimination. Log-log scale has been chosen, as well as the overlay of the 90% prediction interval.

- **Scatter plot of the residuals** displays the PWRES (population weighted residuals) and the IWRES (individual weighted residuals) w.r.t. time, as seen in Figure 15. The residuals should be symmetrically scattered around the horizontal zero-line 0, indicating that residuals behave as independent standardized normal random variables. However, this is not the case here, in particular at high times.

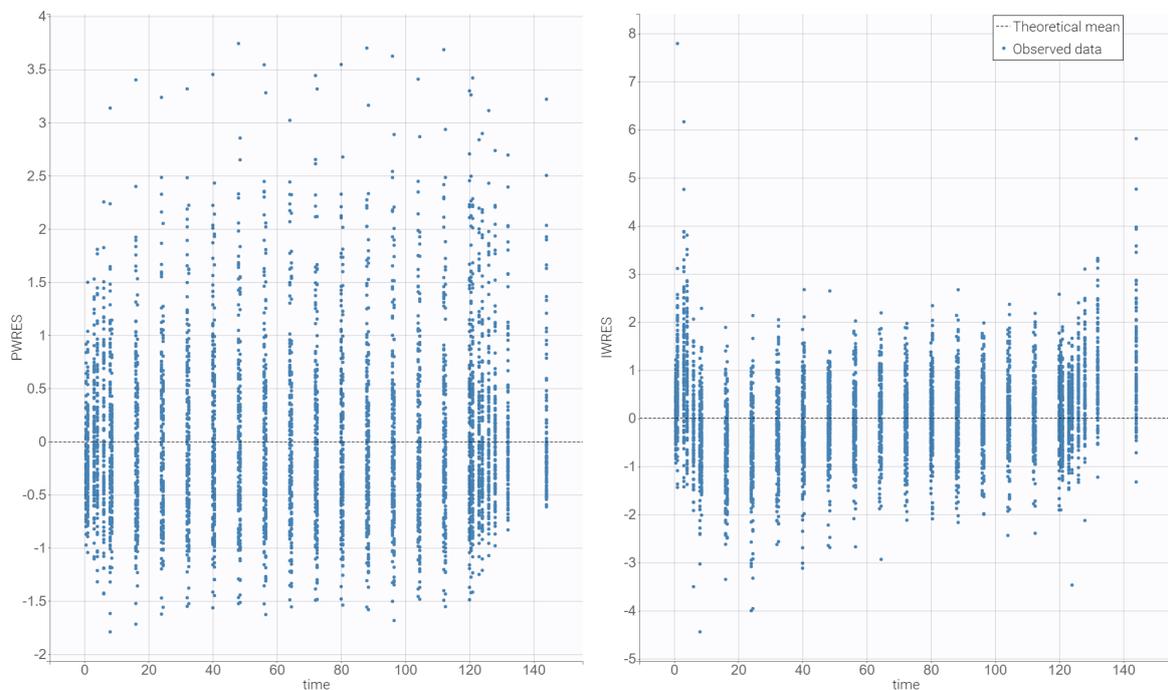


Figure 15: Diagnostic plot: scatter plots of PWRES and IWRES for the one-compartment model with 1st-order absorption without delay and linear elimination.

- **Distribution of the residuals** helps to detect mis-specification in the model. Figure 16 shows that the empirical and theoretical probability density functions (PDF) of the PWRES and IWRES do not agree and indicates mis-specification.

