Xpose 2.0 *User's Manual*

Niclas Jonsson   Mats Karlsson

Department of Pharmacy,
Uppsala University, Sweden

Draft 4
Mar 12, 1998

Contents

1 Introduction

Population analysis using non-linear mixed effects models has become an important tool in the analysis of pharmacokinetic/pharmacodynamic data [1] and the benefits of this type of analysis over the traditional alternatives have to some length been discussed in the literature e.g. [2, 3]. The price one has to pay for these benefits is an increased complexity of the models that are fitted to the data and a corresponding increase in the number of assumptions that has to be made [4]. Assumptions of e.g. independence of residuals and homogeneity of residual variance is of course also a part of model dependent *individual* analysis but the challenge of non-linear mixed effects models lies in the hierarchical nature of the model, i.e. we do not only have the differences between the observed and predicted observations to take in to account, we also have to take into account that individuals may differ, which results in a second type of residuals — the difference between individual parameter values and what the model predicts for the corresponding typical individual in the population. This together with the possibility to include the relationships between demographic data, for example age, weight and clinical laboratory measurements, and parameters, which in a sense is a regression within the regression, makes the task of assessing the goodness of fit of a non-linear mixed effects model a lot harder than for an individual specific model.

Xpose is an S-PLUS based model building aid for population analysis using NON-MEM. It facilitates data set checkout, exploration and visualization, model diagnostics, candidate covariate identification and model comparison. Data set checkout includes visualization of the observed variable(s), covariates and plots to reveal errors in the data file. Model diagnostic plots includes the usual residual plots but also plots to check the validity of assumptions specific to non-linear mixed effects models. Data exploration is also done by various plots but also includes auxiliary screening analyses such as stepwise generalized additive modeling (GAM) and tree based modeling. The stability of the GAM results with respect to covariate model selection as well as the impact of influential individuals and certain types of covariate interactions can be explored using a bootstrap re-sampling procedure. To facilitate documentation, Xpose can produce run summaries and run records. The run records are one page summaries of a run consisting of a combination of goodness of fit plots, parameter estimates and the model file. Run records are tabulated summaries of many runs consisting of user comments, termination messages and objective function values.

The basic idea is that the user should only have to tell Xpose what run that is to be processed to be able to produce anything between simple numerical summaries of, e.g. covariates, to sophisticated bootstrap of the GAM analyses. The only other thing the user has to do is to make NONMEM produce one or more table files that contains the items, for example residuals and predictions, that is to be displayed or used in the Xpose analyses. In other words, Xpose takes care of reading the data into S-PLUS, reformatting of the data to suit each plot and analysis specific requirements and to do anything else that is necessary to accomplish what the user selects from the menus.

1.1 Terms of usage

Any person who wants to use Xpose should be aware of the following:

The users of Xpose are encouraged to report if they think that Xpose is behaving strangely or if they find a bug. Even if we can not take on any official support responsibilities, we are interested in having a useful and bug free program. If we are not aware of the bugs we can't do anything about them... We are also interested in suggestions and comments but again we cannot promise to implement things because one or a few users requests it.

The reason we want people to obtain Xpose from `http://www.biof.uu.se/Xpose` is that we want to know, at least roughly, how many users of Xpose there is but it is also for your own sake. S-PLUS source code is very easy to alter and if your friendly colleague next door offers you a copy of Xpose you can not be sure that it is identical to the original code. Anyway, you may use Xpose freely at your own risk and we do not take any responsibility for how Xpose is used or misused.

1.2 Changes between Xpose 1.1 and Xpose 2.0

The changes between Xpose version 2.0 and the previous version (1.1) are numerous. The main user visible difference is that Xpose 2.0 uses the Trellis Graphics^{TM} library instead of the core graphics of S-PLUS. Another change is that it is no longer strictly necessary with separate table files for, e.g. covariates and parameters. New is also the possibility to save plots, after optional customization of the plot title and axis labels, to a file suitable for inclusion in other programs such as word processors. There are also a large number of new plots specifically designed to check certain assumptions made in non-linear mixed effects modeling as well as diagnostics for GAM models, including a new Bootstrap of the GAM algorithm. Xpose 2.0 is also a lot more flexible than the older version. To make the transition easier for users of the previous version, major differences, that might be confusing to someone expecting Xpose 1.1 behavior, will be pointed out in footnotes.

1.3 About this manual

This manual should be viewed as a beginners guide to Xpose and should, together with the installation instructions provided with the source code, be enough to get a new user going. It will be assumed that Xpose is installed on the system and that the Xpose 2.0 library is in the S-PLUS search path (See the installation instructions for details.) We also assume that the user knows how to interpret the usual goodness of fit plots, i.e. we will not explain all the plots available in Xpose. Instead we will concentrate on the areas that might not be intuitive and we will also give references to the literature where appropriate.

1.4 About the distribution

The Xpose 2.0 distribution contains the source code, the Xpose 2.0 *User's Manual*, installation instructions and example files. The installation of the source code is described in the installation instructions. The *User's Manual* is this document and is distributed in both a postscript and HTML version. Details about the documentation is given in the README file in the `Doc` sub-directory. The example files contains three runs, complete with data file, model files, output files and table files. The details of the examples are given in the README file in the `Examples` sub-directory.

1.5 Acknowledgment

We would like to thank Dr. Lewis Sheiner for the impetus to Xpose. We have used parts of the S-PLUS code provided in the course folder for the "Intermediate Workshop in Population Pharmacokinetic Data Analysis Using the NONMEM System" given by Dr. Lewis Sheiner and Dr. Stuart Beal. Dr. Janet Wade wrote some of the in house S-PLUS scripts we used before the development Xpose and on which some of the plots in Xpose are based. We would also like to thank Dr. Stuart Beal for making available a version of NONMEM with some advanced features. (Please note that this version is not yet generally available, neither is it required for using Xpose.) We would also like to thank the participants in the S-news mailing list, especially Dr. Venables, for many helpful discussions and code examples. We are also grateful for the helpful comments and suggestions sent by Xpose 1.1 users and Xpose 2.0 beta testers.

2 Producing input data for Xpose

Xpose most often uses the contents of one or more NONMEM table files when producing plots and analyses. For the Run summary and Run record Xpose also needs the NONMEM output files and model files. To check a data set the data file is needed. For more details about how Xpose reads model files, output files and data files see Sections 7.1 and 8.

2.1 Creating the table file(s) for Xpose to read

Xpose needs at least one NONMEM table file for input. If present in the current directory, Xpose will also read other table files. These table files should be named in a specific way. The table file names are divided up in two parts. The first part is the *stem*, for example the standard table file has the stem sdtab, and the second part is the *run number*. The standard table file for run number one should therefore be named sdtab1.[1] The other table files that Xpose reads, if present in the current directory and if they match the run number, are patab, cotab, catab, mutab and mytab. Xpose expects these table files to be produced with the NOPRINT and ONEHEADER options on the NONMEM $TABLE record and with the POSTHOC option on the $ESTIMATION record. For example:

```
$EST POSTHOC
$TABLE ID   TIME IPRED IWRES
       NOPRINT ONEHEADER FILE=sdtab1
```

2.1.1 The sdtab

sdtab stands for standard table file. It should contain items describing the overall goodness of fit. The recommended column items and order for this table file is ID, TIME[2], IPRED and IWRES (in addition to these NONMEM by default adds DV, PRED, RES and WRES). IPRED (individual predictions) and IWRES (individual weighted residuals) are not NONMEM defined items and has to be defined by the user. These can be obtained by the following code in the $ERROR block[3]:

```
$ERROR
   IPRED = F
   W     =   ; Your choice: 1           = additive error model
             ;              F           = constant CV error model
             ;              F**THETA(.) = power error model
             ;              (F**2+THETA(.)**2)**0.5
             ;                          = additive plus
             ;                            proportional
             ;                            error model
   IRES  = DV-IPRED
   IWRES = IRES/W
   Y     = IPRED + W*EPS(1)
```

Note that the IWRES = IRES/W line will, in some cases, need reformulation to avoid division by zero.

---

[1] It is possible to alter the table file name structure by the addition of a *table file name suffix*, e.g. .txt, making Xpose look for table files ending in .txt, for example sdtab1.txt. See the installation instructions for details.

[2] Or whatever independent variable that is used in the NONMEM analysis

[3] The code is taken from the notes of the NONMEM intermediate workshop given by Stuart Beal and Lewis Sheiner. It is also somewhat different from the corresponding code in the Xpose 1.1 manual.

2.1.2 The patab

`patab` stands for parameter table file. It should contain the parameter estimates of the fit. The recommended column order is `ID` followed by the parameters of interest, for example the individual `CL` estimates, the ηs and/or the typical individual parameter estimates.

2.1.3 The cotab and catab

`cotab` and `catab` stands for continuous and categorical table file respectively. The table files should contain the covariates, divided up into continuous (`cotab`) and categorical (`catab`) covariates. The reason for this division is that they are treated differently in plots and analyses, e.g. the GAM. The recommended column order is `ID` followed by the covariates.

2.1.4 The mutab

`mutab` stands for multiple response variables table file. In a simultaneous fit of two different types of measurements it is usually a good idea to make, for example, goodness of fit plots for the two measurements separately. For Xpose to be able to differentiate between the two, it is necessary with a flag variable that will be, for example, 1 for pharmacokinetic observations and 2 for pharmacodynamic observations. The `mutab` was in Xpose version 1.1 used to provide such a flag variable. In Xpose 2.0 it is not necessary to provide the flag variable in this way (see Section 15) but `mutab` is kept for compatibility reasons. The recommended column order is `ID`, `FLAG`, `TIME`, `IPRED` and `IWRES`.

2.1.5 The mytab

`mytab` is new to Xpose 2.0 and can be used for items not fitting into the table files described above, for example variables created within NONMEM. Since this table file is completely user defined there is no recommended column order.

2.2 The extra file

The table files described in the previous section are limited to the variables available within NONMEM. NONMEM has a limit of 20 data columns so if the data set to be analyzed has more columns than that it is not possible to access all, e.g., covariates in Xpose. To solve this problem, Xpose can read an extra file (if present in the current directory), called `extra` followed by the run number (for example `extra1`). The format of this file should follow the format of NONMEM table files, i.e. the column headers on the second line and no alphabetic characters below this line, except the `E` or `e` in exponents (the period, `.`, that can be used in NM-TRAN data files does not work, neither do dates or clock times with a colon separating hour and minutes). It is also important that the number of lines in the `extra` file is the same as in the NONMEM table files. Care should also be taken when selecting the column names, see Section 2.4.

2.3 The Xpose data bases

When the user starts Xpose (see Section 3) it prompts for a run number to process. This number will be the current run number until another run number is specified (see Section 5). Each run number is associated with an *Xpose data base* containing the data from the matching table files and extra file. This data base is just a regular S-PLUS data frame with a few add-ons (see Section 2.5) and is called xpdb followed by the run number. The Xpose data base is stored in the current S-PLUS working directory and is hence accessible in later S-PLUS or Xpose sessions[4]. The Xpose data base matching the current run number is called the *current data base*.

