User Guide for PKgraph Package

Xiaoyong Sun[†‡*]

December 1, 2010

[†]Binformatics and Computational Biology Program, [‡]Department of Statistics
Iowa State University, Ames, Iowa 50010, USA

Contents

1 Introduction **2**

2 Installation **2**

3 PKgraph infrastructure **3**
3.1 Graphical user interfaces 3
3.1.1 Main interface 3
3.1.2 Graph interface 4
3.2 Functional module 5

4 Functions **5**
4.1 Project 6
4.2 Configure 6
4.3 Data management 7
4.4 Exploratory data analysis 7
4.4.1 Univariate 8
4.4.2 Bivariate 8
4.4.3 Parallel coordinate plots 9
4.5 PK models 9
4.5.1 Configure model result 10
4.5.2 Parameters 11
4.6 Model validation 11
4.6.1 Influence analysis summary (PsN) 12
4.6.2 Visualization for influence analysis 12
4.6.3 Bootstrap summary (PsN) 13
4.6.4 Visualization for bootstrap 13
4.7 Model comparison 14
4.7.1 Select datasets 15
4.7.2 Configure mapping 15
4.7.3 Comparison 15
4.8 Interactive graphics 16

[*]johnsunx1@gmail.com

5 Example **17**

1 Introduction

Population pharmacokinetic (PopPK) modeling has become increasing important in drug development because it allows unbalanced design, sparse data and the study of individual variation. However, this complexity of the model makes it a challenge to diagnose the fit. Graphics can play an important and unique role in PopPK model diagnostics. The software described in this paper, PKgraph, provides a graphical user interface for PopPK model diagnosis with interactive graphics. It also provides an integrated and comprehensive platform for analysis of pharmacokinetic data including exploratory data analysis, goodness of model fit, model validation and model comparison. It can be used with a variety of modeling fitting software, including NONMEM, Monolix, SAS and R. PKGraph is programmed in R, and uses the R packages lattice, ggplot2 for static graphics, and rggobi for interactive graphics. This R package is supported with a user-friendly graphical user interface so that users can easily control diagnosing with simple clicks. The PKGraph software serves as a supplement to the existing packages: NONMEM, Xpose and PsN for diagnosing models.

PKgraph is a R packaged built on the following R packages: RGtk2, gWidgets, gWidgetsRGtk2, lattice, and ggplot2. It requires R (> 2.0) and GTK+, and runs under Windows, Linux and Mac.

2 Installation

PKgraph needs to install the following programs and R packages:

1. install GTK
For Windows, you can download the GTK Developer's Pack from
http://gladewin32.sourceforge.net/

For Unix, you can fetch the source files for the different libraries from
ftp://ftp.gtk.org/pub/gtk/v2.8/

2. Install RGtk2 (Please see RGtk2 Installation notes if you have problems)
install.packages("RGtk2")

3. install rggobi
a. Download and install ggobi (www.ggobi.org)
b. Install rggobi: *install.packages("rggobi")*

4. Install gWidgets
install.packages("gWidgets")

5. Install cairoDevice
install.packages("cairoDevice")

6. Install gWidgetsRGtk2
install.packages("gWidgetsRGtk2")

7. Install lattice
install.packages("lattice")

8. Install ggplot2
install.packages("ggplot2")

3 PKgraph infrastructure

The software incorporates a key concept: interactive graphics to link various datasets and diagnostics plots. The framework is programmed using RGtk2 and consists of main formats of interfaces, (1) main, containing links to all parts of the software, and handles the basic data management, and links to diagnostic modules, and (2) graph, which provides tools specifically for each diagnostic module. (2). Basic module: data input/output module, configuration module, and data management module.

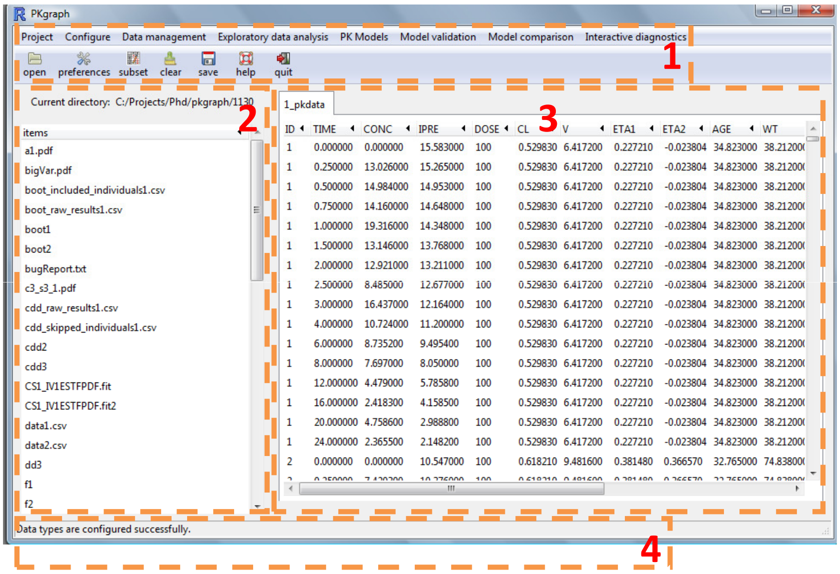
3.1 Graphical user interfaces

3.1.1 Main interface

The main interface (Figure˜1) of PKgraph provide the links to all components of the software. There are four areas: (1) tool area (tool bar and menu bar, top), (2) directory area (middle-left), (3) data area (middle-right) and (4) status bar (bottom).

- The tool area has menu items linking to the basic management modules (project, configuration, data management) and the diagnostic modules (exploratory data analysis, PK models, model validation, model comparison and interactive diagnostics). These are menu items containing numerous functions associated with each of the different types of diagnostics.
- The directory area shows current directory and all of its files. These files might be data files, or code, depending on the modeling software used.
- Clicking on any of the data files, will open them and display them in the data area (3). Choosing the file also brings up a panel allowing for different formats to be read, thus handling all possible modeling software formats. The data files might contain raw data, and model diagnostics such as parameter estimates, fitted values and residuals and these are displayed in the table view of the data area.

- The stats bar displays the progress of the different functions, for example here it says “Data is loaded successfully” to indicate that there were no problems with opening the data file.



Current directory: C:/Projects/Phd/pkgraph/1130

a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2
data1.csv
data2.csv
dd3
f1
f2

ID	TIME	CONC	IPRE	DOSE	CL	V	ETA1	ETA2	AGE	WT
1	0.000000	0.000000	15.583000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.250000	13.026000	15.265000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.500000	14.984000	14.953000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.750000	14.160000	14.648000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	1.000000	19.316000	14.348000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	1.500000	13.146000	13.768000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	2.000000	12.921000	13.211000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	2.500000	8.485000	12.677000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	3.000000	16.437000	12.164000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	4.000000	10.724000	11.200000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	6.000000	8.735200	9.495400	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	8.000000	7.697000	8.050000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	12.000000	4.479000	5.785800	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	16.000000	2.418300	4.158500	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	20.000000	4.758600	2.988800	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	24.000000	2.365500	2.148200	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
2	0.000000	0.000000	10.547000	100	0.618210	9.481600	0.381480	0.366570	32.765000	74.83800

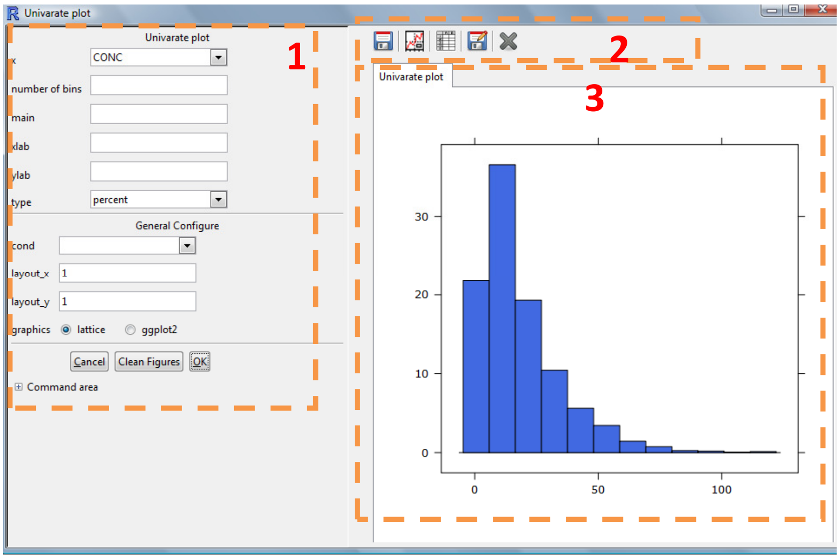
Data types are configured successfully.

Figure 1: Main interface of PKgraph

3.1.2 Graph interface

Selecting an item from a diagnostic module menu brings up a graph interface (Figure˜2). The style of the interface is the same for all diagnostic functionality. It contains three areas: 1) parameter setup area, 2) tool bar, 3) plot area.