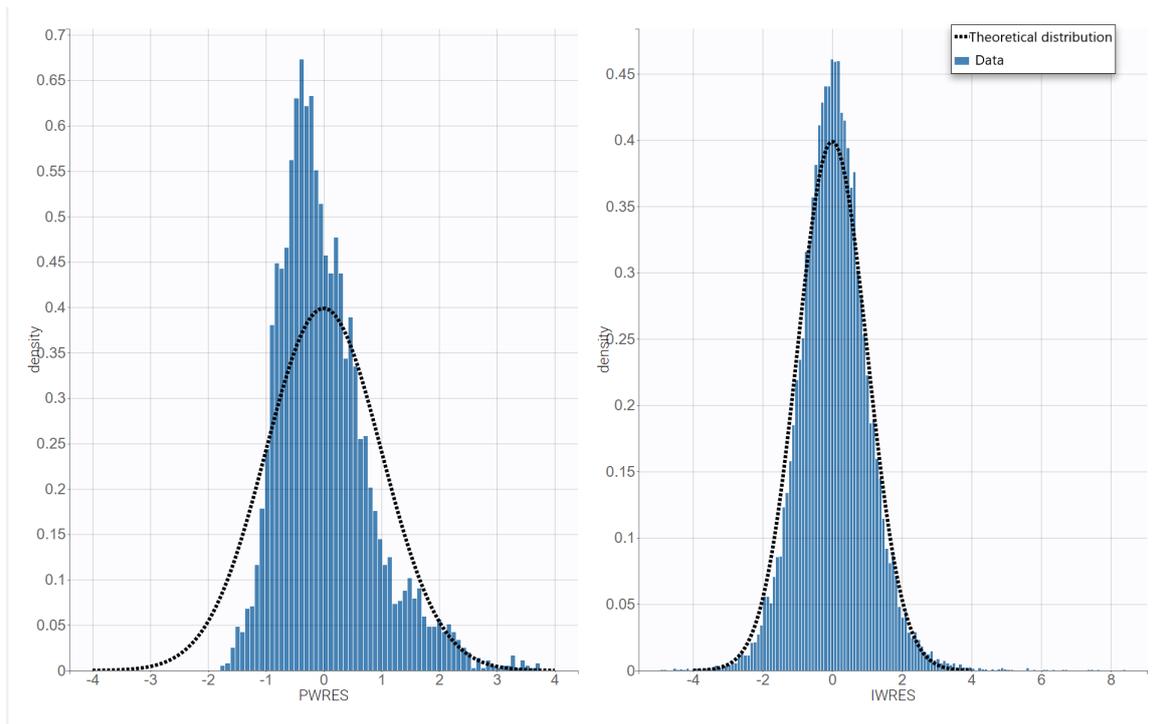


Figure 16: Diagnostic plot: distribution of the empirical (histogram) and theoretical (dashed line) PDF of the PWRES and IWRES for the two-compartment model with 1st-order absorption without delay and linear elimination.

The diagnostic plots show that the one-compartment model underpredicts low values of concentration during the later phase of the elimination. A structural model with two-compartment may improve the predictive power.

3.2 Structural model: two-compartment

The two-compartment model is chosen from the Monolix library: 1st-order absorption without delay, 2 compartments and linear elimination with the parameters (k_a , Cl , V_1 , Q , V_2). Then, the txt file with the model is modified as previously to take into account the bioavailability.

Good initial values should exhibit two phases of elimination. Results of the population estimates for the one-compartment model are used to initialize k_a -pop, Cl -pop and V_1 -pop, while remaining parameters are set to 1, see Figure 17.