When the data is read (see next section) the user are asked for (optional) documentation of the current data base. This can be thought of as a reminder note to identify a certain run without having to consult any other run documetation.

```
Type any documentation for the new data base and finish with
a blank line:
1:Test of documentation feature 2:
```

2.4 How Xpose reads files

When the user specifies a run number, Xpose first checks to see if there is a matching Xpose data base available in the current S-PLUS working directory. If there is, the user is asked if he/she wants to use the already existing data base or if the data base should be recreated from the table files and extra file. If there is no matching data base the user is asked if he/she wants to create it.

When Xpose is told to create a data base, the relevant files are read in the following order (if they exists in the current directory): sdtab, mutab, patab, catab, cotab, mytab and extra. The data from each of these files are put together in one large data frame and columns with duplicated names are deleted[5]. When there are no duplicated items in the current data base, Xpose deletes all lines where the column named WRES has an entry of zero (i.e. keeps all lines with observations) unless told otherwise (see Section 5). The next step is to determine what each column is, or rather the other way around, Xpose tries to define the *Xpose data variables*. This process is described in detail in the next section. When the Xpose data variables are defined the data frame is given the S-PLUS class attribute xpose and is saved to disk.

2.5 Xpose data variables

When Xpose makes a plot of, say, WRES vs PRED it takes the column in the current data base defined to be the *wres*-variable and plots that against the column defined to be the *pred*-variable. The *wres*- and the *pred*-variables are examples of *Xpose data*

---

[4]This is a major difference from Xpose 1.1 were the table files were read each time a plot or analyses was selected from the Xpose menus

[5]The first occurance of a duplicated column is kept, which means that if the extra file contains a column with the same name as any of the columns in the table files, that column will be deleted regardless of if the actual numbers in the columns are the same. With the table files this is not a problem since in NONMEM two different variables can not have the same name.

variables. All plots and analyses available in Xpose are defined in terms of xpose data variables. Since it is possible to redefine these variables in Xpose it means that it is quite easy to change the data that is plotted, e.g. concentrate on one parameter at a time, even if we put all parameters in the `patab`[6].

The default Xpose data variable assignment is done when the table files are read (see previous Section) and is based on the position of the columns in the table files and the column names as described in Table 1.

2.5.1 Xpose data variables without default settings

There are a few Xpose data variables that do not have a default setting. These are the definitions of typical parameter values (*tvpar*), random effects (*ranpar*) and weights (*weight*).

As soon as there are covariates included in the mixed-effects model the typical values of the parameters can be different from one individual to another. To assess the appropriateness of the inter-individual error model (the η model) one can plot the random effects (i.e. ηs) vs the corresponding typical parameter values [4]. To be able to do that in Xpose it is necessary to define the typical parameter values and the corresponding random effects.

If the *weight* variable is defined it will be used in the GAM and the bootstrap of the GAM analyses as prior weights (see Sections 10.5.5 and 11.5).

2.6 Using only one table file for all data items

The reason for the rather strict organization of the table files is that it helps Xpose to decide which column is what. As indicated previously it is not necessary to use the recommended table files nor the recommended table file order since it is possible to (re)define all the Xpose data variables within Xpose. This is a great ease for NON-MEM IV users which no longer needs to use multiple `$PROBLEM` statements to make NONMEM produce more than one table file in a single run. It is recommended that either the `sdtab`, the `mutab` or the `mytab` is used if only one table file is produced. See also Section 5.

3 Starting and quitting Xpose

Start S-PLUS in the directory where you have the NONMEM table files (see section 2.1 and the installation instructions). At the S-PLUS prompt type `xpose2()`. Xpose will start and a graphics window will be opened. Type the number of the run to be processed. If there already is an Xpose data base present that matches the run number, Xpose will ask if you want to use it or recreate the Xpose data base from the table files.

---

[6]This is one of the major changes in Xpose between version 2.0 and 1.1. In 1.1 we would have to look at plots of all parameters defined in the `patab` and the only way available to change what was plotted was to re-arrange the table files

Table 1: Description of the default Xpose data variable assignments. The variables appear in the order they are defined when Xpose creates a data base.

Xpose data variable	Xpose usage	Default definition[a]
id	Used to identify individuals	1. A column with the name ID 2. The first column in the first existing file in input file order[b]
idlab	Used to label data individuals	Same as the *id*-variable
dv	Use in plots with with the dependent variable	1. A column with the name DV 2. The fourth column from the end in the first existing file in input file order[b]
res	Not used in the default set-up of plots	1. A column with the name RES 2. The second column from the end in the first existing file in input file order[b]
wres	Used in plots with weighted residuals	1. A column with the name WRES 2. The last column in the first existing file in input file order[b]
pred	Used in plots with population predictions	1. A column with the name PRED 2. The third column from the end in the first existing file in input file order[b]
idv	Used in plots with the independent variable	1. A column with the name TIME 2. The second column in sdtab 3. The third column in mutab
ipred	Used in plots with the individual predictions	1. A column with the name IPRE 2. The third column in sdtab 3. The fourth column in mutab
iwres	Used in plots with the individual weighted residuals	1. A column with the name IWRE 2. The fourth column in sdtab or fifth column in mutab
flag	Used to make plots of subsets of the data[c]	The second column in mutab
curflag	Indicates the level of the flag to be processed[c]	Not defined
occ	Not used in the the default set-up of plots	Not defined
prameters	Used in plots and analyses of parameters	Columns two from the beginning to the fifth from the end in patab
covariates	Used in plots and analyses of covariates	Columns two from the beginning to the fifth from the end in cotab and catab[d]
miss	Used to indicate missing data[e]	Default set to -99

[a] Alternative are numbered in the order they are tried

[b] The input file order is sdtab, mutab, patab, catab, cotab, mytab and extra

[c] See Section 15

[d] Covariates found in catab will be coded as factors in S-PLUS

[e] See Section 6.5.1
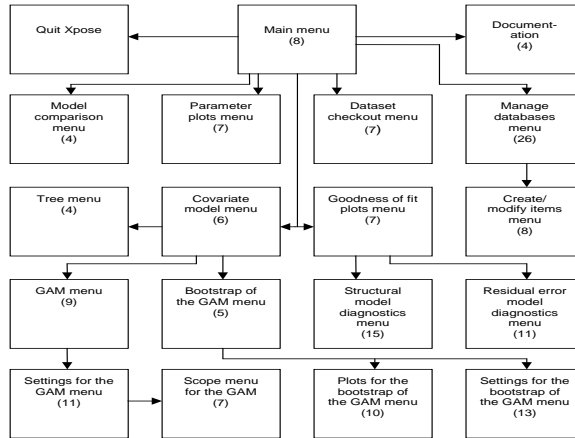
Figure 1: The Xpose road map. Each box, except the Quit option, is a menu. The number of options in each menu is indicated by the number in parenthesis.

To quit Xpose select option 8 from the MAIN MENU (See Section 4). Xpose asks if you just want to quit Xpose or to quit both Xpose and S-PLUS.

4   Moving around in Xpose

The user interface is a system of text based menus with up to 4 levels and 167 menu option (Figure 1).

Below is the MAIN MENU:

```
MAIN MENU

1: Documentation
2: Manage databases
3: Data checkout
4: Goodness of fit plots
5: Parameters
6: Covariate model
7: Model comparison
8: Quit
```

```
Selection:
```

Typing a number at the `Selection:` prompt selects the corresponding option from the menu. Some options leads to new menus while others lead to a plot or an analysis. In all menus except the `MAIN MENU`, option number one leads to the menu above in the system of menus.

5   Redefining and creating variables

It is possible to redefine Xpose data variables as well as to create new data items within an Xpose session. This is done from the `MANAGE DATABASES MENU` (option 2 from the `MAIN MENU`).

```
DATABASE MANAGEMENT MENU

1: Return to the main menu
2: Change run number/database
3: List items in current database
4: View the documentation for the current data base
5: Change id variable
6: Change independent variable
7: Change dependent variable
8: Change pred variable
9: Change ipred variable
10: Change wres variable
11: Change iwres variable
12: Change variable to label data points with
13: Change flag variable
14: Change current value of the flag variable
15: Change occasion variable
16: Change missing data variable
17: Change parameter scope
18: Change covariate scope
19: Change typical parameter scope
20: Map random effects to typical parameters
21: Change name of a variable
22: Change weight variable
23: Copy variable definitions from another data base
24: Create/modify items in the current data base
25: Turn color printing on
26: Keep all lines when reading table files
Selection:
```

Option 2 will prompt the user for a new run number to process, similar to the initial question when Xpose is started.

Option 3 will list the items in the current data base, including the Xpose data variable definitions:

```
The current run number is 4 .
```

```
The data base contains the following items:
ID TIME IPRE IWRE DV PRED RES WRES VID CL V KA Q VSS ETA1
ETA2 ETA3 ETA4 ETA5 VIST SEX SMOK CIGS AGE HT WT CMPL CRCL

The following variables are defined:

Id variable: ID
Label variable: ID
Independent variable: TIME
Dependent variable: DV
Parameters: CL
Covariates: SEX WT CRCL AGE
(Continuous: WT CRCL AGE )
(Categorical: SEX )
Missing value label: -99
Wres variable: WRES
Iwres variable: IWRE
Res variable: RES
Ipred variable: IPRE
```

This option tells us the current run number (4 in this case), what items we have in the current data base and which of the Xpose data variables that are defined.

Option 4 displays the documentation, if any, for the current data base.

Options 5-20 can be used to redefine the default settings for the xpose data variables.

Option 21 is used to change variable names, i.e. the default name in plots will be the the newly defined name instead of the four letter name obtained from the NONMEM table file. Unfortunately there are some limitations in Xpose (or S-PLUS rather) regarding the names that can be given to a variable. The names should not contain spaces[7] or other "strange" characters like parentheses. Note that this does only apply when *renaming* a variable in this way. In the specification of axis labels and plot titles it works perfectly fine with any ASCII string (see Section 13).

With option 22 it is possible to define prior weights used in the GAM and the bootstrap of the GAM analyses (see Sections 10.5.5 and 11.5).

If only one table file is used it is usually necessary to manually change some of the Xpose data variable definitions. Having to do so for each run can easily become tedious. With option 23 it is possible to copy variable definitions from one Xpose data base to another.

Option 24 leads to the `CREATE/MODIFY DATA ITEMS MENU` (See Section 5.1.)

Option 25 will toggle the usage of color when printing and exporting plots (see Section 13).

Option 26 affects the way Xpose reads table files. By default, Xpose will not put table file lines with `WRES` equal 0 into the current data base. There are, however, instances when this behavior is not desired and by selecting this option before reading a set of table files, Xpose will keep all table file lines.

---

[7]Spaces do actually work in some plots. If you want to have spaces in your variable names, just try it, it might work!

5.1 Creating and/or modifying data items

```
CREATE/MODIFY DATA ITEMS MENU

1: Return to the previous menu
2: Add a Time After Dose item to the current data base
3: Add averaged covariates to the current data base
4: Add exponentiated values of an item to the current data base
5: Add the logarithm of an item to the current data base
6: Add absolute values of an item to the current data base
7: Change covariates categorical <-> continuous
8: Copy data item from another data base
Selection:
```

This menu, which can be reached from the MANAGE DATABASES MENU contains options that creates or modifies data items in the current data base.