- The parameter area setup allows choice of variable, plot labels, layout for trellis or facetted plots. A choice of lattice or ggplot2 graphics is provided.
- The tool bar allows users to save plots, open plots for interactive graphics (**ctrl + b for brushing data**), display subset selection from ggobi, save subset selection from ggobi and close ggobi.
- The plot area displays the figure, and multiple figures if more than one are created.



x: CONC

type: percent

layout_x: 1

layout_y: 1

graphics: ● lattice ○ ggplot2

Figure 2: Graph interface of PKgraph

3.2 Functional module

Functional module matches the menu items in PKgraph toolbar. It includes the following menu itmes:

- Project
- Configure
- Data management
- Exploratory data analysis
- PK models
- Model validation
- Model comparison
- Interactive graphics

In the next sections, I will go through each menu item in detail.

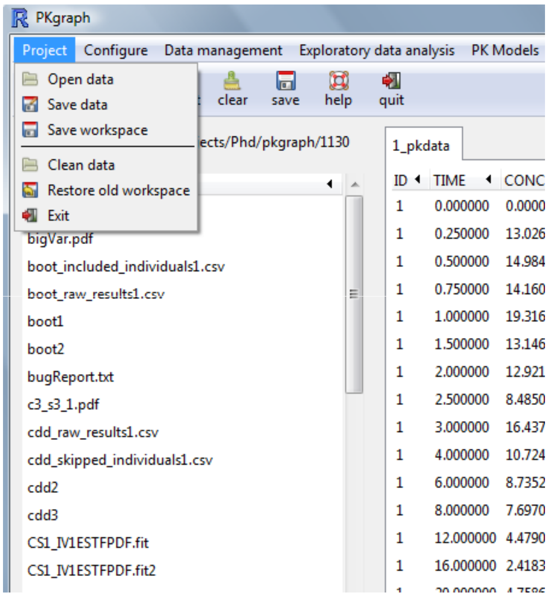
4 Functions

In this section, I will go through each function in the menu item of toolbar.

4.1 Project

This menu itme is in charge of input, output and save data. It has the following functions (Figure˜3),

- *Open data*: open modeling fit result from NONMEM, Monolix, SAS, R or other software. It has options to setup the data format, start line and separation symbol.
- *Save a file*: save a file.
- *Save a workspace*: save a workspace for later usage. It generally saves a group of lists for configuration and related data.
- *Clean data*: clean all loaded data.
- *Restore old workspace*: restore the workspace from the data and list you saved from previous step.
- *Exit*: exit from PKgraph.



ects/Phd/pkgraph/1130

bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2

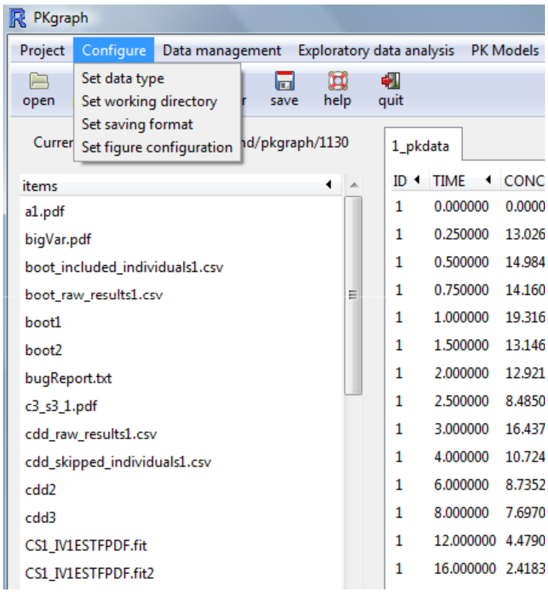
ID	TIME	CONC
1	0.000000	0.0000
1	0.250000	13.026
1	0.500000	14.984
1	0.750000	14.160
1	1.000000	19.316
1	1.500000	13.146
1	2.000000	12.921
1	2.500000	8.4850
1	3.000000	16.437
1	4.000000	10.724
1	6.000000	8.7352
1	8.000000	7.6970
1	12.000000	4.4790
1	16.000000	2.4183

Figure 3: Menu items in *Project*

4.2 Configure

This menu itme is utilized to configure PKgraph. It has the following functions (Figure˜4),

- *Set data type*: set the ID, TIME, CONC variables for current PK data.
- *Set working directory*: change current directory.
- *Set saving format*: set up saving format for figures, including pdf, jpg, tiff, png bmp, win.metafile, and figure width and height. If figure width and height is not configured, the default one will be used.
- *Set figure configuration*: color and loess can be selected for figures.



items
a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2

ID	TIME	CONC
1	0.000000	0.0000
1	0.250000	13.026
1	0.500000	14.984
1	0.750000	14.160
1	1.000000	19.316
1	1.500000	13.146
1	2.000000	12.921
1	2.500000	8.4850
1	3.000000	16.437
1	4.000000	10.724
1	6.000000	8.7352
1	8.000000	7.6970
1	12.000000	4.4790
1	16.000000	2.4183

Figure 4: Menu items in *Configure*

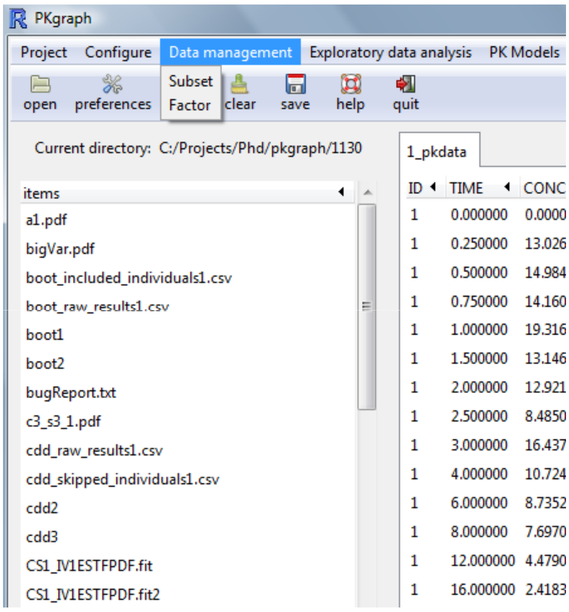
4.3 Data management

This menu item is utilized to manage data. It has the following functions (Figure˜5),

- *Subset*: subset current data.
- *Factor*: factor categorical variables. Graphical packages require the variable to be factor type in order to display the categorical symbol in figures.

4.4 Exploratory data analysis

This menu item is utilized to explore data and screen patterns. It has the following functions (Figure˜6),



Current directory: C:/Projects/Phd/pkgraph/1130

items
a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2

ID	TIME	CONC
1	0.000000	0.0000
1	0.250000	13.026
1	0.500000	14.984
1	0.750000	14.160
1	1.000000	19.316
1	1.500000	13.146
1	2.000000	12.921
1	2.500000	8.4850
1	3.000000	16.437
1	4.000000	10.724
1	6.000000	8.7352
1	8.000000	7.6970
1	12.000000	4.4790
1	16.000000	2.4183

Figure 5: Menu items in *Data management*

- *Univariates*: plot univariate varaiables.
- *Bivariates*: plot bivariate variables.
- *Parallel coordinate plot*: Parallel coordinate plot for multivariate variables.
- *Scatterplot matrix*: Scatterplot matrix for multivariate variables.

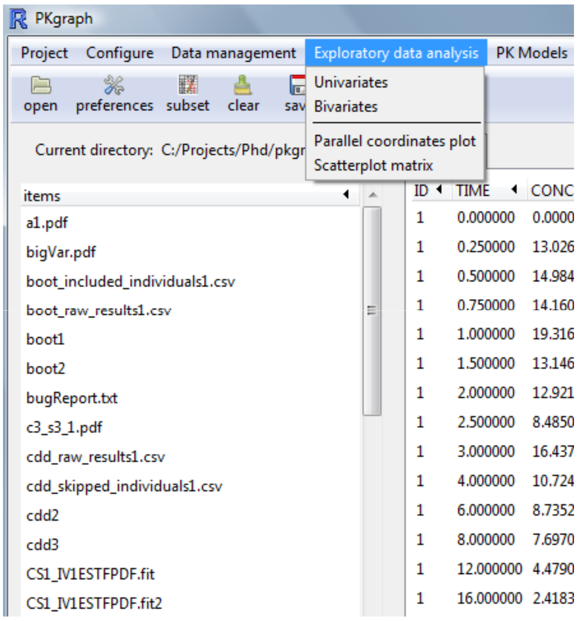
4.4.1 Univariate

When clicking this menu item, users will generate a "graph interface"(Figure˜2). In this interface, users can specify all parameters in the left area of window. In the right area of window, it has the following four buttons on the top:

- *save*: save figures.
- *ggobi*: open the plot in ggobi for interaction on the plot.
- *synchronize*: synchronize subset selection from ggobi to the data in R.
- *save synchronized data*: save synchronized subset.
- *close*: close ggobi.

4.4.2 Bivariate

This menu item also generates a "graph interface". It is similar to the Univariate interface, except that users will have two variables instead of one.



a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2

ID	TIME	CONC
1	0.000000	0.0000
1	0.250000	13.026
1	0.500000	14.984
1	0.750000	14.160
1	1.000000	19.316
1	1.500000	13.146
1	2.000000	12.921
1	2.500000	8.4850
1	3.000000	16.437
1	4.000000	10.724
1	6.000000	8.7352
1	8.000000	7.6970
1	12.000000	4.4790
1	16.000000	2.4183

Figure 6: Menu items in *Exploratory data analysis*

4.4.3 Parallel coordinate plots

This menu item provides access to lattice function: *parallel* function from lattice package.