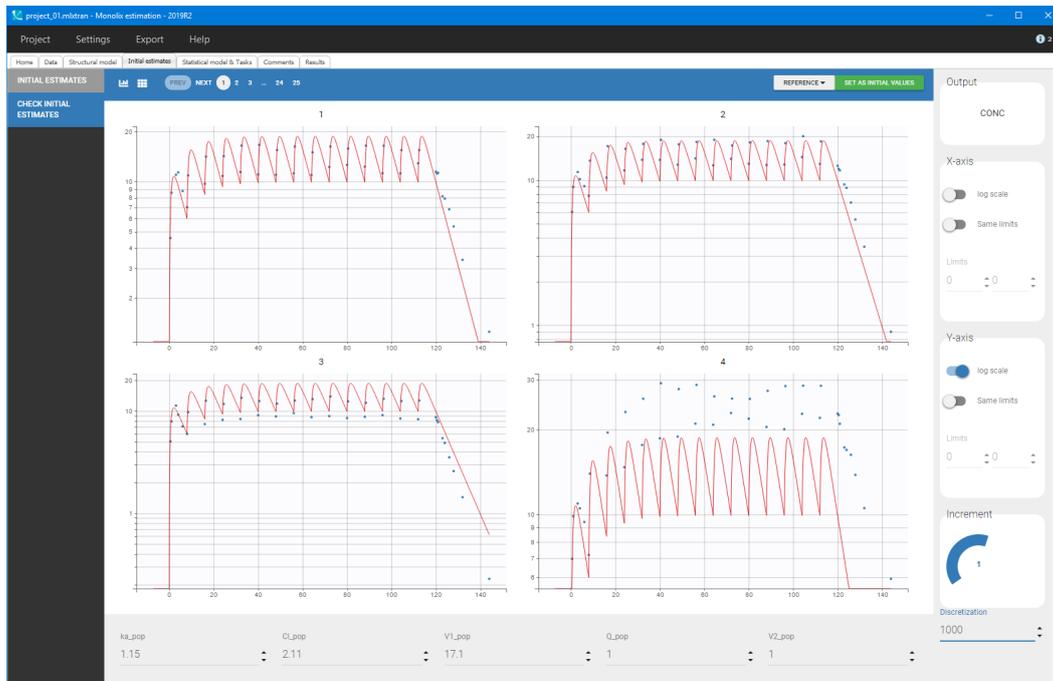


Figure 17: Visual way to choose initial parameters: concentration data points for several individuals (in blue) and model prediction (red solid line) for the two-compartment model.

Using the default statistical model, all Monolix tasks are initialized by clicking the green button RUN. Figure 18 presents the individual fits for the same patients as shown previously, and Figure 19 shows the observations vs. predictions plot. Adding a second compartment improves the prediction at low concentrations and removes the mis-specifications.

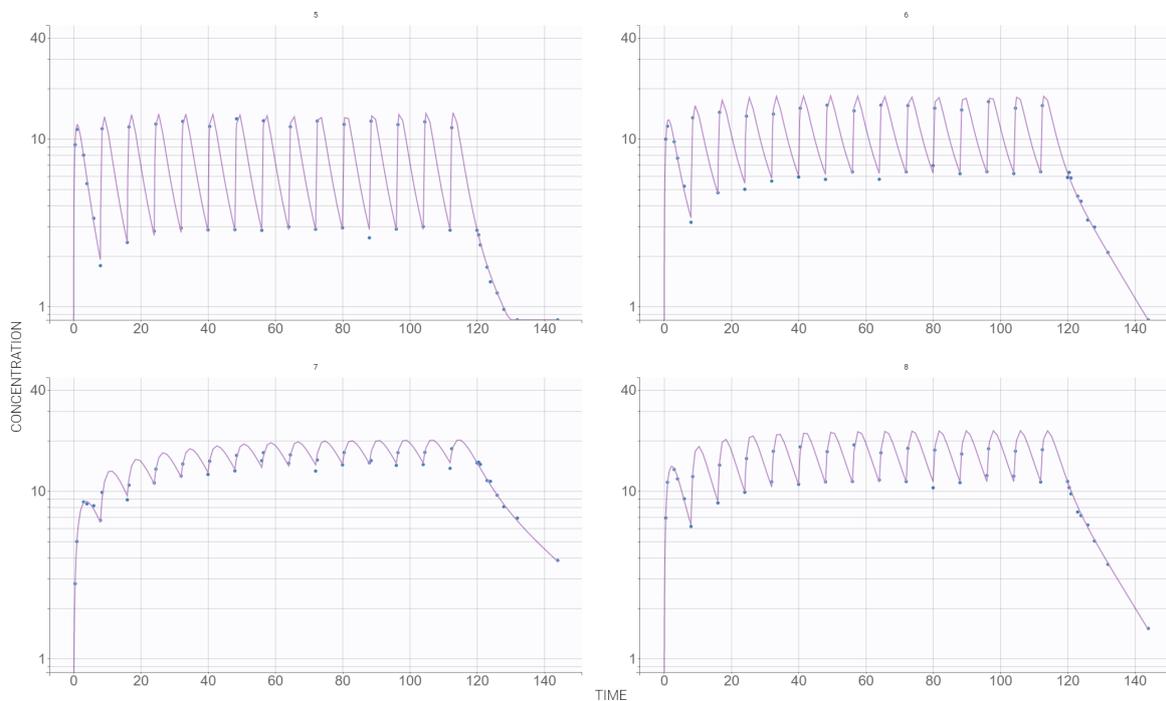


Figure 18: Diagnostic plot: individual fits for the two-compartment model with first-order absorption without delay and linear elimination. Log-scale for all individuals have been chosen in the tab "Settings".