Option 2 creates a time after dose data item (the Xpose data variable is *tad*). This is useful if we are dealing with a long and/or variable dosing history, i.e. when the times of the observations are relative to the first dose rather than the time since the last dose. Selecting this option will create a new item in the current data base called TAD. Redefining the *idv* variable to TAD will make the plots using the independent variable, e.g. the basic goodness of fit plots, use time after dose instead of the time since the first dose. This option will try to read either of sdtab, mutab or mytab in that order. The column in the read table file that matches the name of the independent variable in the current data base will be used for the calculations. If no such column is found, Xpose will use the second column for the sdtab and mytab or the third column for the mutab.

Option 3 will add columns with within individual averages of continuous covariates that varies within one or more individuals. These "new" covariates will get the name of the original covariate with an m appended, e.g. the within individual average of serum creatinine levels (SECR) will be SECRm.

Option 4 will add a column with the exponentiated values of a user specified item in the current data base. The name of the new column will be the name of the original column prepended by exp, e.g. expPRED.

Option 5 will add a column with the logarithm of a user specified item in the current data base. The name of the new column will be the name of the original column prepended by log, e.g. logDV. This option will not work if there are values in the column that are less than or equal to zero

Option 6 will add a column with the absolute values of a user specified item in the current data base. The name of the new column will be the name of the original column surrounded by vertical bars, e.g. |ETA1|.

With option 7 it is possible change continuous covariates into categorical covariates and vice versa.

With option 8 it is possible to copy a data item from another xpose data base to the current data base. The new column will be called the same as the variable copied appended by the run number of the data base it was copied from, e.g. CL1 (the CL column of the data base for run number 1).

6 Plotting conventions

This section describes some details about the way that Xpose draws plots and also a description of the more unusual plot types.

6.1 Plotting symbols and individual curves

Many of the plots in Xpose do not use the regular type of plotting symbols, like ◆ or ❑, but rather a number, which by default is the ID number (Figure 2). This is to facilitate the identification of individuals and to make it possible to follow individuals between plots. The data item used as the plotting symbol can be specified in the database management menu. It is, for example, perfectly possible to specify the variable SEX as the variable to label data points with, making it directly visible if, e.g. males in general have larger WRES than females. It is also possible to turn the usage of ID numbers off (See Section14.)

In some plots, individual data points are connected with a line (Figure 2) to make it easier to spot trends in, e.g. goodness of fit plots. To decide which points that belongs to a certain individual, Xpose uses the Xpose data variable *id*. This is default set to the ID, but can be redefined in the database management menu to be something else. The data item to be used as plotting symbol and the data item identifying individuals can be specified separately (see Section 14).

6.2 Smoothes

One way to make it easier to discern underlying patterns in a plot with many data points is to add a smooth non-parametric curve. The idea with such a curve is that it should smoothly follow the general trend in the data without making assumptions about the functional form of the possible relationship between the variables plotted. Such smoothes are frequently used in the plots in Xpose (Figure 2). The curves are obtained using the `loess()` function in S-PLUS. `loess` uses locally weighted linear or quadratic regression in overlapping intervals of the data [5]. The `loess` curves in Xpose use a locally linear model and an interval span of 2/3 of the data range.

6.3 Visual reference grids

In some plots there is a *visual reference grid* (Figure 3). These are not intended for reading off individual data values but rather for making the comparison with another plot or panel easier.

6.4 Trellis Graphics

Trellis Graphics^TM is a new graphics library in S-PLUS. It has been developed by William S. Cleveland and coworkers and is described in the book *Visualizing Data* [6]. The theory behind Trellis Graphics^TM is based on modern research on exploratory data
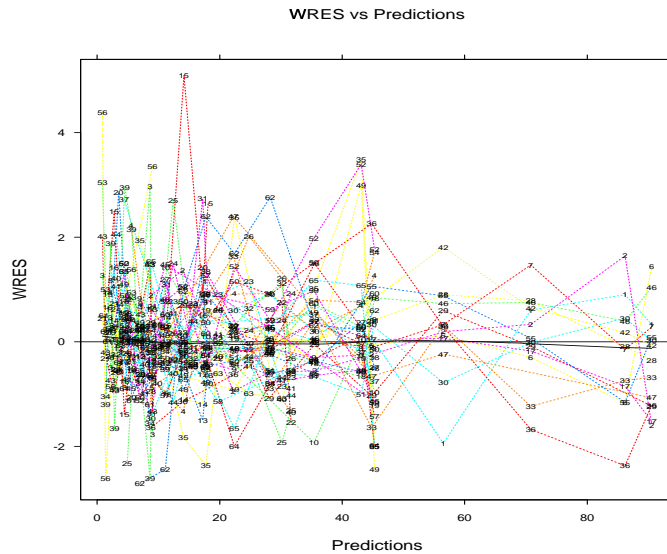
Figure 2: A plot of WRES vs predictions. The points are labeled with the ID number and each individuals points are connected by a dashed line (also based on the ID number). Added to the plot is a smooth curve to enhance any underlying pattern.

analysis and visual perception [7, 8]. Xpose does *not* utilize the full power of Trellis Graphics[TM]. To do so it would be necessary with a far more flexible graphical user interface than the text based menu system that is currently implemented in Xpose. Nevertheless, Trellis Graphics[TM] offers a lot of advantages (and implementation drawbacks!) over the core S-PLUS graphics and will be the default graphics methodology in S-PLUS from version 4 and onward.

6.4.1 Multi-panel conditioning

One of the big advantages of Trellis Graphics[TM] is the ease with which plots of one variable vs another given a third can be drawn. This is in the core S-PLUS graphics called a coplot but is in the Trellis Graphics[TM] terminology called a multi-panel conditioning plot [9]. The idea behind this type of plot is to extend the regular two dimensional diagrams to three or more dimensions (more than three dimensions is possible by specifying more than one given variable, however, Xpose uses only one given variable) and thus be able to visualize relationships that would otherwise be hard to see.

Figure 4 is an example of a multi-panel conditioning plot of CL vs HCTZ (a categori-
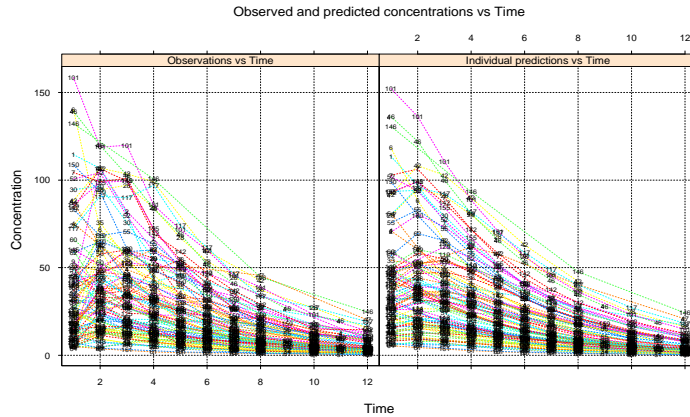
Figure 3: A plot of the observed and predicted concentrations vs time. The observations and predictions are plotted vs time in separate panels using the same scales on the x- and y-axes respectively. A visual reference grid has been added to facilitate comparison of the plots in the two panels.

cal covariate indicating whether the patient was also treated with hydrochlorothiazide) given `SEX` (the given variable). In this case the conditioning is based on the levels in the variable `SEX` (1 or 2). It is also possible to condition on *intervals* of a continuous covariate (Figure 5). The top panel, called the given panel, shows the intervals. The x-axis of the given panel gives the range of the given variable and the horizontal bars shows the intervals. The multi-panel plot at the bottom is the actual plot, divided up in the intervals. The lowest bar in the given plot corresponds to the lower left panel, the second lowest bar corresponds to the lower right panel and so on from bottom left to top right.

In Xpose, the given intervals are constructed so that there are approximately the same number of points in each panel, with approximately half of the points in common between the intervals. The latter is to have a smooth transition between the panels so that, e.g. a smooth through the data in each panel can be followed from lower left to top right. Xpose will determine the number of intervals from the total number of data points. With fewer than 50 data points there will be two intervals, with between 50-75 data points there will be three intervals, with between 75-100 there will be 4 intervals and with more than 100 points there will be six intervals.

6.4.2 Box and whiskers plots

The plot in Figure 4 is a box and whiskers plot (in the core S-PLUS graphics called a boxplot). These plots are in Trellis Graphics[TM] plotted horizontally rather than vertically. The solid dots in the middle of the boxes are the medians, which measures the center, or location, of the distributions. The length of the box, which is the inter-quartile range (i.e. the middle 50% of the data are enclosed in the box), is a measure
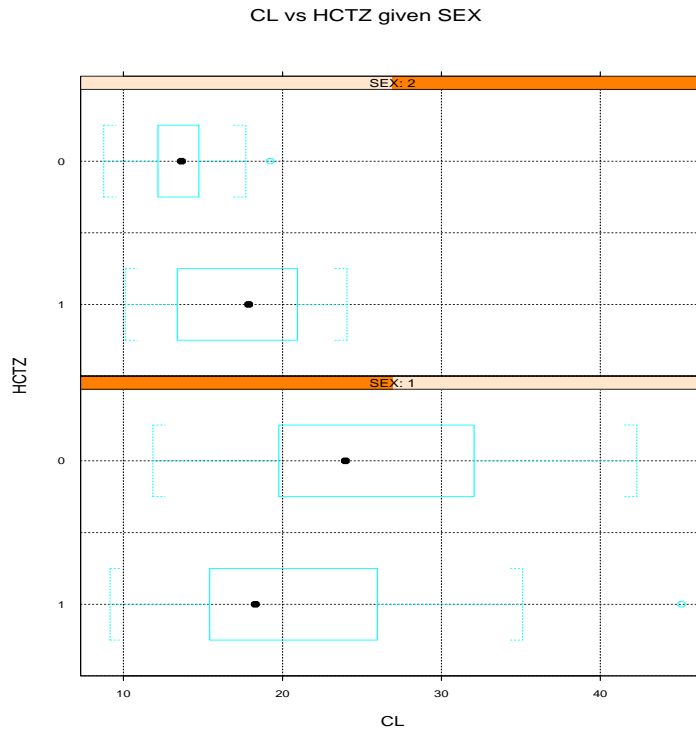
Figure 4: A multi-panel conditioning plot of the relationship between CL and hy-drochlorothiazide (HCTZ) usage given SEX . See text for further details.

of the spread of the distribution. If the inter-quartile range is small the middle data are tightly packed around the median. If the inter-quartile range is large, the middle data spread out far from the median. The relative distance of the upper and lower quartiles from the median gives information about the shape of the distribution. Unequal dis-tances indicates that the distribution is skewed. The whiskers encode the upper and lower adjacent values. The upper adjacent value is the largest observation that is less than or equal to the upper quartile times 1.5. The lower adjacent value follow the same, but opposite, calculations. Outside values, which are observations beyond the adjacent values, are graphed individually. The adjacent and outside values provide summaries of the spread and shape at the tails of the distribution.