4.5 PK models

This menu item is utilized to check model assumptions and goodness of fit. The guideline follows Census menu (http://census.sourceforge.net/). It has the following functions (Figure˜7),

- *Configure model result*: This is the key step to match data variables to default metric system. By this step, data from any platform (NONMEM, Monolix, SAS, R) can be interpreted graphically in figures.
- *Individual plots*: Bivariate plot for each individual.
- *Basic goodness of fit plots*: Goodness of fit plot is one of key tools to check model fitting. These kinds of plots will give an overall perspective of model performance, including scatter plot for concentration versus PRED, concentration versus IPRED, PRED versus time and IPRED versus time.
- *Parameters*: Generally, there are assumptions for distribution of parameters during modeling process. The histogram is utilized to check this

distribution. In addition, the correlation of parameters has significant effect on modeling performance, and it can be checked by scatter plots or a scatterplot matrix.

- *Random effects*: The assumptions for random effects also need to be tested for distribution and correlation by histogram, scatter plots or a scatterplot matrix.
- *Structural model*: Structural model can be diagnosed by PRED versus concentration conditioned on time, IPRED versus concentration conditioned on time, WRES versus time, WRES versus PRED, PRED versus concentration conditioned on covariates, IPRED versus concentration conditioned on covariates.
- *Residual error model*: Two assumptions are related to this submodel: 1) homoscedastic variability; 2) symmetrically distributed residuals. To test these assumptions, we applied the following techniques: 1) histogram for distributions of WRES; 2) histogram for individual distribution of WRES; 3) scatterplot of |WRES| versus PRED to check the shape of residual; 4) scatterplot of |WRES| versus PRED conditioned on covariates to screen the covariate effects; 5) autocorrelation of WRES.
- *Covariate model*: Parameters, ETA and WRES are of great use to help screen proper covariates. We can utilize the following methods to check covariate models: 1) scatter plot for parameters versus covariates, ETAs versus covariates, WRES versus covariates; 2) scatterplot matrix of covariates.

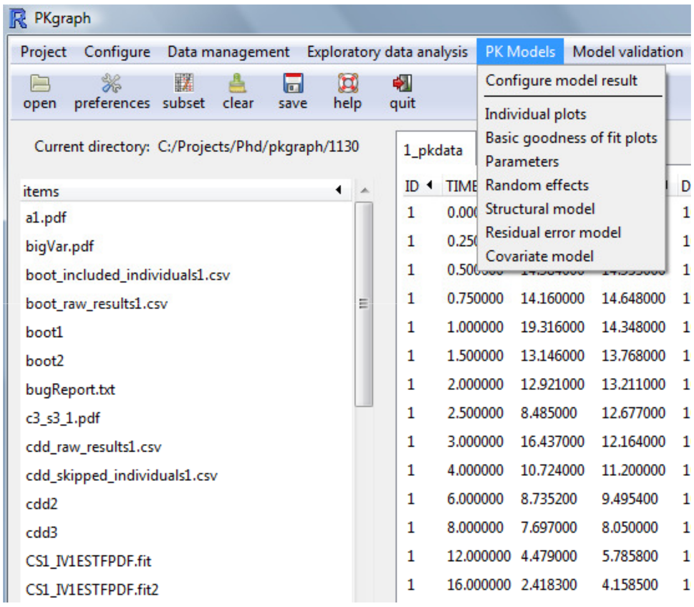
Users have to configure data variable first before going to specific model diagnostics. To illustrate the usage of this menu item, I will use two menu times: *Configure model result* and *Parameters* to demonstrate the model diagnostics.

4.5.1 Configure model result

The interface for this function is shown in Figure˜8. The fixed column is column name from data, and the selectable column is variable name from the default metric system (Table˜1).

Package variable	**Description**
ID	Patient ID
TIME	Time after dose
CONC	The concentration of drug in the body
PRED	Prediction generated from model fitting
RES	Residual
WRES	Weighted residual
IPRED	Individual prediction
IWRES	Individual weighted residual
COV	Covariates

Table 1: Package metric system



Current directory: C:/Projects/Phd/pkgraph/1130

items

- a1.pdf
- bigVar.pdf
- boot_included_individuals1.csv
- boot_raw_results1.csv
- boot1
- boot2
- bugReport.txt
- c3_s3_1.pdf
- cdd_raw_results1.csv
- cdd_skipped_individuals1.csv
- cdd2
- cdd3
- CS1_IV1ESTFPDF.fit
- CS1_IV1ESTFPDF.fit2

ID	TIME			D
1	0.00			1
1	0.25			1
1	0.50			1
1	0.750000	14.160000	14.648000	1
1	1.000000	19.316000	14.348000	1
1	1.500000	13.146000	13.768000	1
1	2.000000	12.921000	13.211000	1
1	2.500000	8.485000	12.677000	1
1	3.000000	16.437000	12.164000	1
1	4.000000	10.724000	11.200000	1
1	6.000000	8.735200	9.495400	1
1	8.000000	7.697000	8.050000	1
1	12.000000	4.479000	5.785800	1
1	16.000000	2.418300	4.158500	1

Figure 7: Menu items in *PK models*

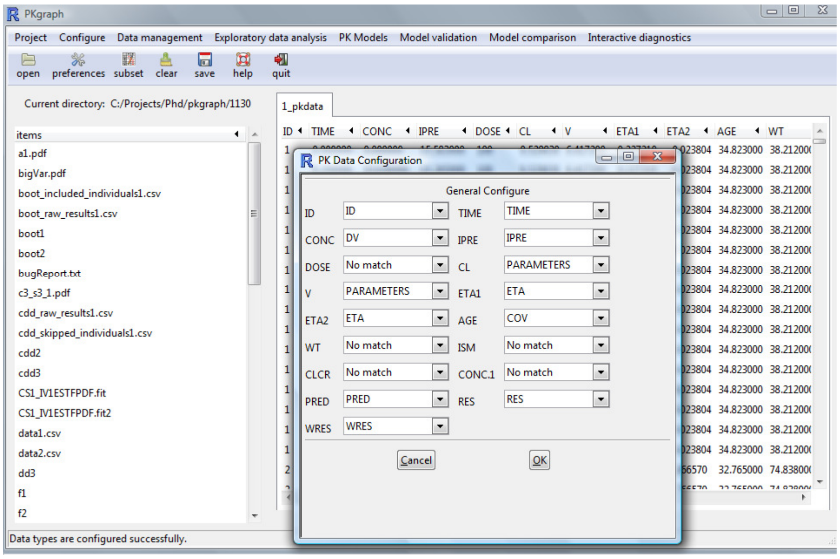
4.5.2 Parameters

The interface for this function is shown in Figure˜9. After users choose proper parameters in the left window, the system will produce all figures automatically. Users can pick specific figures for diagnosing with functions in the toolbar.

4.6 Model validation

Resampling methods has been extensively employed in the model validation. Currently, bootstrap targets for confidence interval, case deletion diagnostics identify influential cases, and stochastic simulation is utilized to compare models (PsN). PKgraph mainly focuses on case deletion diagnostics and bootstrap. It provides the following functions (Figure˜10),

- *Influence analysis summary (PsN)*: analyze PsN *cdd* results.
- *Visualization for influence analysis*: apply parallel coordinate plots and multidimensional scaling to visualize data from case deletion diagnostics (multiple NONMEM runs).
- *Bootstrap summary (PsN)*: analyze PsN *boot* results.
- *Visualization for bootstrap*: visualize data from bootstrap (multiple NONMEM runs).



Current directory: C:/Projects/Phd/pkgraph/1130

items
a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2
data1.csv
data2.csv
dd3
f1
f2

ID: ID
TIME: TIME
CONC: DV
IPRE: IPRE
DOSE: No match
CL: PARAMETERS
V: PARAMETERS
ETA1: ETA
ETA2: ETA
AGE: COV
WT: No match
ISM: No match
CLCR: No match
CONC.1: No match
PRED: PRED
RES: RES
WRES: WRES

Data types are configured successfully.

Figure 8: *Configure model result* in *PK models*

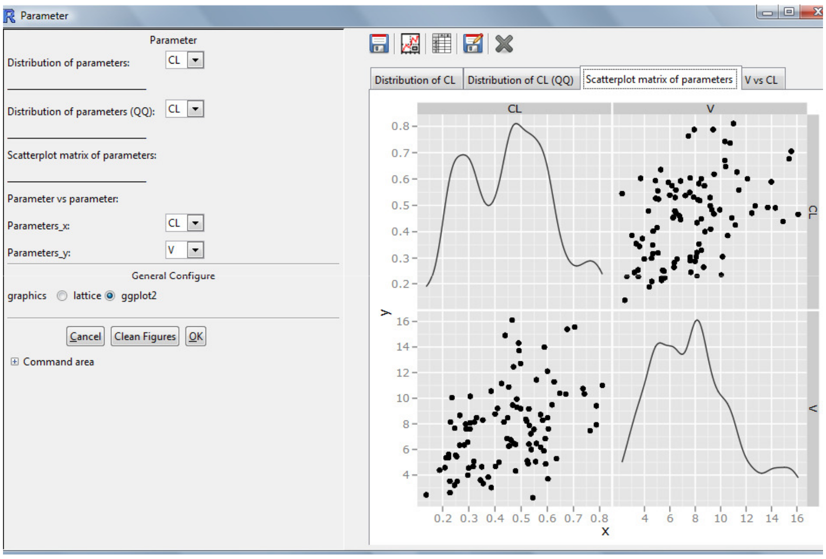
4.6.1 Influence analysis summary (PsN)

This function is specifically for PsN *cdd* results (Figure˜11). It takes two result files from PsN: *raw_results1.csv* and *skipped_individuals1.csv*, and generates a scatter plot for cov.raito versus cov.score.