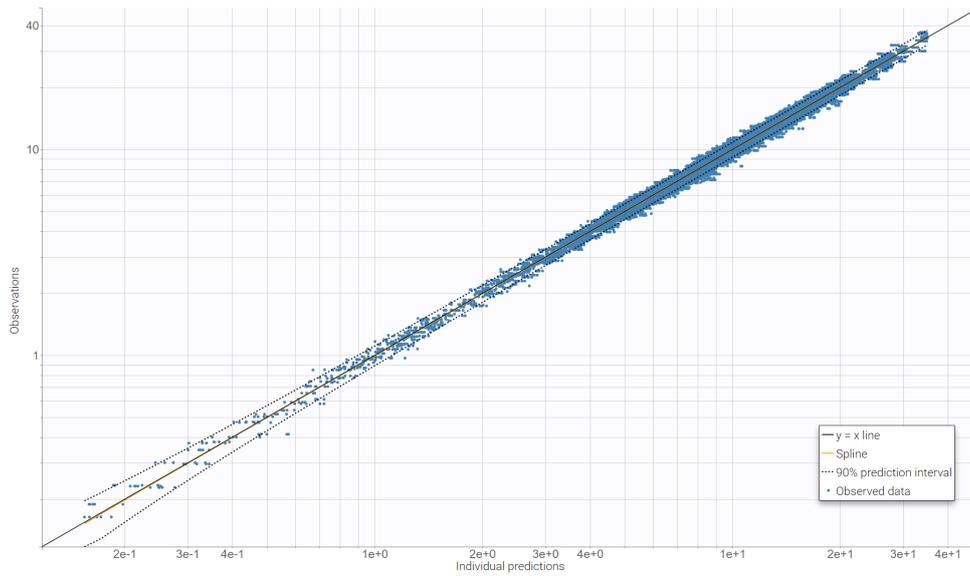


Figure 19: Diagnostic plot: observations vs. predictions for the two-compartment model with 1st-order absorption without delay and linear elimination. Log-log scale has been chosen, as well as the overlay of the 90% prediction interval.

The two models, with one and two compartments, are compared in Sycomore and presented in Figure 20. Parameter estimates have small standard errors in both cases, but the BICc criterion is much lower for the second model.

The VPC plot gives a graphical assessment to check whether simulations from the model of interest are able to reproduce both the central trend and the variability in the observed data, when plotted versus an independent variable (typically time). The observed empirical percentiles for the model with two compartments remain within the prediction intervals, which means that the model properly captures the data, see Figure 21.

Project Name	project_00	project_01
-2*LL (Importance sampling)	11117.38	8042.35
BICc (Importance sampling)	11172.91	8123.77
Fisher tasks		
Fixed effects		
	S.A.	
ka_pop	1.15	7.05%
V_pop	17.07	5.79%
Cl_pop	2.11	2.94%
V1_pop		11.74
Q_pop		1.59
V2_pop		7.46
Standard deviation		
	S.A.	
omega_Cl	0.294	7.05%
omega_V	0.577	7.53%
omega_ka	0.7	0.41
omega_Q		0.655
omega_V1		0.374
omega_V2		0.348
Residual error parameters		
	S.A.	
a	0.453	3.26%
b	0.0277	5.04%

Figure 20: Sycomore: comparison between the one compartment model (on the left) and the two compartment model (on the right). Ist-order absorption without delay and linear elimination is common for both models.

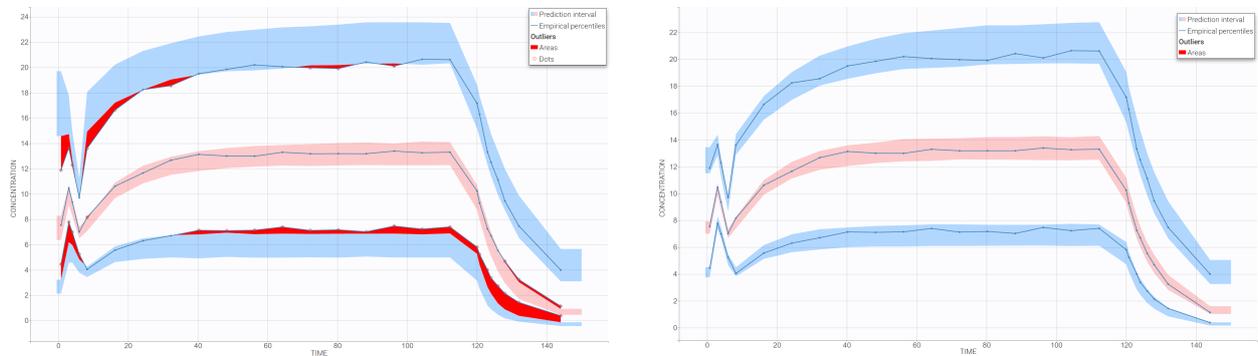


Figure 21: VPC comparison between the one compartment model (on the left) and the two compartment model (on the right). Ist-order absorption without delay and linear elimination is common for both models.