The groups in each of the panels of Figure 4 are ordered so that the HZTC group with the lowest grand median (i.e. regardless of the division in males and females) is plotted first, that is lowest, in each panel. This is called *main effects ordering* and is part of the Trellis Graphics[TM] philosophy [6]. It also makes it easier to see subgroup deviations from the general trend . This ordering is used in almost all box and whiskers plots in Xpose with the notable exception of the GAM plots (see Section 10).
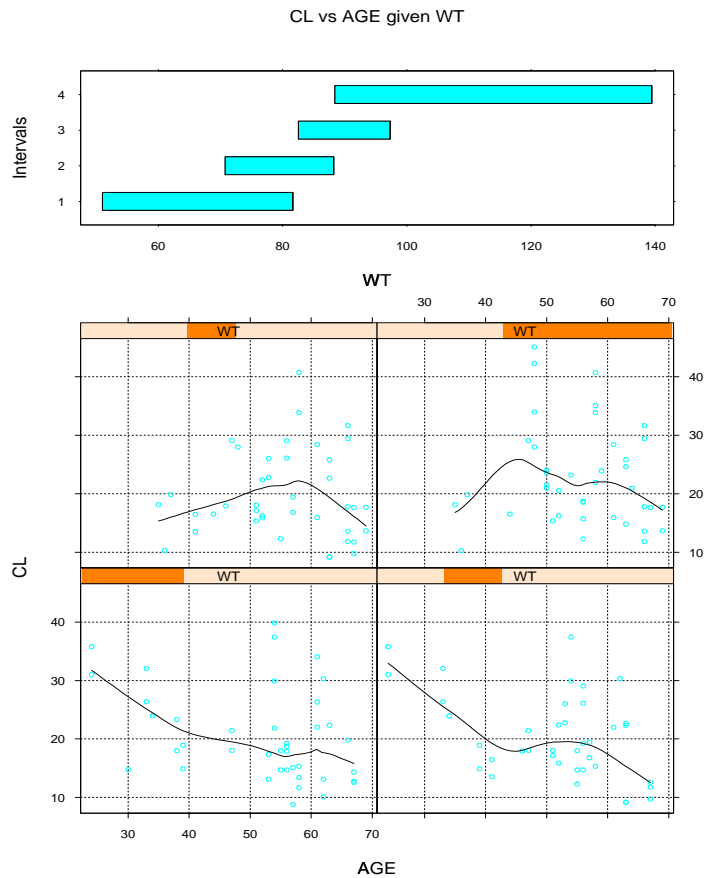
Figure 5: A multi-panel conditioning plot of CL vs AGE given intervals of WT . See text for further details.

6.5 What data is plotted?

In the process of reading the input table files and the extra files (see Section 2.4) Xpose discards the lines in the table files with a WRES entry equal to zero. This is to get rid of the lines corresponding to dosing events, which do not contain any information about the fit. The remaining data, henceforth referred to as all data, corresponds to the observation events in the NONMEM data file.

In many plots we are interested in all data, for example plots including WRES. In plots and analyses of the parameters, on the other hand, we are usually only interested in one data point per individual. In Xpose, this data point is taken from the first line in each individual. For example, if we want to plot CL vs CRCL (creatinine clearance), Xpose will plot the CL and CRCL values on the first line of each individual, regardless of whether the CRCL change within the individual. This behavior is fine as long as the covariates do not change within the individuals. If there are small within individual changes in the covariates or if the change is of no interest to the analysis, we

can let Xpose calculate within individual averages of the continuous covariates (see Section 2.4) and then use the averaged covariates instead of the corresponding real covariates when making the plots. An alternative strategy is to construct a new variable that change value every time the within individual varying covariate change (this can presently not be done within Xpose). Making the plot with the Xpose data variable *id* set to this new variable will display all covariate values.

There are other means to influence the way Xpose selects the data to be plotted or analyzed. In the GAM we can tell Xpose to use all data instead of the first data line for each individual (see Section 10). It is also possible to make plots and analyses on a subgroup of the data (see Section 15).

6.5.1   Handling of missing data

If the Xpose data variable *miss* is set (the default value is -99 but can be changed in the database management menu), Xpose will treat the value of this variable as missing data. The way this works in practice depends on the plot or analysis requested. In plots of one data item vs another, for example a parameter vs a covariate, individuals with a value of either the covariate *or* the parameter equal to the *miss* variable, will be excluded from the plot. In plots or analyses that handles blocks of data, e.g. the GAM or multi panel conditioning plots, individuals will be excluded that have at least one data entry equal to the *miss* variable. For example in the GAM, if an individual have a missing value of, say, `CRCL`, that individual will be excluded completely from the GAM analysis.

7   Documenting runs

A new feature in Xpose 2.0 is the possibility to document runs in other ways than in the form of plots. Run documentation can in Xpose 2.0 be done in two ways. First, a single run can be documented using a *run summary*. Second, a series of runs can be summarized in a *run record*. These documentation features can be reached from the `DOCUMENTATION MENU` on the `MAIN MENU`.

```
DOCUMENTATION MENU

1: Return to previous menu
2: Change run number
3: Summarize run
4: Construct a run record
Selection:
```

7.1   The run summary

The run summary is a one page summary of a run consisting of goodness of fit plots, NONMEM termination messages, parameter estimates and, optionally, the model file. An example run summary is displayed in Figure 6[8]. The run summary is intended as

---

[8]The code in the figure works only for NONMEM V

a way to document runs. It is, however, not a substitute for the NONMEM output file, which contains loads of other, useful, information.

At the top of the run summary are four basic goodness of fit plots. Below the plots, in the left column, are the termination messages from NONMEM, followed by the value of the objective function. Next follows a table with parameter estimates. in the case of the θ, the estimates are reported as they come in the NONMEM output file. The variance estimates (ω^2 and σ^2) are converted to standard deviations and the covariances (off-diagonal ω^2s and σ^2s) are converted to correlations. The standard error estimates are reported as the coefficient of variations (CVs) of the corresponding parameter estimates, i.e. the standard error divided by the parameter estimate. For the ω^2s and σ^2s the CV refers to the variance estimate and not to the standard deviation. Xpose can handle a maximum of 40 θs and 30 ω^2s and σ^2s.

Below the table with parameter estimates is the model file, possibly continuing in the right column. The inclusion of the model file is optional (see below). The run summary is only one page long so if the model file is longer than what will fit in the left and right column, it will be truncated.

A run summary can be requested with option 3 on the DOCUMENTATION MENU. When the user selects this option, Xpose first asks the user if he or she wants to send the output directly to the printer[9] (this is different from all other options where option to print the plot is available first after the plot has been shown on screen, see Section 13). When the user has specified where the run summary is to be sent, Xpose prompts for the name of the model file. Pressing return will accept the default model file name[10] and typing 0 means that no model file should be included in the run summary. If the default file name is not right the user has to type the correct file name followed by return. The next question is about the name of the output file name. Pressing return accepts the default output file name[11] and typing 0 will return to the MAIN MENU. If the default file name is not right the user has to type the correct file name followed by return. An example of how a run summary can be requested is given below:

```
DOCUMENTATION MENU

1: Return to previous menu
2: Change run number
3: Summarize run
4: Construct a run record
Selection: 3
Do you want to send the summary output to the printer y(n)?

Type the name of the model file (0=skip model file,return=run4.mod)
run4.mod
Type the name of the output file (0=exit,return=run4.lst)
run4.lst
```

---

[9]The run summary output is optimized for printing and will consequently not look very good on screen

[10]The default model file name is constructed from the combination of run , the run number and .mod , for example run1.mod .

[11]The default output file name is constructed from the combination of run , the run number and .lst , for example run1.lst .

Summary of run 3
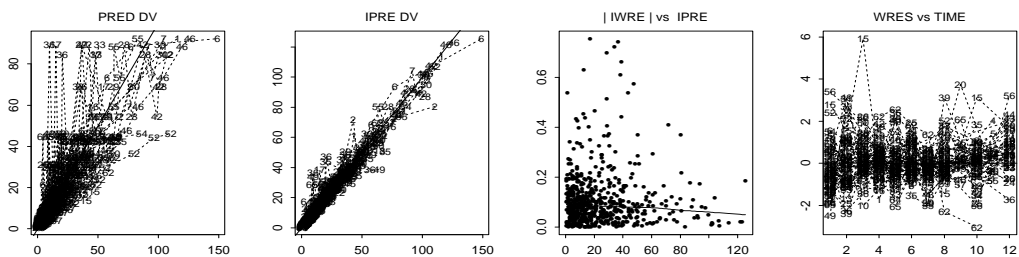

```
MINIMIZATION SUCCESSFUL
NO. OF FUNCTION EVALUATIONS USED:  223
NO. OF SIG. DIGITS IN FINAL EST.:  3.1


Objective: 2186.855


TH1     20.7     OM2:2   0.43
TH2     85.1     OM3:3   1.2
TH3     1.75     SI1:1   0.17
TH4     0.916
OM1:1   0.37



$PROB Simulated data distributed with Xpose 2
;;
;; As run2 but with the covariance step and NM V tables
;;
;;
$INPUT ID AMT TIME DV SS II SEX AGE RACE HT SMOK CON1 CON2 CON3 WT SECR D
$DATA data1.dta IGNORE=#
$SUBROUTINE ADVAN2 TRANS2
$PK
    SXCL = 1
    IF(SEX.EQ.2) SXCL = THETA(4)
    TVCL = THETA(1)*SXCL
    TVV  = THETA(2)
    TVKA = THETA(3)
    CL   = TVCL *EXP(ETA(1))
    V    = TVV  *EXP(ETA(2))
    KA   = TVKA *EXP(ETA(3))
    S2=V
$THETA (0,20) (0,87) (0,2) (0,0.5)
$OMEGA 0.12 0.14
$OMEGA 1.8
$ERROR
    DEL = 0
    IF(F.LE.0) DEL = 1
    IPRED = F
    IRES  = DV - F
    W     = F + DEL
    IWRES = IRES/W
    Y     = IPRED + W*EPS(1)
$SIGMA  0.23
$EST NOABORT POSTHOC
; Table statement for NONMEM IV. The data variables has to be defined
; within Xpose.
;$TABLE ID TIME IPRED IWRES CL V KA AGE HT WT SECR SEX RACE SMOK
;       CON1 CON2 CON3
;         NOPRINT ONEHEADER FILE=mytab3
; Table statements for NONMEM V
$TABLE ID TIME IPRED IWRES
        NOPRINT ONEHEADER FILE=sdtab3
$TABLE ID  CL V KA ETA1 ETA2 ETA3
        NOPRINT  ONEHEADER FILE=patab3
$TABLE ID AGE HT WT SECR
        NOPRINT  ONEHEADER FILE=cotab3
$TABLE ID SEX RACE SMOK CON1 CON2 CON3
        NOPRINT  ONEHEADER FILE=catab3
```

Figure 6: An example of a run summary.

7.2 The run record

A run record is basically a table with summary information about many runs. It can be thought of as an index to the run records. The information that Xpose includes in a run record is the run number, user supplied comments, NONMEM termination messages and objective function values.

The run numbers to be processed are supplied by the user. Only continuous ranges of run numbers are permitted (see below).

The user supplied comments are taken from the model files. The comment character in NM-TRAN model files is the semi-colon (;). User supplied comments for the run record should be preceded with two semi-colons (;;—). In other words, all text after two semi-colons present in the model file will be extracted and included in the run record. Below is an example:

```
$PROBLEM Example of user supplied comments
;; This line will be extracted into the run record.
;; As will this.
;   This line will not be included.
$INPUT ...
$DATA  ...
....
```

It is usually a good idea to include the comments just below the $PROBLEM statement.