4.6.2 Visualization for influence analysis

This function is to visualize data from case deletion diagnostics (multiple NONMEM runs). Let's use multiple NONMEM run form PsN (Figure˜12), and find file directory for these runs. Then we can select parameters as shown in Figure˜13. These parameters include:

- *Target directory path*: the path for multiple NONMEM runs. It is a required parameter.
- *Simulation folder pattern*: the common folder name for multiple NONMEM runs. For this example, it is *NM_run*. It is a required parameter.
- *NONMEM result file name*: the file name for NONEM fitted result. In each NONMEM run, there should be a file with this name. It is required parameter.
- *Patient ID*: the ID for each subject. It is a required parameter.
- *Plot variable*: the variable you use to detect difference among patients. For this example, we choose *CL*. It is a required parameter.



Parameter

Parameter

Distribution of parameters: CL

Distribution of parameters (QQ): CL

Scatterplot matrix of parameters:

Parameter vs parameter:

Parameters_x: CL

Parameters_y: V

General Configure

graphics ○ lattice ● ggplot2

Figure 9: *Parameters* in *PK models*

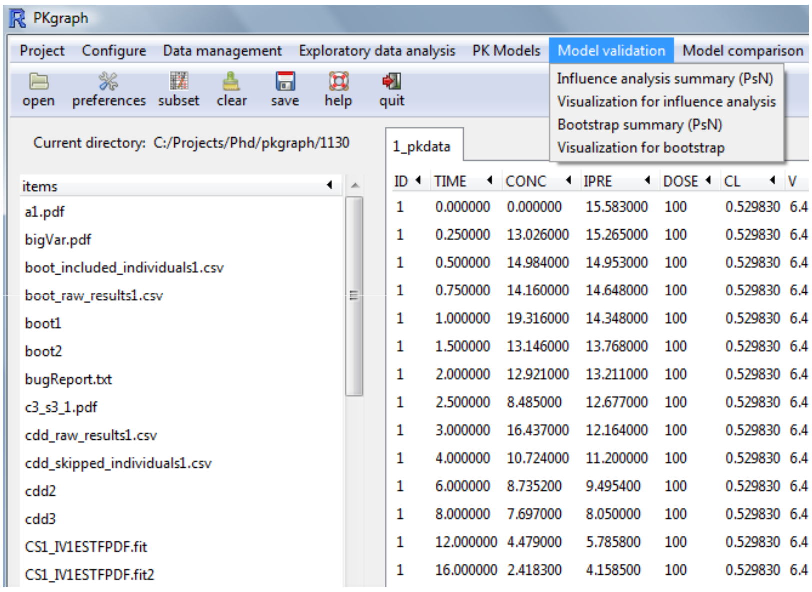
4.6.3 Bootstrap summary (PsN)

This function is specifically for PsN *boot* results (Figure˜14). It takes two result files from PsN: *raw_results1.csv* and *included_individuals1.csv*, and generates related plots.

4.6.4 Visualization for bootstrap

This function is to visualize data from boostrap (multiple NONMEM runs). Let's use multiple NONMEM run form PsN (Figure˜15), and find file directory for these runs. Then we can select parameters as shown in Figure˜16. These parameters include:

- *Target directory path*: the path for multiple NONMEM runs. It is a required parameter.
- *Bootstrap folder pattern*: the common name style for multiple NONMEM runs. For this example, it is *NM_run*. It is a required parameter.
- *NONMEM result file name*: the fit result for each NONMEM run. In this example, it is *CS1_IV1ESTFPDF-1.fit* . It is a required parameter.
- *Bootstrap key table path*: the path for bootstrap key file, which is file describing the sampling schema for patient IDs. It is a required parameter.



Current directory: C:/Projects/Phd/pkgraph/1130

items
a1.pdf
bigVar.pdf
boot_included_individuals1.csv
boot_raw_results1.csv
boot1
boot2
bugReport.txt
c3_s3_1.pdf
cdd_raw_results1.csv
cdd_skipped_individuals1.csv
cdd2
cdd3
CS1_IV1ESTFPDF.fit
CS1_IV1ESTFPDF.fit2

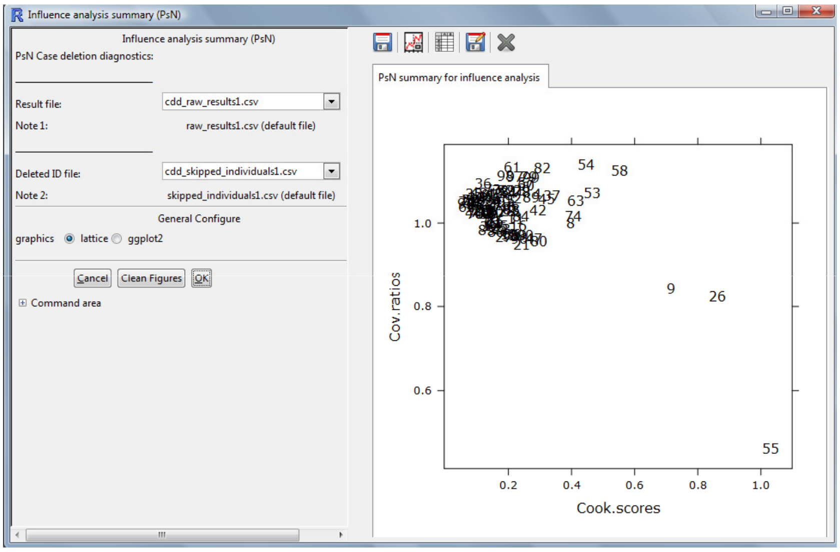
ID	TIME	CONC	IPRE	DOSE	CL	V
1	0.000000	0.000000	15.583000	100	0.529830	6.4
1	0.250000	13.026000	15.265000	100	0.529830	6.4
1	0.500000	14.984000	14.953000	100	0.529830	6.4
1	0.750000	14.160000	14.648000	100	0.529830	6.4
1	1.000000	19.316000	14.348000	100	0.529830	6.4
1	1.500000	13.146000	13.768000	100	0.529830	6.4
1	2.000000	12.921000	13.211000	100	0.529830	6.4
1	2.500000	8.485000	12.677000	100	0.529830	6.4
1	3.000000	16.437000	12.164000	100	0.529830	6.4
1	4.000000	10.724000	11.200000	100	0.529830	6.4
1	6.000000	8.735200	9.495400	100	0.529830	6.4
1	8.000000	7.697000	8.050000	100	0.529830	6.4
1	12.000000	4.479000	5.785800	100	0.529830	6.4
1	16.000000	2.418300	4.158500	100	0.529830	6.4

Figure 10: Menu items in *Model validation*

- *Bootstrap key table name*: The file describes the sampling schema for patient IDs In this example, it is *included_individuals1.csv*. It is a required parameter.
- *Patient ID*: the ID for each subject. It is a required parameter.
- *Plot variable*: the variable you use to detect difference among patients. For this example, we choose *CL*. It is a required parameter.
- *xlabel*: the name label for each NONMEM run. It is optional.

4.7 Model comparison

In this process, there are three main steps: 1) select datasets; 2) configure mapping; 3) comparison (Figure˜17). The first step is to select datasets for comparison. Currently the program only supports comparison of two models. Then users proceed to configure mapping by matching column names or variable names from two data sets. These matching variables are generally the variables from original data sets and they are not related to model fitting. When all parameters are set, the program offers three choices for comparison: “histogram comparison” (distribution comparison), “scatter plot comparison” and “transform comparison”.



Influence analysis summary (PsN)

PsN Case deletion diagnostics:

Result file: cdd_raw_results1.csv

Note 1: raw_results1.csv (default file)

Deleted ID file: cdd_skipped_individuals1.csv

Note 2: skipped_individuals1.csv (default file)

General Configure

graphics ● lattice ○ ggplot2

Figure 11: Influence analysis summary (PsN)

4.7.1 Select datasets

This function is to select datasets available in the PKgraph data area. Figure˜18 shows there are three data sets available, including fit result 2: 2_CS1_IV1ESTFPDF.fit (fit with additive error model) and fit result 3: 3_ CS1_IV1ESTFPDF.fit2 (proportional error model). In this example, we will compare these two models.

4.7.2 Configure mapping

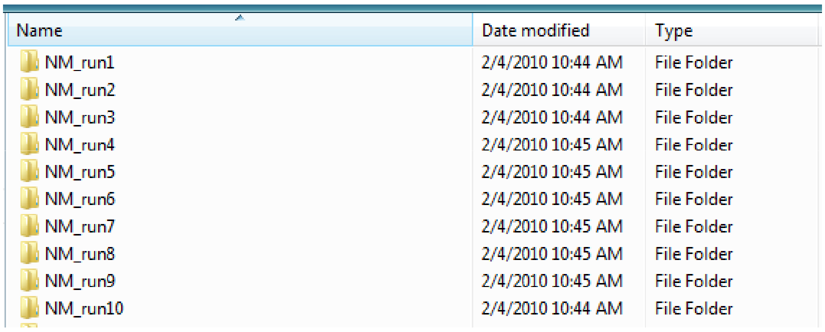
This step will join two fit results. As a result, users have to match the original data variables between two fit results. For example (Figure˜19),

- *Matching variables*: *ID, Time, Concentration, WT, AGE*, etc must be matched in this step. These variables do not change with different models.
- *Non-matching variables*: *RES, PRED, WRES*, etc are fit results, and should NOT be matched. These variables change with different models.