3.3 Statistical model

Goodness of individual fits for the two-compartments model is satisfactory. The data set does not contain any covariates, so the next step is to assess if there are any [correlations between the random effects](#). In Monolix there are several tools to guide the user in exploring how to improve a statistical model: the diagnostic plots, the tests and the proposals displayed in the "Results" section.

The diagnostic plots, see Figure 22, show strong correlations between the random effects of the parameters V1, ka and Cl (coefficients ~ 0.4).

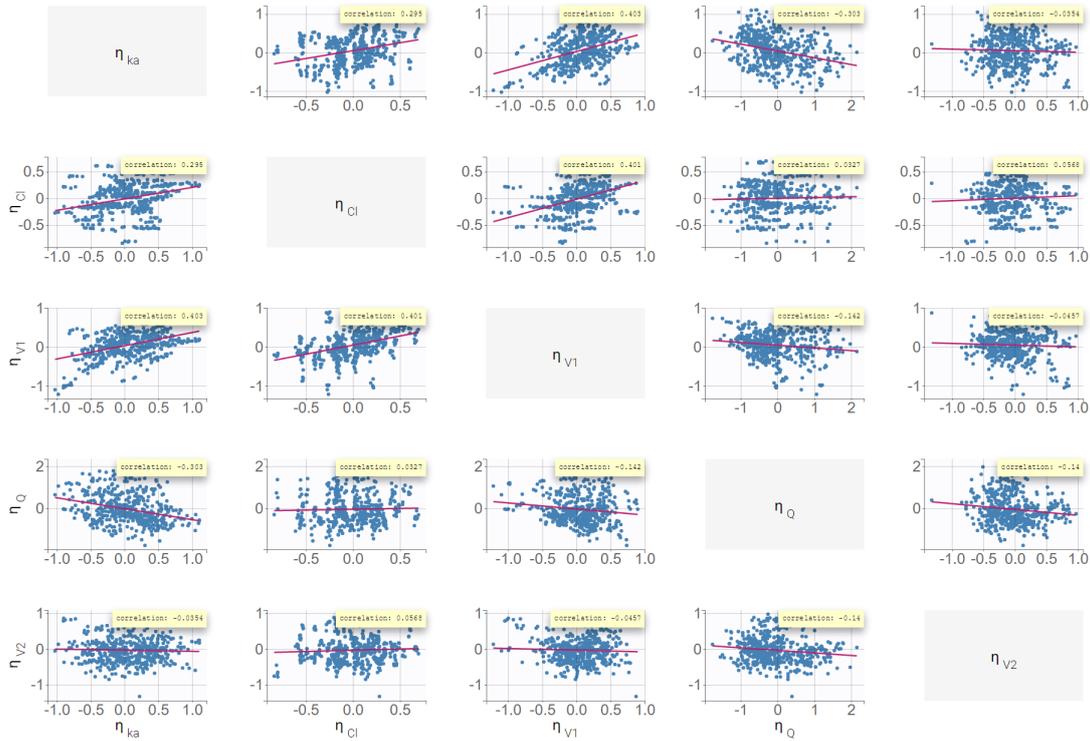


Figure 22: Diagnostic plots: correlations between random effects. The coefficients for the correlations between η_{V1} , η_{Cl} and η_{ka} are around 0.4).

The significance of the correlations is confirmed by the **t-test**, which is available in the "Results" tab and presented in Figure 23. A small p-value indicates that the null hypothesis - the expectations of correlations between (η_{V1}, η_{ka}) , (η_{V1}, η_{Cl}) and (η_{Cl}, η_{ka}) are zero - can be rejected and the correlations are significant.