The termination messages and objective function values are taken from the NONMEM output files.

The run record is not displayed on screen. Instead, it is either sent directly to the printer or to a file suitable for insertion into a word processor. The latter option is to make it possible for the user to influence the formatting of the run record, for example with page numbers. The items in the run record file are tab-separated, i.e. when the run record file has been imported into the word processor the user has to select the imported text and use a convert text to table option (present in for example Microsoft Word). Further more, separate lines with in one entry in the run record, e.g. the user supplied comments in the example above or the NONMEM termination messages, are separated by two hash characters (##). This means that the process of importing a run record file into a word processor can be summarized in the following four steps:

1. Construct the run record file (see below)

2. Import the table into a word processor

3. Convert the imported file to a table

4. Replace all ## entries in the table to *manual line feeds* or equivalent.

5. Format the table

Below is an example session:

```
DOCUMENTATION MENU

1: Return to previous menu
2: Change run number
3: Summarize run
4: Construct a run record
Selection: 4
Please type the range of run numbers to create a run record for
Range starts at run number (1):1
Range ends at run number:3
Do you want to export the documentation to a file?
n(y)n
```

8 Checking a data set

In the Data checkout menu (option 3 on the Main menu) there is an option called `Check a data set`[12] With this option it is possible to plot the `ID` number vs the values of each column in an NM-TRAN data file to find any gross errors.

The format of the data file has some restrictions. The first column of the data file will be plotted vs the values in the other columns. Hence, to plot the `ID` vs the other columns it is necessary to have the `ID`-number in the first column. When the data file is read by S-PLUS it will regard dates and colon separated clock times as character data, which cannot be plotted. To avoid the error these columns would give rise to if Xpose tries to plot them, it is necessary to give them the name `XXXX`, i.e. when Xpose have read the data file it avoids all columns with this name. A similar problem will arise if NULLs in the NM-TRAN data file are coded by a dot (`.`). Since NULLs are legitimate in some of the columns of greatest interest (e.g. `AMT` and `DV`) they are better replaced by zeros.

When the user selects the `Check a data set` option from the `DATA CHECKOUT MENU` he or she is first prompted for the name of the data file. This should, obviously, be the name of an existing NM-TRAN data file, followed by return. Xpose then prompts for the line number at which the column headers are[13]. The default is 2, which is the same as a NONMEM table file would have. It is however perfectly fine to specify another line number, making it possible to keep comments regarding, e.g. the contents of the data file, at the top of the file. Below is an example session:

```
DATACHECKOUT MENU

1: Return to the main menu
2: Numerically summarize the covariates
3: Histograms of the covariates
4: Histograms of the covariates given categorical covariates
5: Scatterplot matrix of covariates
6: Check a data set
7: DV vs the independent variable
```

---

[12]A similar option was available in Xpose version 1.1. In the new version of Xpose these plots are much faster, unfortunately with the loss of some, but not very informative, information.

[13]This is new compared to Xpose 1.1. It is no longer necessary to follow the NONMEM table file convention with the column headers on the second line.

```
Selection: 6
Please type the name of the data file you want to check
data4
Please type the line number of the line in the data file
containing the column headers (2):4
```

The contents of the data file is plotted using dot plots, one for each column (Figure 7). On the y-axis are the ID numbers and on the x-axis are the AGE values. Each entry for each value of ID will get one dot on the corresponding line.
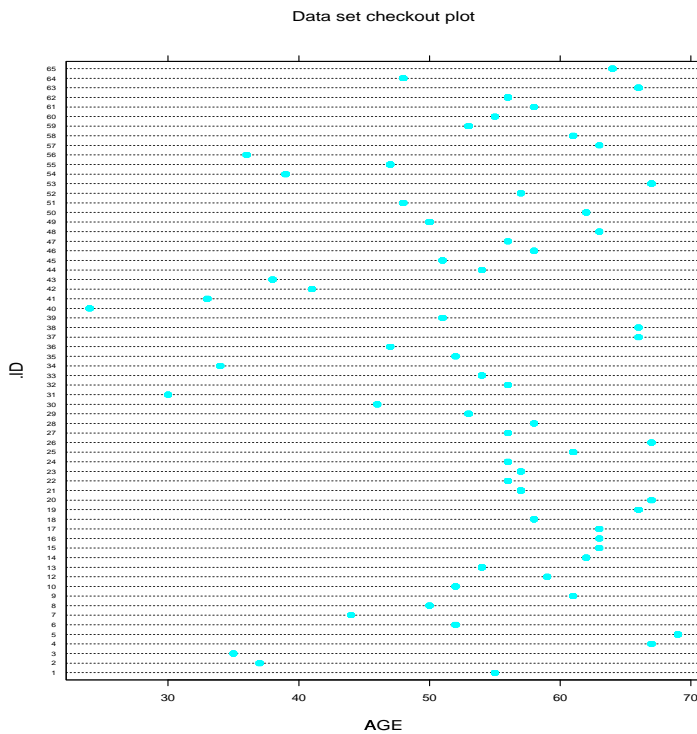
Figure 7: A dot plot of ID vs AGE .

9 Comparing models

To discriminate between rival models is an important aspect of model building. In NONMEM it is possible to use the difference in the objective function value to discriminate between hierarchical, or nested, models. In Xpose it is possible to visualize the differences between two model fits. Option 7 in the MAIN MENU leads to the MODEL COMPARISON MENU:

```
MODEL COMPARISON MENU

1: Return to previous menu
```

```
2: Basic model comparisons
3: Additional model comparisons
4: Delta PRED/IPRED/WRES/IWRES vs covariates
Selection:
```

The plots in this menu plots, for example, `WRES` from the current data base vs the `WRES` from another, reference, data base. When one of the plot options is selected Xpose prompts the user for a reference run number. Similarly to changing the run number in the `DATABASE MANAGEMENT MENU`, Xpose checks if there is an existing xpose data base matching the run number. If there is, the user gets the option to use the data base or to recreate it from the table files. If no matching data base is found the user is asked if he or she wants to create it from the table files. Note that the same data has to be used in the two runs to make the comparison meaningful.

10   The GAM

Direct inclusion of covariate relationships in the model is one of the important benefits with population analysis over individual specific analysis. Identifying the relevant covariates can, however, be a time consuming task. To speed this up, Mandema and coworkers [10, 11] suggested the use of a stepwise generalized additive modeling procedure (GAM[14]).

The GAM can be run in Xpose by selecting the `COVARIATE MODEL MENU` in the `MAIN MENU` (option 6) and then the `GAM MENU` (option 4):

```
GAM MENU

1: Return to previous menu
2: Run a GAM
3: Summarize GAM fit
4: Plot GAM results
5: Akaike plot
6: Studentized residuals for GAM fit
7: Individual influence on GAM fit
8: Individual influence on GAM terms
9: Settings for the GAM
Selection:
```

The GAM in Xpose 2.0 offer a great deal of flexibility[15] and a number of GAM specific plots that are not directly available in S-PLUS.

---

[14]We use the term GAM to refer to the *stepwise* generalized additive modeling procedure in contrast to a generalized additive model, which is just single model without any variable selection.

[15]There are a lot of differences between the way the GAM is implemented in Xpose 1.1 and Xpose 2.0, some are user visible while others, perhaps the major part of the differences, are not. It is still possible to run the GAM in the same way, that is, using the same options, as in Xpose 1.1, and the GAM results, given a certain data set, should be identical.

10.1   Overview of the GAM

This section contains a brief description of how the GAM works in S-PLUS. It is intended to give enough background for the following sections that describes how the GAM can be used in Xpose. For a more formal description of this feature in S-PLUS, please refer to the S-PLUS manuals [5] or to the book *Statistical models in S* by Chambers and Hastie [12]. For a technical description of generalized additive models in general, the book by Hastie and Tibshirani [13] is a good start.

The GAM can be used to find a subset of the available explanatory variables (i.e. the covariates) that are most useful in explaining the variability in the dependent variable (e.g. `CL`). This is done by trying combination of models of the explanatory variables in a stepwise fashion. The model discrimination is made by comparison of the Akaike information criteria (AIC). The order in which the models are tried is defined by the *model scope*. For each explanatory variable, a hierarchy of possible models is defined, for example: not included, a linear model and a non-linear model. The initial model is usually a combination of the first models in the scope but can be any combination of the models defined in the scope. In each step, for each covariate, the models up and down in the hierarchy are tried and the model that decreases the AIC statistic the most are retained to the next step. The search is terminated when no model can decrease the AIC any further.

10.2   Running the GAM

To run the GAM in Xpose we need to have a current data base in which we have covariates and at least one parameter defined. The default scope, for each covariate, is not included, a linear regression model and a natural cubic spline with one internal breakpoint (e.g. `ns(AGE,df=2)`). For categorical covariates the third model is omitted. When implementing a covariate model in NONMEM it is usually a good idea to center the covariates to the middle of the data (this makes it easier to get reliable estimates of the intercept in a linear model). In Xpose, continuous covariates are by default centered around the median covariate value. Selecting option 2 in the GAM Menu starts the GAM analysis.

Xpose saves the GAM objects created by the GAM analysis. This makes it possible to go back and look at the results of a GAM run at a later time. These objects, *Xpose gam objects* are given names which are a combination of `gam.xpose.`, the parameter name and the run number, for example `gam.xpose.CL.1`. If the GAM is requested for a run number and parameter with an already existing Xpose gam object, the user has to confirm that he or she wants to delete the old object.

10.3   Displaying the results of a GAM run

The results of a GAM run can be displayed either numerically or graphically.
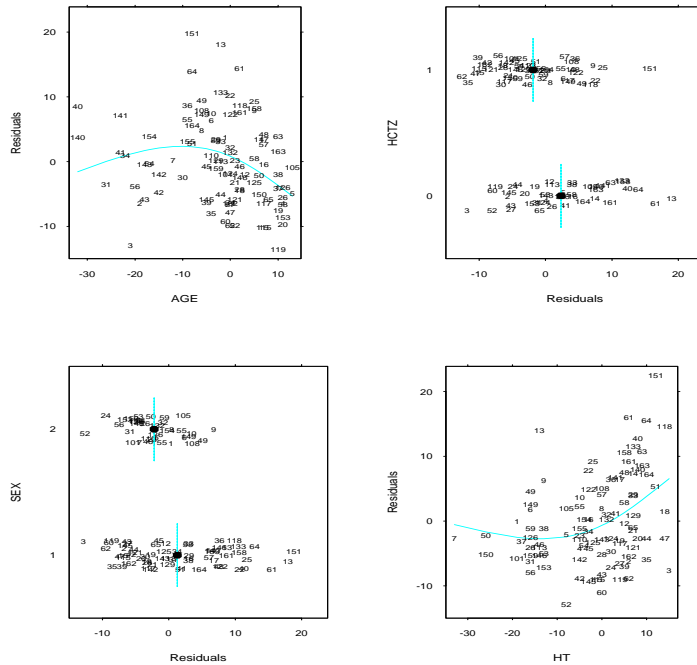
Figure 8: Plot of GAM results. Each panel shows the relationship between a covariate and CL . Data points are labeled by the ID number

10.3.1 Numerical summary of a GAM fit

Selecting option 3 on the GAM MENU produces a numerical summary of an Xpose gam object in the current S-PLUS working directory. This summary consists of the output from the S-PLUS function gam.summary plus information about the Xpose GAM settings used (see Section 10.5) to produce the Xpose gam object.