After mapping, a new dataset joining two fit results will show in data area of main interface.

4.7.3 Comparison

"histogram comparison" enables to compare distributions of matching parameters from two models. "scatter plot comparison" provides a environment to compare matching parameters by scatter plot. "transform comparison" transforms



Name	Date modified	Type
NM_run1	2/4/2010 10:44 AM	File Folder
NM_run2	2/4/2010 10:44 AM	File Folder
NM_run3	2/4/2010 10:44 AM	File Folder
NM_run4	2/4/2010 10:45 AM	File Folder
NM_run5	2/4/2010 10:45 AM	File Folder
NM_run6	2/4/2010 10:45 AM	File Folder
NM_run7	2/4/2010 10:45 AM	File Folder
NM_run8	2/4/2010 10:45 AM	File Folder
NM_run9	2/4/2010 10:45 AM	File Folder
NM_run10	2/4/2010 10:44 AM	File Folder

Figure 12: Multiple NONMEM runs for case deletion diagnostics

data by ratio or log ratio in order to visualize the difference between variables from two models. All these models can be linked directly to ggobi for interactive diagnostics by clicking second button in the tool bar area on the top right panel.

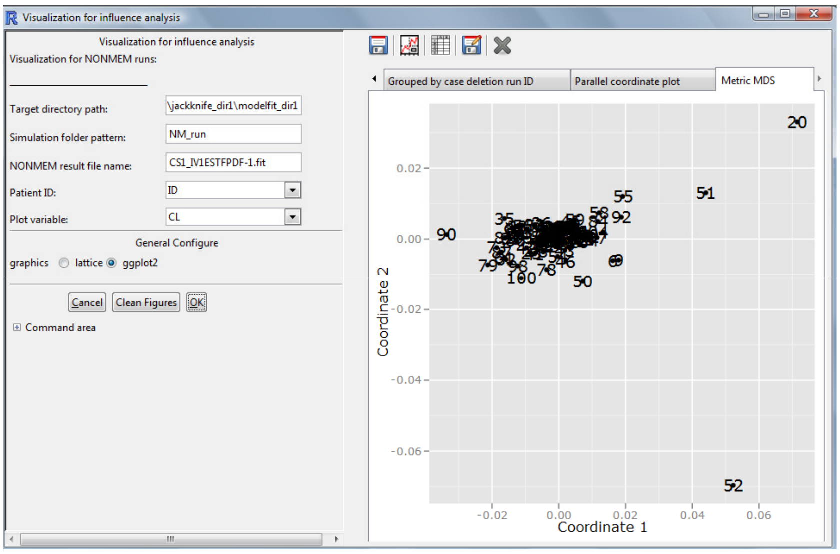
All variable names for model 1 will have additional ".x" label, and all variable names for model 2 will have additional ".y" label.

Let us look at "histogram comparison" as one example. First, we need to make sure that current data set is "4_ModelComparison" (Figure˜20); second, we click "histogram comparison". The result is shown in (Figure˜21) for comparing *CL*.

4.8 Interactive graphics

This functional module incorporates a unique feature: interactive graphics into every step of model diagnostics. It targets to link diverse data sets in one integrative platform. Users can have access to this feature through *ggobi* button in the graph interface. In addition, users have flexibility to apply this feature to achieve their specific goals. In the toolbar, there is option: *interactive graphics*, designed for this purpose. It includes three steps: select datasets; configure mapping; and diagnostics. By linking diverse data sets with a key variable, users can seek patterns by brushing, linking and diagnosing patterns conveniently.

In ggobi, the main operation for brushing data is *ctrl + b*. By moving the



Visualization for influence analysis

Visualization for NONMEM runs:

Target directory path: \jackknife_dir1\modelfit_dir1

Simulation folder pattern: NM_run

NONMEM result file name: CS1_IV1ESTFPDF-1.fit

Patient ID: ID

Plot variable: CL

General Configure

graphics ○ lattice ● ggplot2

Figure 13: Parameters and results for case deletion diagnostics

brushing rectangle, the users can select interesting subsets in ggobi. More information is available in http:
www.ggobi.org if needed.

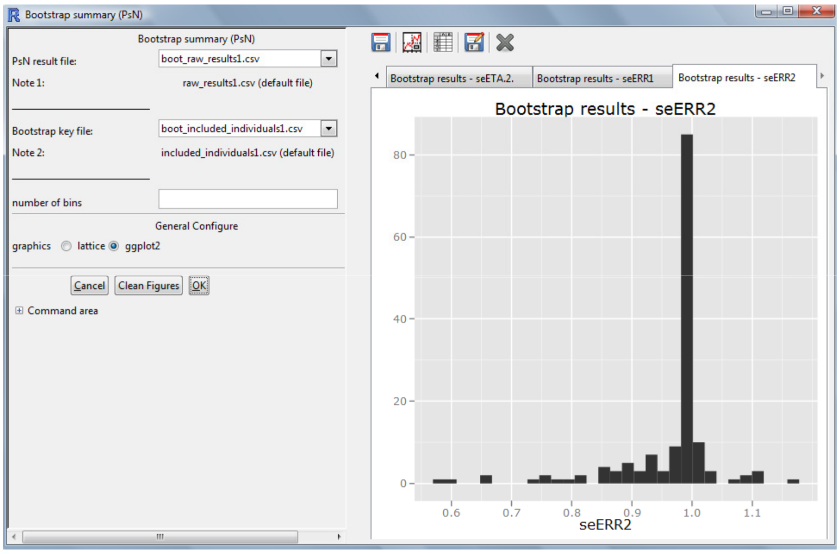
5 Example

One dataset from NONMEM is utilized to demonstrate PKgraph. This data set has 100 patients with covariates: ISM (gender), AGE, and WT. The data is fitted with one compartment model with zero order absorption and first order elimination.

As a text file, the fitting result from NONMEM is imported into PKgraph for further investigation and analysis. In the “open” dialog, we set up file format for reading with default parameters, and as a result, the input data shows up on the right panel while a message, “Data is loaded successfully” appears in the status bar at bottom of panel.

Alternatively, to make the input process flexible, users can input data into R first and then load data from “Data from R environment” in the “open” dialog. All the fitted results from a wide variety of software including NONMEM, SAS, etc can be loaded into this package.

```
> library(PKgraph)
> data(pkdata)
```



Bootstrap summary (PsN)

PsN result file: boot_raw_results1.csv

Note 1: raw_results1.csv (default file)

Bootstrap key file: boot_included_individuals1.csv

Note 2: included_individuals1.csv (default file)

number of bins

General Configure

graphics ○ lattice ● ggplot2

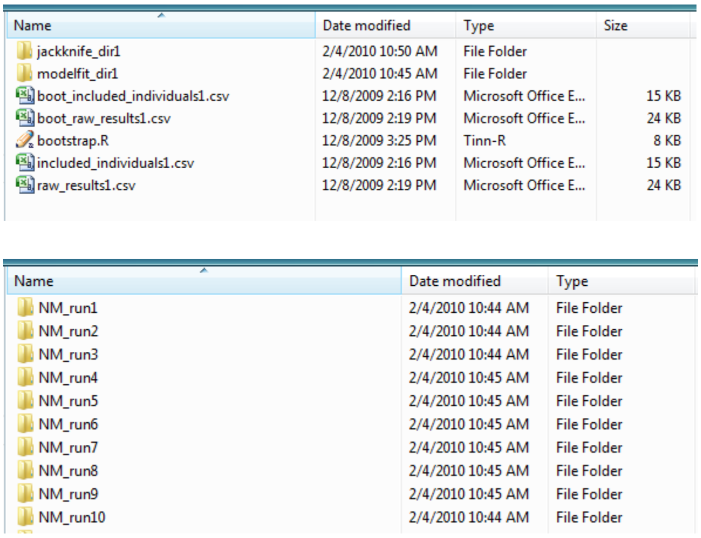
Cancel
Clean Figures
OK
Command area
Bootstrap results - seETA2.
Bootstrap results - seERR1
Bootstrap results - seERR2
Bootstrap results - seERR2
80
60
40
20
0
0.6
0.7
0.8
0.9
1.0
1.1
seERR2

Figure 14: Bootstrap summary (PsN)

```
> PKgraph()
```

To further explore data, first, we choose “Bivariates” from “Exploratory Data Analysis” located at menu bar to check the scatter plots of interested variables(Figure˜22, Figure˜23). The option “cond” from the functional model interface helps user to draw conditional plots to seek patterns for subgroups. Certainly, users can also select “ggplot2” graphic package with different taste of figure. Next, we can take advantage of interactive techniques to look at maximum concentration by clicking second image button on the right panel. This will start ggobi and load related data. GGobi includes two windows: console window and plot window. In order to link figures together, users need to open all interested figures by “Display” option in the menu bar. The following figure clearly shows that maximum concentration comes from male patients (value: 1). To look at these data in detail, we go back to the figure graphical user interface and click third image button to check selected data set in ggobi. The selected data set pops up and links to patient with ID: 55. We repeat the same procedure for other variables to check patterns.