Joint distribution of the random effects
test if the random effects are correlated

Correlation test (t-test):

	STATISTICS	P-VALUE
eta_ka eta_Cl	2.78245	0.00635169
eta_ka eta_V1	3.74692	0.000301087
eta_ka eta_Q	-3.49965	0.000699874
eta_ka eta_V2	-0.520229	0.604065
eta_Cl eta_V1	4.26166	4.63392e-5
eta_Cl eta_Q	0.364909	0.715958
eta_Cl eta_V2	0.587441	0.558245
eta_V1 eta_Q	-1.57193	0.119157
eta_V1 eta_V2	-0.490538	0.624839
eta_Q eta_V2	-1.89018	0.0616589

Figure 23: Statistical test: correlation t-test to check if the random effects are correlated.

Monolix **Proposal**, which evaluates how different correlation models fit the current simulated parameters, also suggests to add these correlations to the statistical model, see Figure 24.

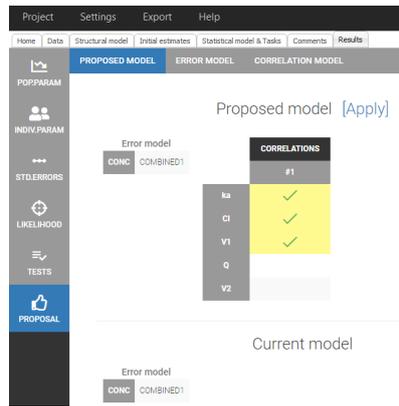


Figure 24: Monolix "proposal" to add correlations between parameters.

The correlations are added in the "Statistical model and tasks" tab, in the section dedicated to the Individual model, by clicking corresponding boxed in the column CORRELATIONS, as presented in Figure 25.

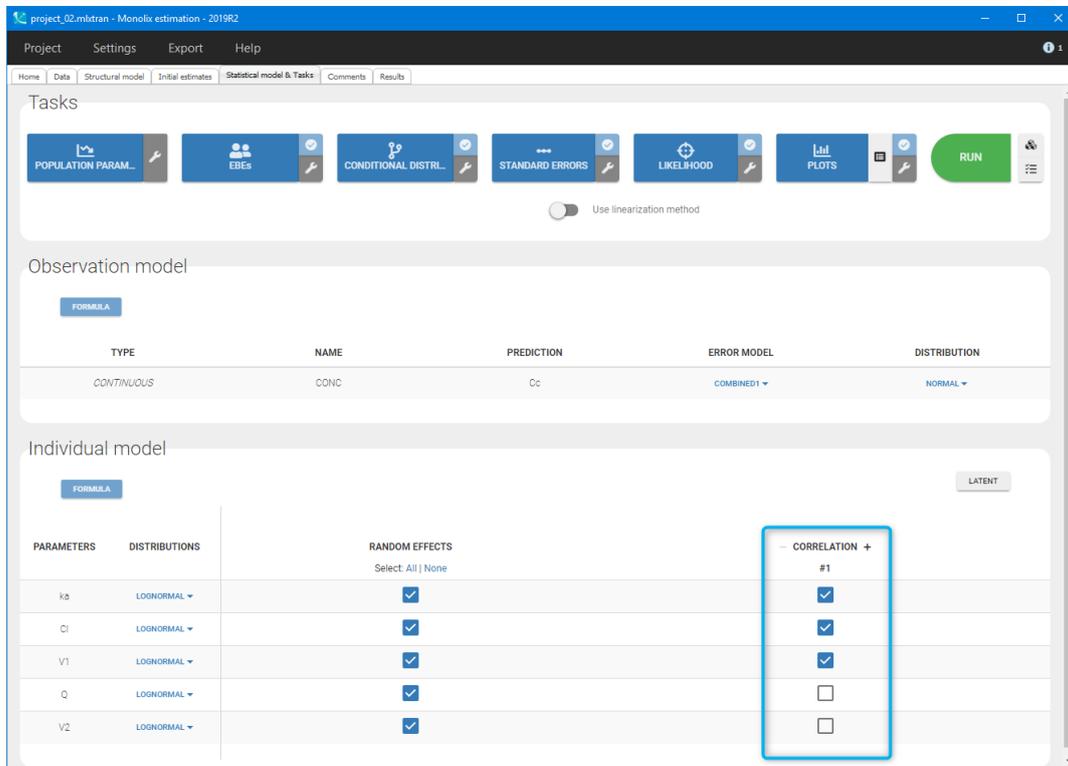


Figure 25: Statistical model: adding correlations.