10.3.2 Displaying the GAM results graphically

It is also possible to view the GAM results graphically. Option 4 on the GAM MENU, will produce plots of the final GAM model (Figure 8). Individual data points are labeled with the *id* variable and a smooth is drawn in the plots with continuous covariates. The fitted values of relationships with categorical covariates are indicated by a solid black dot and a vertical blue line.

10.4 Obtaining diagnostics for a GAM run

Using stepwise procedures, such as the GAM, is not without criticism (e.g. G.E.P Box, Technometrics (8) 1966 625-629) and even if the GAM has proven to be useful for the

Figure 9: Akaike plot of a GAM run. On the y-axis are the most important models tested by the GAM and on the x-axis are the corresponding AIC values.

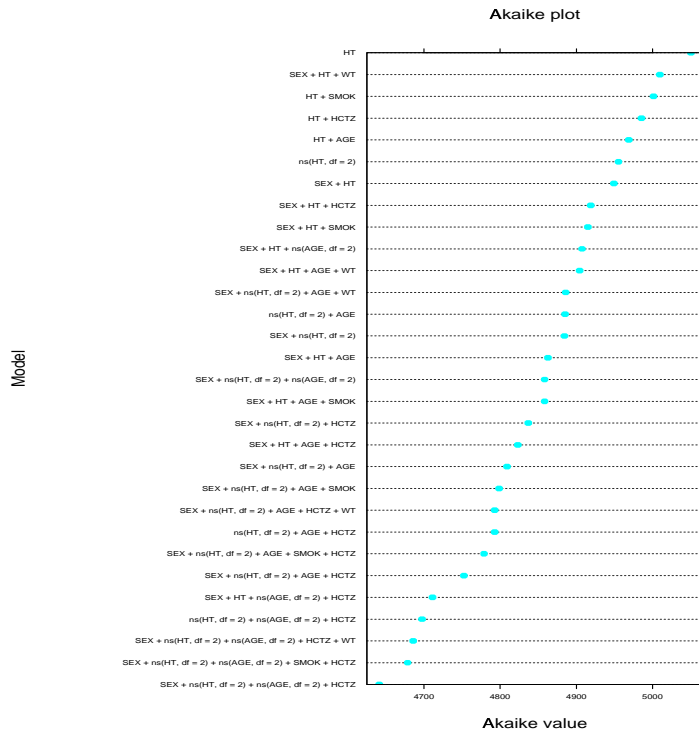identification of important covariates, one should be careful in the interpretation of the results. One of the things that can influence the GAM results are outlying individual data points. These can both hide and create covariate relationships [14]. Xpose has a number of options to check the final model for unduly influential individuals. These options uses quite involved calculations of approximate case-deletion diagnostics for the final GAM model [15]. A detailed description of the way these are implemented in Xpose is given in Jonsson and Karlsson [14] and in the Appendix (to be written).

For options 6, 7 and 8 it is necessary to use parametric smoothers, i.e. `ns`, `bs` and the hockey stick models[16] (see Section 10.6).

10.4.1 The Akaike plot

Option 5 on the GAM menu will produce a dot plot of the 30 most important models tried in the stepwise search for the final model (Figure 9). On the y-axis are the models (using S-PLUS syntax) and on the x-axis are the corresponding AIC value. With this plot it is possible to evaluate how much better the final model is compared to the other models. It can be viewed as a measure of the uncertainty in the final GAM model.

---

[16]Using the parametric models will make the generalized additive models to generalized linear models, and the calculations involved in option 6, 7 and 8, only holds for generalized linear models.
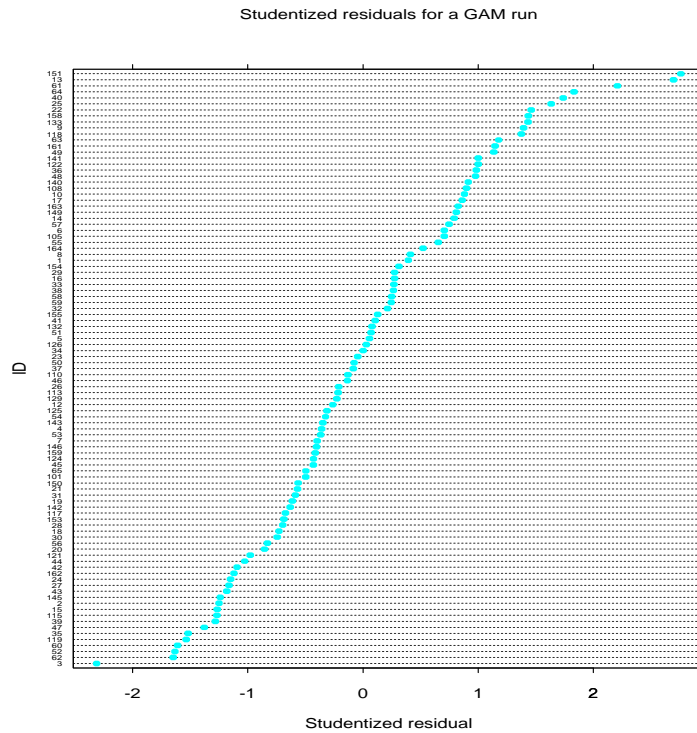
Figure 10: Studentized residuals for a GAM run. On the y-axis are the ID number and on the x-axis are the residuals. The ID numbers are ordered according to the size of the residuals.

10.4.2 Studentized residuals for a GAM fit

Option 6 on the GAM Menu will produce a dot plot of the studentized residuals for each individual. A studentized, or jacknifed, residuals, is the residual between an individual data point (parameter) and the prediction from a fit with that particular data point omitted. These residuals are normalized so that the variance should be 1 and the mean zero, making them suitable for the detection of influential points.

An example of this plot is given in Figure 10. On the x-axis are the values of the studentized residuals and on the y-axis are the ID numbers as given by the *id*-variable. The ID numbers are ordered according to the size of the residual.

10.4.3 Individual influence on the GAM fit

Option 7 will produce a plot (Figure 11) of the Cooks distance vs leverage [14, 15]. Cooks distance is a measure of the influence a certain data point has, i.e. how much the fit will change if that data point is omitted from the analysis. A high value indicates a high influence. The leverage is a measure of how a data point influences the certainty with which the fit is obtained. A high value indicates a higher leverage. A high leverage data point need not be influential and vice versa. A point with a high value of
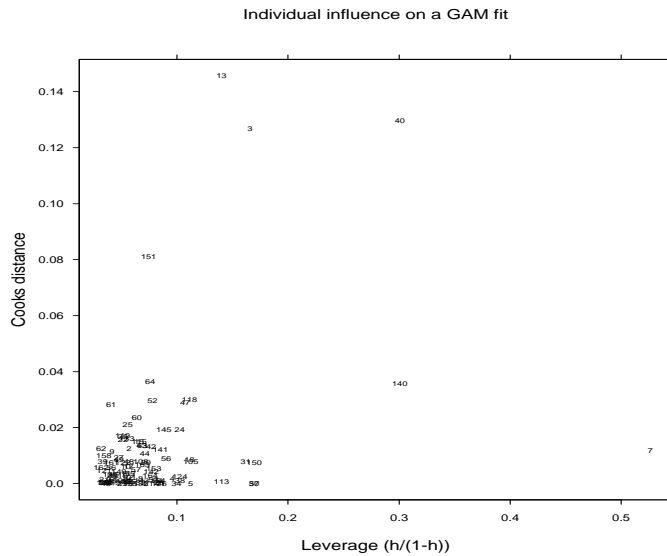
Figure 11: Cooks distance vs leverage for a GAM fit. Data points are labeled by the ID number

Cooks distance and leverage is very important to the fit and often affects the covariate selection [14]. This property also seem to "travel" into the NONMEM model, i.e. if an individual point has a high leverage and influence in the GAM he or she will also be important for the covariate model in NONMEM [14].

10.4.4   Individual influence on the GAM terms

Option 8 will produce plots of the Cooks distance for each covariate term in the final GAM model. This plot can be used to identify which of the covariate terms that are affected by the individuals detected with option 7.

10.5   Changing the default behavior of the GAM

Option 9 leads to the GAM SETTINGS MENU:

```
GAM SETTINGS MENU

1: Return to previous menu
2: List current settings
3: Estimate dispersion factor (on/off)
```

```
4: Stepwise search for covariates (on/off)
5: Use all lines (on/off)
6: Use weights (on/off)
7: Exclude individuals
8: Specify starting model
9: Normalize to median (on/off)
10: Set model scope
```

The options in this menu can be used to change the way the GAM analysis behaves.

10.5.1 List current GAM settings

Option 2 will list the current GAM settings. These settings will be effectual until changed in the GAM SETTINGS MENU or until S-PLUS is quit. The settings with which a certain Xpose GAM object was obtained can be listed using the SUMMARIZE A GAM RUN option in the GAM MENU (see Section 10.3).

10.5.2 Estimation of the dispersion factor

This option is a toggle, e.g. if the setting is *not* to estimate the dispersion factor (the default) and the option is selected from the menu, the GAM will estimate it, and vice versa.

The dispersion factor is used in the definition of the AIC and is equal to the "cost", divided by 2, per degree of freedom incurred by adding or dropping a term from the GAM model. If this option is not set, the dispersion factor is taken to be the scaled Pearson chi-squared statistic for the starting model[17] (see Section 10.5.7). If the option is set, the dispersion factor is estimated according to the following scheme: For each covariate, the models without covariate, a linear function of the covariate and non-linear function of the covariate (a natural cubic spline with one internal breakpoint) are compared. The dispersion factor is taken from the gam fit of the model resulting from the combination of the significant models according to this univariate analysis.

10.5.3 Stepwise search for covariates

Option 4 will turn the stepwise search on or of (a toggle). The stepwise search is on by default. When it is off, the GAM will fit the model specified by the starting model. This can be useful when a covariate model have been implemented in NONMEM and the user wants to obtain approximate case deletion statistics for that particular covariate model. Option 8 (Specify starting model) can be used to specify the model to be fit.

10.5.4 Use all lines

Option 5 will toggle the use of only one line per individual on or off. If the setting is off, all lines in the Xpose data base will be used.

---

[17]In Xpose 1.1 the dispersion parameter was not estimated.

10.5.5 Use weights

Option 6 will turn the use of prior weights on or off. If the setting is on, Xpose will use the Xpose data variable *weight* (see Section 2.5) as prior weights in the GAM analysis.

10.5.6 Exclude individuals from the GAM analysis

If a certain individual is identified as having a large influence and leverage on the final GAM model, it is interesting to know whether the exclusion of the individual will change the covariate selection by the GAM. Option 7 can be used to exclude one or more individuals from the GAM analysis. Selecting the option will prompt the user for the `ID` numbers of the individuals to be excluded. The `ID` number(s) to specify is the number appearing in the plots of the GAM results.