Next, we utilize “PK model” option to check model assumptions and diagnose model fitting. The program provides default names such as ID, TIME, COV, etc in order to automatically generate diagnosing results. After we match data variables to the default names, we can proceed to automatically generate routine goodness of fit plots for interested models. Figure ˜24 is one of the results for structural model diagnostics.



Name	Date modified	Type	Size
jackknife_dir1	2/4/2010 10:50 AM	File Folder	
modelfit_dir1	2/4/2010 10:45 AM	File Folder	
boot_included_individuals1.csv	12/8/2009 2:16 PM	Microsoft Office E...	15 KB
boot_raw_results1.csv	12/8/2009 2:19 PM	Microsoft Office E...	24 KB
bootstrap.R	12/8/2009 3:25 PM	Tinn-R	8 KB
included_individuals1.csv	12/8/2009 2:16 PM	Microsoft Office E...	15 KB
raw_results1.csv	12/8/2009 2:19 PM	Microsoft Office E...	24 KB

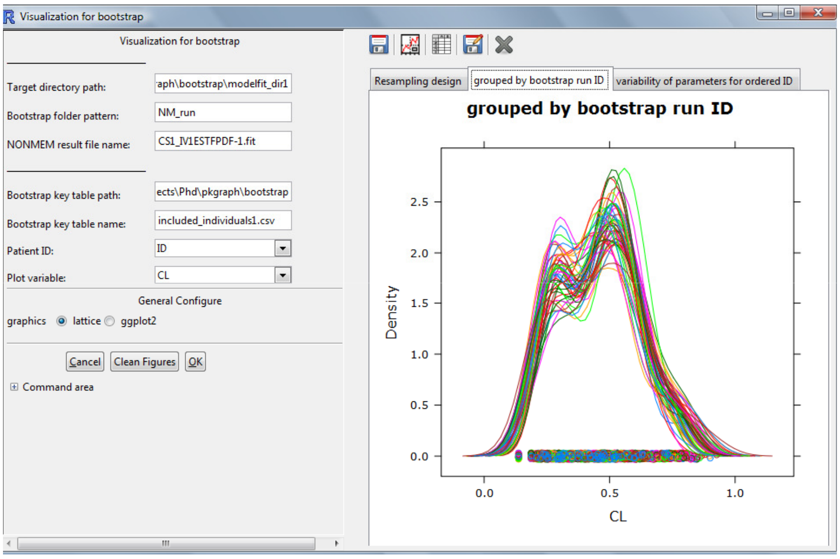
Name	Date modified	Type
NM_run1	2/4/2010 10:44 AM	File Folder
NM_run2	2/4/2010 10:44 AM	File Folder
NM_run3	2/4/2010 10:44 AM	File Folder
NM_run4	2/4/2010 10:45 AM	File Folder
NM_run5	2/4/2010 10:45 AM	File Folder
NM_run6	2/4/2010 10:45 AM	File Folder
NM_run7	2/4/2010 10:45 AM	File Folder
NM_run8	2/4/2010 10:45 AM	File Folder
NM_run9	2/4/2010 10:45 AM	File Folder
NM_run10	2/4/2010 10:44 AM	File Folder

Figure 15: Multiple NONMEM runs for bootstrap

To further look at the influential cases from same data set, we can link them together by "model validation" option in menu bar. In this process, we have 100 NONMEM runs available at directory: C:\ Projects\modelfit_dir1 using PsN function: cdd. Let's input the path of these NONM runs, and select plot variable as "CL". After clicking "OK", we will have the parallel coordinates plot showing the CL variables for all NONMEM runs. From Figure ˜25, we can see some patients have more influential effects on CL when records from these patients are deleted.

Let's identify these influential cases with interactive graphics. Figure ˜26 clearly demonstrates that these influential cases come from patient 52 and 20 based on multidimensional scaling and parallel coordinate plots.

In addition, we compare additive error model (2_CS1_IV1ESTFPDF.fit) with proportional error model (3_ CS1_IV1ESTFPDF.fit2) by "model comparison" function in the menu bar. By comparing the distribution of two models, Figure ˜27 does not find significant difference between two models for CL. In addition, using gender as a conditional variable, we found first model always gave a higher peak value for both male and female.


Visualization for bootstrap
Resampling design
grouped by bootstrap run ID
variability of parameters for ordered ID
grouped by bootstrap run ID
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
CL
General Configure
Cancel
Clean Figures
OK
⊞ Command area

Visualization for bootstrap

Target directory path: aph\bootstrap\modelfit_dir1

Bootstrap folder pattern: NM_run

NONMEM result file name: CS1_IV1ESTFPDF-1.fit

Bootstrap key table path: ects\Phd\pkgraph\bootstrap

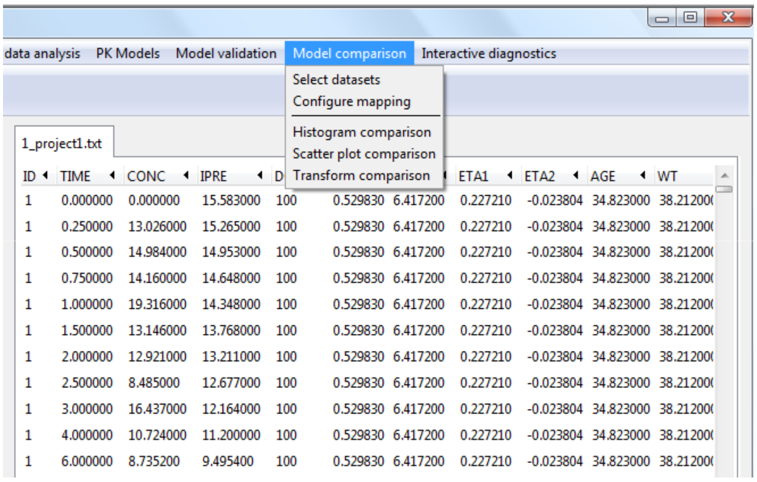
Bootstrap key table name: included_individuals1.csv

Patient ID: ID

Plot variable: CL

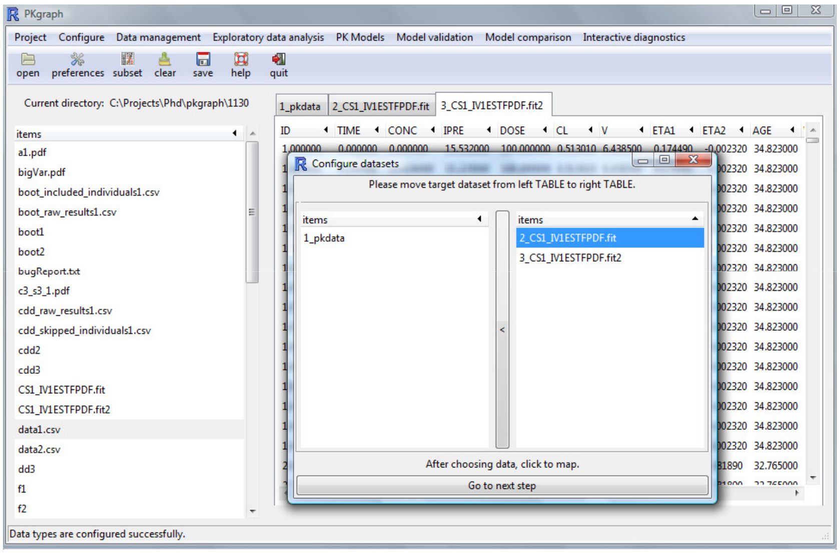
graphics ● lattice ○ ggplot2

Figure 16: Parameters and results for bootstrap visualization



ID	TIME	CONC	IPRE	D			ETA1	ETA2	AGE	WT
1	0.000000	0.000000	15.583000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.250000	13.026000	15.265000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.500000	14.984000	14.953000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	0.750000	14.160000	14.648000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	1.000000	19.316000	14.348000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	1.500000	13.146000	13.768000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	2.000000	12.921000	13.211000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	2.500000	8.485000	12.677000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	3.000000	16.437000	12.164000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	4.000000	10.724000	11.200000	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200
1	6.000000	8.735200	9.495400	100	0.529830	6.417200	0.227210	-0.023804	34.823000	38.21200

Figure 17: Menu items in *Model comparison*



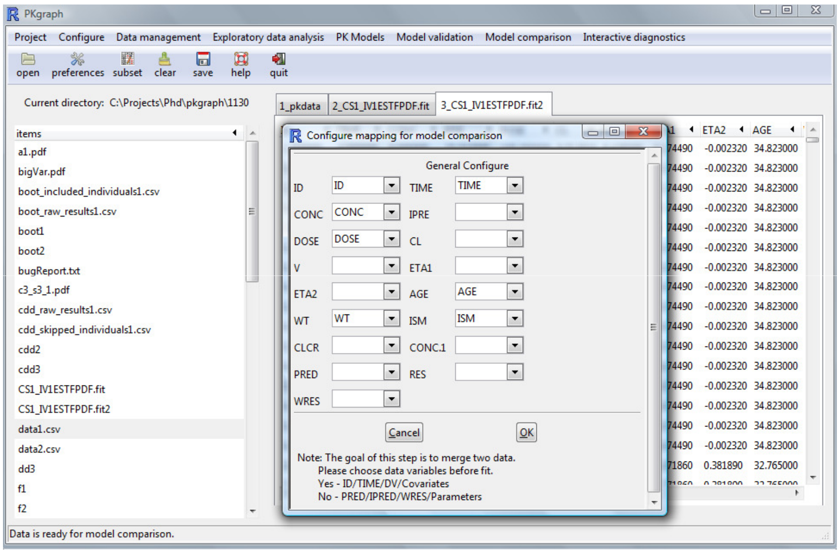
Current directory: C:\Projects\Phd\pkgraph\1130

Please move target dataset from left TABLE to right TABLE.