After adding the correlations, the project is saved as "project.03.mlxtran" and all the tasks are run again. Improving the statistical model improves the information criteria BICc by 8 points (from 8124 to 8117).

4 Summary of results

This case study focused on a development of a population PK model of a hypothetical drug. The structural part is a two compartment model with first-order absorption without delay and linear elimination. Model parameters were estimated using the SAEM algorithm implemented in Monolix. Subsequently, using diagnostic plots and statistical tests, significant correlations were found and added to the statistical model, which improved the BICc information criteria.

The Sycomore application, part of the MonolixSuite, is designed to organize modeling workflow. It clearly displays summary of results, see Figure 26, and allows to compare different models side by side, see Figure 27.

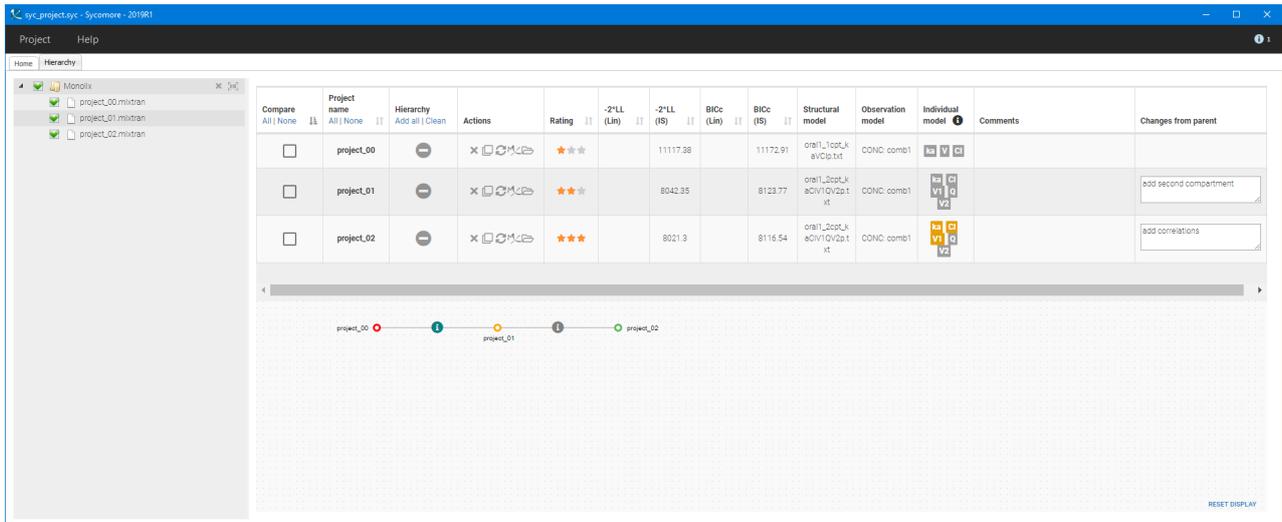


Figure 26: Final work-flow managed in Sycomore.

Project Name	project_01	project_02
Rating	☆☆☆	☆☆☆
Comments		
SAEM	✓	✓
Likelihood results		
-2*LL (Linearization)	▶	▶
BICc (Linearization)		
-2*LL (importance sampling)		8021.3
BICc (importance sampling)		8123.77
Fisher tasks		
Linearization	▶	▶
Stochastic approximation	✓	✓
Fixed effects		
	S.A.	S.A.
ka_pop	0.741	4.2%
Cl_pop	2.03	2.93%
V1_pop	11.74	3.81%
Q_pop	1.59	7.31%
V2_pop	7.46	3.62%
Standard deviation		
	S.A.	S.A.
omega_Cl	0.293	7.05%
omega_Q	0.685	8.21%
omega_V1	0.374	7.61%
omega_V2	0.348	7.98%
omega_ka	0.41	7.92%
corr_V1_Cl		0.337
corr_ka_Cl		0.295
corr_ka_V1		0.622
Residual error parameters		
	S.A.	S.A.
a	0.0179	33.55%
b	0.0489	1.85%

Figure 27: Sycomore: comparison between two models: with and without correlations.