10.5.7 Specifying a starting model for the GAM

The starting model for the GAM is the model from which the stepwise search is started. It is also used to define the dispersion factor used in the search. Selecting option 8 will prompt the user for the *individual model terms* that makes up the starting model. These terms should be the terms used by S-PLUS to specify the starting model. For example, if we want to use a linear relationship with `AGE` and a non-linear relationship with `HT` as the starting model, the corresponding command in S-PLUS would be `gam(CL~AGE+ns(HT,df=2))` (assuming we wanted the natural cubic spline with one internal break point as the non-linear model). The terms that should be given to Xpose when specifying this model are consequently `AGE` and `ns(HT,df=2)`.

This option can also be used to specify the model to fit if the stepwise search is turned off (see Section 10.5.3).

10.5.8 Normalize the covariates to the median

Option 9 will turn the normalization of continuous covariates to the median on or off. Median normalization is the default, i.e. the parameter estimates of the final GAM model (see Section 10.3) can be used as initial estimates in the NONMEM model. Note that the median normalization is done after individuals with missing covariate values are omitted (see Section 6.5.1).

10.6 Setting the model scope for the GAM

For a continuous covariate, the default model scope in the GAM is: not included, a linear model and a natural cubic spline with one internal breakpoint[18]. For a categorical covariate, the third, non-linear, model is omitted.

Option 11 leads to yet another menu in which the user can alter the default model scope.

---

[18]In Xpose 1.1 this is the only available model scope

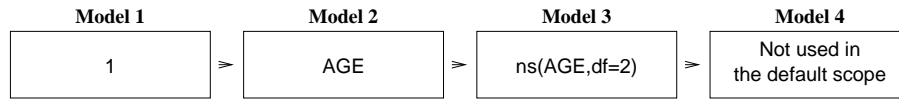| **Model 1**		**Model 2**		**Model 3**		**Model 4**
1	⪰	AGE	⪰	ns(AGE,df=2)	⪰	Not used in the default scope

Figure 12: Picture of the default scope for a continuous covariate

```
GAM SCOPE MENU

1: Return to previous menu
2: Set maximum number of models
3: Default smoothers and their arguments
4: Limit model scope for specific covariates
5: Change models for specific covariates
6: Show current GAM scope
7: Use default scope
Selection:
```

10.6.1 The GAM scope and its default settings in Xpose

The scope for a given covariate is a hierarchy of models that the GAM will test. It is hierarchic in the sense that the GAM will never look beyond one step up or down from the current model, i.e. for the default scope, which is depicted in Figure 12, the GAM will never try the non-linear model (Model 3) unless it first finds the linear model (Model 2). The scope shown in Figure 12 is the default scope for a continuous covariate. For a categorical covariate it is not possible to fit a non-linear model. A 1, as in Model 1 in Figure 12, is used to denote that the covariate is not included in the GAM model, the name of the covariate, AGE in Figure 12, denote a linear relationship and ns(AGE,df=2) denotes a natural cubic spline with one internal breakpoint. The latter is an example of the non-linear models that can be defined in the GAM scope.

In Xpose it is possible to alter the scope in one or more ways: increasing or decreasing the number of levels in the model hierarchy, change the default smoother (i.e. ns), exclude one or more of the covariates from one or more levels in the model hierarchy and change the model used for a specific covariate in a specific level of the hierarchy.

Note that changing the model scope can alter the final GAM model substantially and it is consequently possible to "find" the covariate model one wishes to find. An example of the rational for giving the user the increased flexibility is as follows: Assume that we have a non-linear relationship between a parameter and a covariate and that the (substantial) curvature is off center (see Section 10.6.2), i.e. the parametric smoothers will have problems detecting it. Ignoring this fact, i.e. being content with a linear relationship, is likely lead to structures in the unexplained variability that might induce other, false, relationships, or, and worse, might hide true relationships. In this case it is desirable to be able to specify another, more flexible, non-linear function that captures the non-linearity without being bound to the breakpoints of the parametric smoothers. A similar, but opposite, situation might occur if we have outlying individuals that induces a non-linear relationship even if the true relationship is clearly linear.

10.6.2 Smoothers in the GAM scope

The idea with generalized additive models compared to generalized linear models is to be able to use non-parametric smoothers, e.g. spline functions that use the observed data for its definition. The default smoother, or non-linear relationship, in the GAM as implemented in Xpose, is a natural cubic spline with one internal breakpoint, i.e. a parametric smoother. Consequently, the way that the GAM in Xpose is defined by default makes it a generalized linear model (parametric smoothers is a requisite of the approximate case-deletion diagnostics available, see Section 10.4).

There are basically five different smoothers that can be specified. Two are non-parametric, i.e. `lo` and `s`, and three are parametric (at least in the sense that it is possible to produce the GAM diagnostics if they are used), i.e. `ns`, `bs` and the *hockey stick* models (see Section 10.6.6). The last smoother can only be specified using one of the options in the `GAM SCOPE MENU` and will be described together with that option (see Section 10.6.6). The other smoothers are regular S-PLUS smoothers and the S-PLUS manuals gives further details.

The parametric smoothers have to have the breakpoints specified, i.e. where the centers of the curvature are. The default value of these breakpoints are the median (for one breakpoint) and the quartiles (for more than one breakpoint). If the curvature is much off center it might be a good idea to try one of the non-parametric smoothers or to specify other breakpoints

10.6.3 Setting the maximum number of models in the scope

The default number of models for a continuous covariate is 3. It is possible to decrease this to 2 or to increase it to 4. When the maximum number of models is changed, the default scope is also changed. Reducing the number of models to 2 omits the non-linear model (Figure 12), increasing the number of models to 4 adds a second non-linear model (a natural cubic spline with two internal breakpoints, i.e. a more flexible model than one breakpoint spline). Changing the maximum number of models does not affect the scope of the categorical covariates.

10.6.4 Changing the default smoothers and their arguments

It is possible to change the default smoothers for each hierarchic level of the scope for the continuous covariates. Xpose will ask the user for a smoother to use for each level in scope. The following Xpose output changes the default smoother for the third model from `ns` to `s`:

```
GAM SCOPE MENU

1: Return to previous menu
2: Set maximum number of models
3: Default smoothers and their arguments
4: Limit model scope for specific covariates
5: Change models for specific covariates
6: Show current GAM scope
7: Use default scope
```

```
Selection: 3

Type the name of the smoother you want to use for the *first*
covariate model in the GAM model scope (q=exit,d=default,
0=no covariate effect,1=linear model):

Type the name of the smoother you want to use for the *second*
covariate model in the GAM model scope (q=exit,d=default,
0=no covariate effect,1=linear model):

Type the name of the smoother you want to use for the *third*
covariate model in the GAM model scope (q=exit,d=default,
0=no covariate effect,1=linear model):s

Type any *argument* to the smoother for the third covariate model
in the GAM model scope (multiple argument should be comma
separated, q to exit, d =default):
```

(Note that the word smoother refers to any model of: not included, linear as well as the actual smoothers.)

For each model the user can specify the default (d), no covariate effect (0), a linear model (1) and the name of the smoother (e.g. s). If a smoother is specified the user is also prompted for the arguments to the smoother. For the s in the above example the default argument was accepted (df=4) but could have been set to, for example, df=3.

10.6.5 Limit the scope for specific covariates

With option 4 it is possible to limit the scope, or in other words, exclude a continuous covariate from a specific scope level. For example, if we have two continuous covariates AGE and WT and we want to skip the linear model for AGE in the default scope, the Xpose session might look something like this:

```
GAM SCOPE MENU

1: Return to previous menu
2: Set maximum number of models
3: Global smoothers and their arguments
4: Limit model scope for specific covariates
5: Change models for specific covariates
6: Show current GAM scope
7: Use default scope
Selection: 4
The following continuous covariates are defined in the
current data base
AGE WT

Type the names of the covariates you want to exclude
from the first model in the model scope and finish with an empty
line:
1:
```

```
Type the names of the covariates you want to exclude
from the second model in the model scope and finish with an empty
line:
1: AGE
2:

Type the names of the covariates you want to exclude
from the third model in the model scope and finish with an empty
line:
1:
```

For each level in the model scope the user is asked for the name(s) of the covariate(s) that is to be omitted. Answering with a return (and no covariate name) will leave the scope unaffected for all covariates.

10.6.6   Change models for specific covariates

With option 5 it is possible to specify the scope models for individual continuous covariates. Below is an example session for the covariate `AGE` and `WT`:

```
GAM SCOPE MENU

1: Return to previous menu
2: Set maximum number of models
3: Global smoothers and their arguments
4: Limit model scope for specific covariates
5: Change models for specific covariates
6: Show current GAM scope
7: Use default scope
Selection: 5
The following continuous covariates are defined in the
current data base
AGE WT
Type the name of the covariate you want to specify a special
model scope for (q to quit): AGE

Type the name of the *smoother* you want to use in the first model
of the model scope (q to exit, d=default, 0= exclude
and 1= linear model, R=hockey stick right, L= hockey stick left):

Type the name of the *smoother* you want to use for the second
model in the model scope (q to exit,d=default,
0=exclude,1=linear model, R=hockey stick right, L= hockey stick left):

Type the name of the *smoother* you want to use for the third
model in the model scope (q to exit,d=default,
0=exclude,1=linear model, R=hockey stick right, L= hockey stick left):R

(New covariate)

Type the name of the covariate you want to specify a special
model scope for (q to quit): q
```

The first the user has to do is to type the name of the first covariate to specify models for, in the above case AGE. Xpose then steps through the levels of the model scope and prompts the user for a model. The options are the default (d), exclude (0), a linear model (1), hockey stick right or left (R and L respectively) or the name of a smoother. In the latter case the user will be prompted for any arguments similar to option 3 on the GAM SCOPE MENU (see Section 10.6.4). The hockey stick right and left models fits a slope to the left or right hand side of the median covariate value, leaving the fit on the opposite side horizontal at the "intercept" in the middle of the data. These models are fairly inexpensive non-linear alternatives to the normal smoothers and can sometimes capture non-linearities that would otherwise not be detected.

When Xpose has stepped through the scope levels for the first covariate the user is prompted for a new covariate name. The answer should be either the name of the next covariate or q if no more covariate specific adjustments of the scope is needed.

10.6.7   Viewing the current GAM scope

Option 6 prints the current GAM scope with, if any, adjustments made by the user.

10.6.8   Re-setting the scope

Option 7 resets the scope to the default.

11   The bootstrap of the GAM

11.1   Overview the Bootstrap of the GAM

The bootstrap is a proven statistical methodology, which relies more on brute computational power rather than clever analytical solutions. A good starting point to the bootstrap literature is *Introduction to the Bootstrap* by Efron and Tibshirani [16].

The bootstrap of the GAM can be used to assess the importance of the covariates and to obtain similar diagnostics information to that available to the GAM, the difference being that the information is obtained for all covariates, not only for the "significant" ones. In addition, the bootstrap of the GAM can also give information about how covariates interact with respect to inclusion/exclusion from the covariate model.