After choosing data, click to map.

Data types are configured successfully.

Figure 18: *Select datasets* in *Model comparison*



Current directory: C:\Projects\Phd\pkgraph\1130

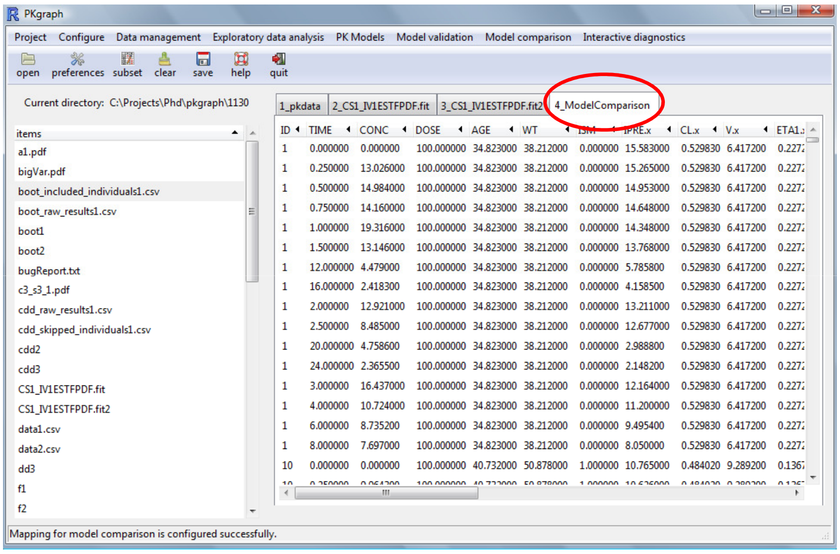
General Configure

ID	ID	TIME	TIME
CONC	CONC	IPRE	
DOSE	DOSE	CL	
V		ETA1	
ETA2		AGE	AGE
WT	WT	ISM	ISM
CLCR		CONC.1	
PRED		RES	
WRES			

Note: The goal of this step is to merge two data.
Please choose data variables before fit.
Yes - ID/TIME/DV/Covariates
No - PRED/IPRED/WRES/Parameters

Data is ready for model comparison.

Figure 19: *Configure mapping* in *Model comparison*

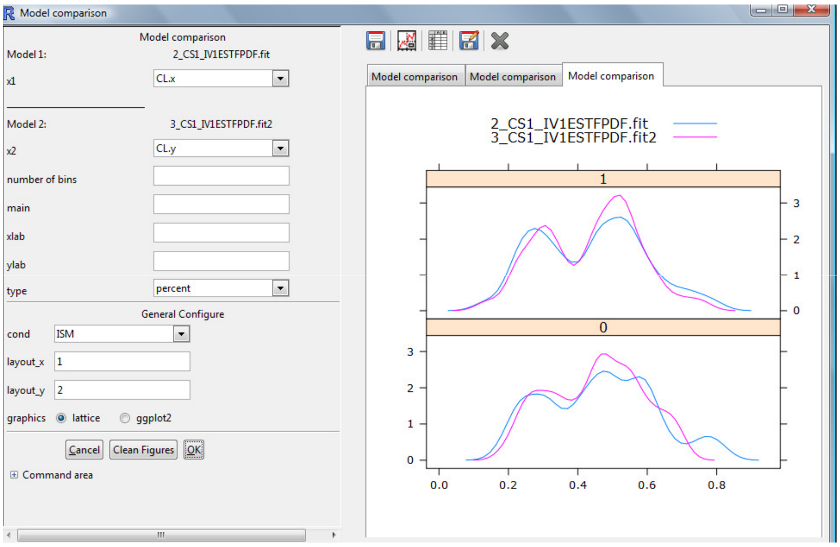


Current directory: C:\Projects\Phd\pkgraph\1130

ID	TIME	CONC	DOSE	AGE	WT	ISM	IPRE.x	CL.x	V.x	ETA1.x
1	0.000000	0.000000	100.000000	34.823000	38.212000	0.000000	15.583000	0.529830	6.417200	0.227
1	0.250000	13.026000	100.000000	34.823000	38.212000	0.000000	15.265000	0.529830	6.417200	0.227
1	0.500000	14.984000	100.000000	34.823000	38.212000	0.000000	14.953000	0.529830	6.417200	0.227
1	0.750000	14.160000	100.000000	34.823000	38.212000	0.000000	14.648000	0.529830	6.417200	0.227
1	1.000000	19.316000	100.000000	34.823000	38.212000	0.000000	14.348000	0.529830	6.417200	0.227
1	1.500000	13.146000	100.000000	34.823000	38.212000	0.000000	13.768000	0.529830	6.417200	0.227
1	12.000000	4.479000	100.000000	34.823000	38.212000	0.000000	5.785800	0.529830	6.417200	0.227
1	16.000000	2.418300	100.000000	34.823000	38.212000	0.000000	4.158500	0.529830	6.417200	0.227
1	2.000000	12.921000	100.000000	34.823000	38.212000	0.000000	13.211000	0.529830	6.417200	0.227
1	2.500000	8.485000	100.000000	34.823000	38.212000	0.000000	12.677000	0.529830	6.417200	0.227
1	20.000000	4.758600	100.000000	34.823000	38.212000	0.000000	2.988800	0.529830	6.417200	0.227
1	24.000000	2.365500	100.000000	34.823000	38.212000	0.000000	2.148200	0.529830	6.417200	0.227
1	3.000000	16.437000	100.000000	34.823000	38.212000	0.000000	12.164000	0.529830	6.417200	0.227
1	4.000000	10.724000	100.000000	34.823000	38.212000	0.000000	11.200000	0.529830	6.417200	0.227
1	6.000000	8.735200	100.000000	34.823000	38.212000	0.000000	9.495400	0.529830	6.417200	0.227
1	8.000000	7.697000	100.000000	34.823000	38.212000	0.000000	8.050000	0.529830	6.417200	0.227
10	0.000000	0.000000	100.000000	40.732000	50.878000	1.000000	10.765000	0.484020	9.289200	0.136
10	0.250000	0.064200	100.000000	40.732000	50.878000	1.000000	10.626000	0.484020	9.289200	0.136

Mapping for model comparison is configured successfully.

Figure 20: Current data set for *Model comparison*



Model comparison

Model 1: 2_CS1_IV1ESTFPDF.fit

x1: CL.x

Model 2: 3_CS1_IV1ESTFPDF.fit2

x2: CL.y

number of bins:

main:

xlab:

ylab:

type: percent

General Configure

cond: ISM

layout_x: 1

layout_y: 2

graphics: ● lattice ○ ggplot2

Figure 21: *histogram comparison* for *Model comparison*

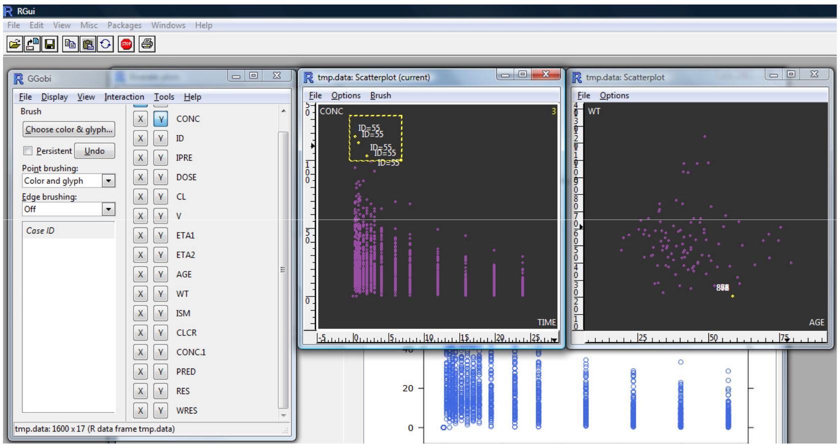
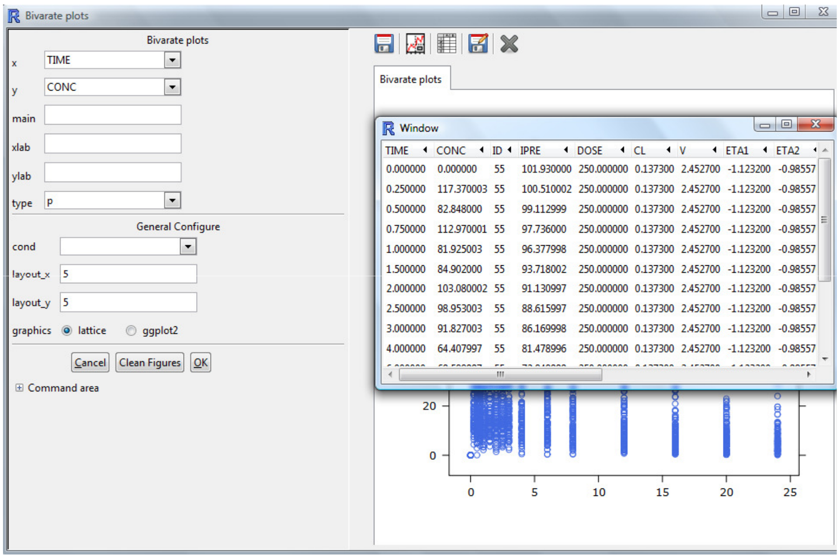


Figure 22: Exploratory data analysis. Peak is identified with brushing. This patient is from light weight and middle age group.



Bivarate plots

x: TIME

y: CONC

main:

xlab:

ylab:

type: p

General Configure

cond:

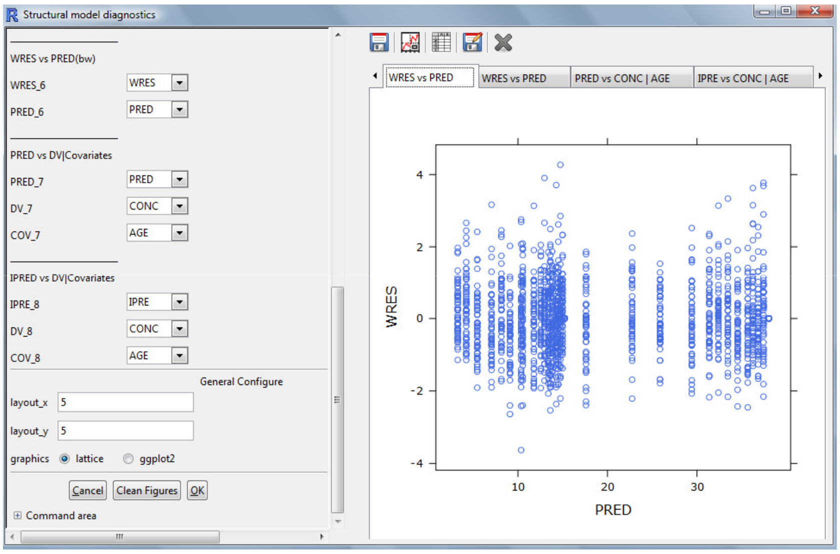
layout_x: 5

layout_y: 5

graphics: ● lattice ○ ggplot2

TIME	CONC	ID	IPRE	DOSE	CL	V	ETA1	ETA2
0.000000	0.000000	55	101.930000	250.000000	0.137300	2.452700	-1.123200	-0.98557
0.250000	117.370003	55	100.510002	250.000000	0.137300	2.452700	-1.123200	-0.98557
0.500000	82.848000	55	99.112999	250.000000	0.137300	2.452700	-1.123200	-0.98557
0.750000	112.970001	55	97.736000	250.000000	0.137300	2.452700	-1.123200	-0.98557
1.000000	81.925003	55	96.377998	250.000000	0.137300	2.452700	-1.123200	-0.98557
1.500000	84.902000	55	93.718002	250.000000	0.137300	2.452700	-1.123200	-0.98557
2.000000	103.080002	55	91.130997	250.000000	0.137300	2.452700	-1.123200	-0.98557
2.500000	98.953003	55	88.615997	250.000000	0.137300	2.452700	-1.123200	-0.98557
3.000000	91.827003	55	86.169998	250.000000	0.137300	2.452700	-1.123200	-0.98557
4.000000	64.407997	55	81.478996	250.000000	0.137300	2.452700	-1.123200	-0.98557

Figure 23: Exploratory data analysis. The detailed information for this patient is selected for investigation.



Structural model diagnostics

WRES vs PRED(bw)

WRES_6 WRES

PRED_6 PRED

PRED vs DV|Covariates

PRED_7 PRED

DV_7 CONC

COV_7 AGE

IPRED vs DV|Covariates

IPRE_8 IPRE

DV_8 CONC

COV_8 AGE

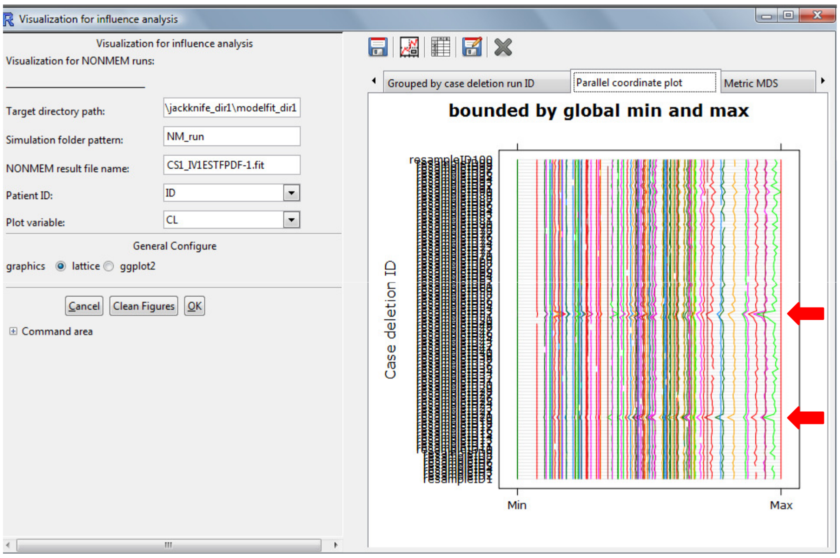
General Configure

layout_x 5

layout_y 5

graphics ● lattice ○ ggplot2

Figure 24: Structural model diagnostics.



Visualization for influence analysis

Visualization for influence analysis

Visualization for NONMEM runs:

Target directory path: \jackknife_dir1\modelfit_dir1

Simulation folder pattern: NM_run

NONMEM result file name: CS1_IV1ESTFPDF-1.fit

Patient ID: ID

Plot variable: CL

General Configure

graphics ● lattice ○ ggplot2

Figure 25: Influence analysis

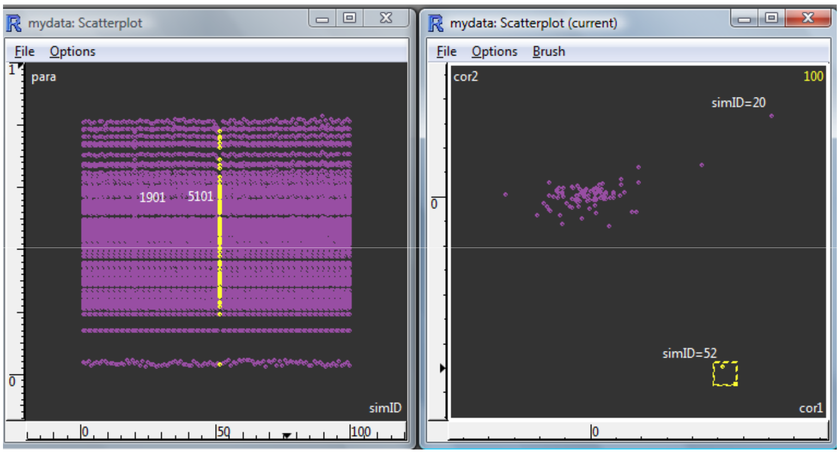
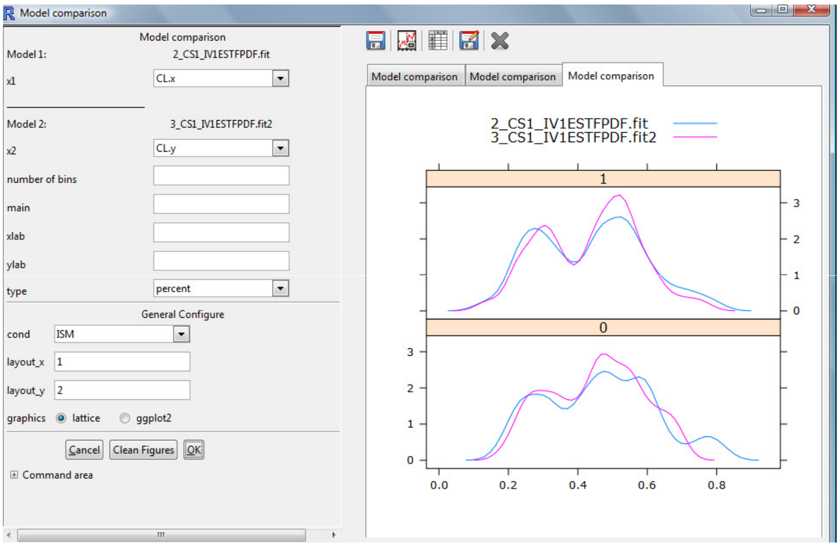


Figure 26: Influence analysis: linking results from multidimensional scaling and parallel coordinate plots.



Model comparison

Model comparison

Model 1: 2_CS1_IV1ESTFPDF.fit

x1: CL.x

Model 2: 3_CS1_IV1ESTFPDF.fit2

x2: CL.y

number of bins:

main:

xlab:

ylab:

type: percent

General Configure

cond: ISM

layout_x: 1

layout_y: 2

graphics: ● lattice ○ ggplot2

Figure 27: Histogram comparison for comparing distributions of CL from two models.