In the bootstrap of the GAM, the GAM is run a large number of times on bootstrap realizations of the original data set. (The term *data set* will in this section be used to refer to the data set of individual parameter values and covariates, usually one entry, or line, per individual.) The bootstrapped data sets are constructed by randomly sample with replacement from the original data set and the GAM is then run on each of them. The basic results from a bootstrap of the GAM run is obtained simply by counting the number of times each of the covariates are selected, regardless of functional form of the relationship, as well as how many times the continuous covariates are selected as a non-linear function[19]. This "number of times" are presented as relative frequencies,

---

[19]Xpose 1.1 did not differentiate between linear and non-linear models.

i.e. the number of times a certain covariate or covariate model was selected divided by the number of bootstrapped data sets.

The bootstrap of the GAM, as it is implemented in Xpose, has been described by Jonsson and Karlsson [14] and similar uses of the bootstrap have been described by Mick and Ratain [17], Sauberei and Schumacher [18] and Ette [19].

To use the bootstrap of the GAM select option 6 in the `MAIN MENU` and then option 5 from the `COVARIATE MODEL MENU`. This will lead to the `BOOTSTRAP OF THE GAM MENU`:

```
BOOTSTRAP OF THE GAM MENU

1: Return to the previous menu
2: Run the bootstrap of the GAM
3: Summarize a bootstrap of the GAM run
4: Plot menu
5: Settings for the bootstrap of the GAM
Selection:
```

11.2   Running the bootstrap of the GAM

It is necessary to have covariates and at least one parameter defined in the current data base to run the bootstrap of the GAM.

Option two will run the bootstrap of the GAM. Xpose saves the bootstrap of the GAM run in *Xpose bootgam objects*. They are named by a combination of `bootgam.xpose.`, the parameter name and the run number, e.g. `bootgam.xpose.CL.1`. If there already is an Xpose bootgam object in the current directory that matches the parameter name and run number, the user has to confirm that he or she wants to overwrite it.

One of the important questions in the bootstrap of the GAM is the number of bootstrap iterations to make, i.e. the number of bootstrapped data sets to use. In Xpose 2.0 there are basically three different strategies. The first is to use a fixed, user specified, number of iterations[20]. The other two strategies uses algorithms to determine when the inclusion frequencies of the covariates are stable, i.e. Xpose determines the number of iterations needed. Section 11.5 describes these algorithms in more detail and also gives information about how to adjust their behavior.

11.3   Summarizing a bootstrap of the GAM run

Bootstrap of the GAM runs can be viewed both numerically and graphically. Option 3 on the `BOOTSTRAP OF THE GAM MENU` will produce a numerical summary of the run:

```
Convergence algorithm: Fluctuation ratio

Convergence criteria: 1.0368 (target= 1.04 )
Number of iterations: 150
```

---

[20]In Xpose 1.1 the user always had to specify the number of iterations to use.

```
Initial dispersion not estimated.
No start model specified.
Median normalization on.
Seed number: 286

Model size:
 Min. 1st Qu. Median Mean 3rd Qu. Max.
    1       3      3 3.18       4    5


     Prob Nonlin FrNonlin
  HT 0.747  0.44     0.589
 AGE 0.613  0.247    0.402
HCTZ 0.6    0        0
 SEX 0.493  0        0
SMOK 0.42   0        0
  WT 0.307  0.12     0.391
```

The first half or so of the output contains information about the settings used in the run and the second half contains results. First is descriptive statistics on the size of all the final GAM models. Below that comes a table with three columns containing the relative inclusion frequencies of the covariates and the non-linear covariate models, followed by the latter divided by the former, i.e. a measure of how important the non-linear relationship was compared to the total number of times the covariate was found.

11.4 Plotting the results

The results from a bootstrap of the GAM run can also be viewed graphically. Selecting option 4 on the BOOTSTRAP OF THE GAM MENU leads to the PLOT MENU FOR THE BOOTSTRAP OF THE GAM:

```
PLOT MENU FOR THE BOOTSTRAP OF THE GAM

1: Return to previous menu
2: Inclusion frequencies
3: Most common covariate combinations
4: Distribution of model size
5: Inclusion stability - covariates
6: Inclusion stability - non-linear
7: Inclusion index of covariates
8: Inclusion index of non-linear models
9: Inclusion index of covariates/individuals
10: Inclusion index of non-linear models/individuals
Selection:
```

11.4.1 Plotting the inclusion frequencies

Option 2 will produce a plot of the relative inclusion frequencies of the covariates and the non-linear covariate relationships (Figure 13).

The relative inclusion frequencies plotted in the top panel in Figure 13 were obtained by counting the number of times each covariate was selected by the GAM *regardless*
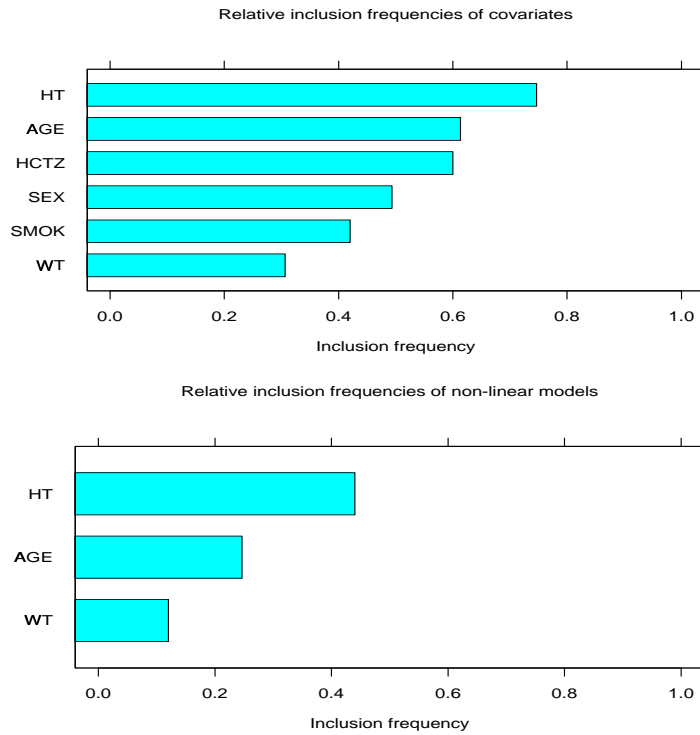
Figure 13: Barplot of the relative inclusion frequencies of the covariates (top panel) and the non-linear covariate models (lower panel).

of the functional form of the covariate relationship. The relative inclusion frequencies plotted in the lower panel were obtained by counting the number of times each covariate was included in the GAM model as a non-linear relationship.

11.4.2 Plotting the most common covariate combinations

Option 3 will plot the most common one, two, three and four covariate combinations in the GAM model (Figure 14).

The frequencies for these combinations are obtained by counting how many times a certain combination of covariates are found in the same GAM model. The top left panel in Figure 14 is the same as the top panel in Figure 13 and the other panels shows the four most common combinations of two, three and four covariates respectively.

11.4.3 The distribution of model sizes

Another piece of information we can get from a bootstrap of the GAM run are the sizes, i.e. the number of covariates, of the final GAM models for the bootstrapped data sets. Option 4 will display this distribution (Figure 15).

The idea here is to learn something about the typical number of covariates necessary to
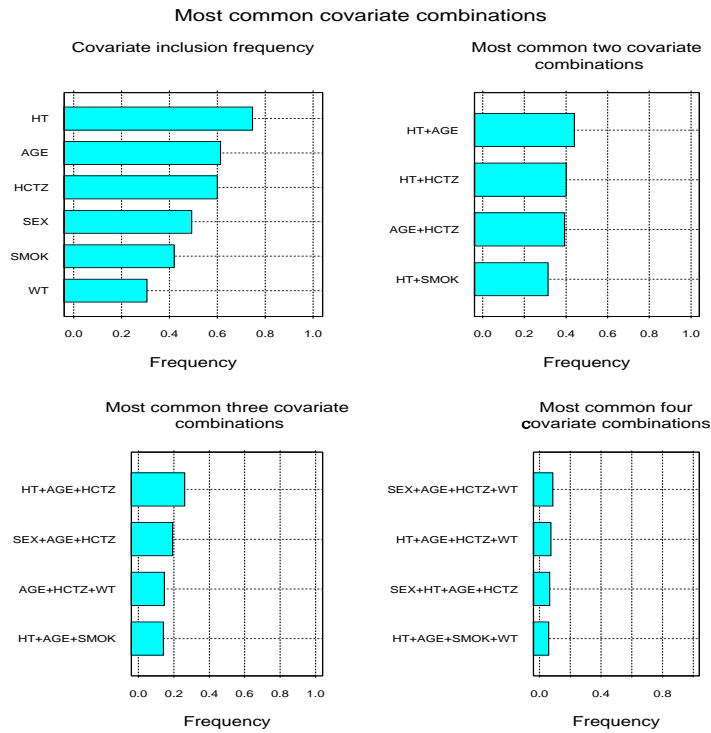
Figure 14: The most common one, two, three and four covariate combinations.

explain the variability in the parameter.

11.4.4 The stability of the inclusion frequencies

One of the important questions when running the bootstrap of the GAM is the number of bootstrapped data set to use (how to set the number of data sets to use is described in Section 11.5). To assess whether the number of data sets used in a particular run is enough to make the inclusion probabilities stable, option 5 and 6 can be used. Option 5 will plot the relative inclusion frequencies vs the bootstrap iterations (Figure 16) and option 6 will make the the same plot for the inclusion frequency of the non-linear covariate models. In these plots the panels are ordered from lower left to the top right according to the total inclusion frequency of each covariate.

11.4.5 Inclusion index of covariates

One interesting possibility with the bootstrap of the GAM is to investigate the *inclusion/exclusion interactions* between the covariates, i.e. will the inclusion of one covariate in the GAM model lead to the inclusion or exclusion of another covariate. Option 7 produces a plot of the *inclusion index* of covariates (Figure 17).
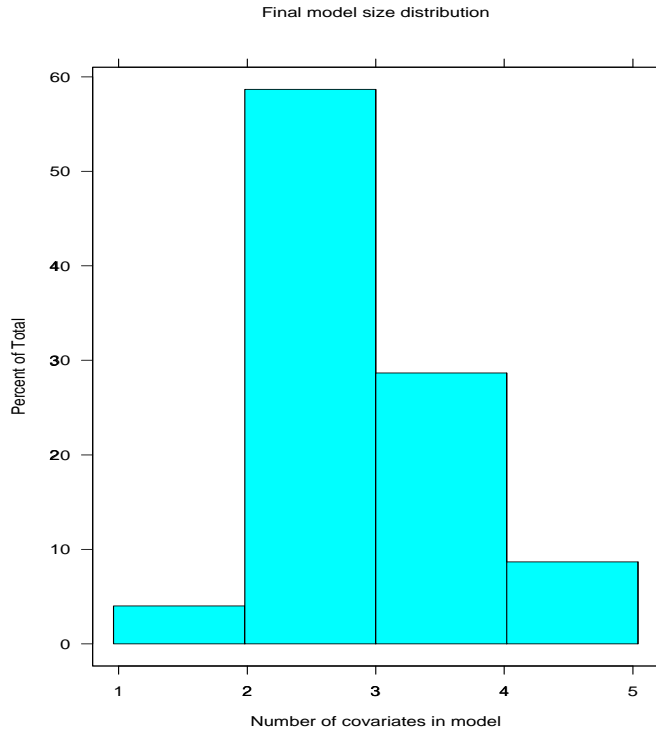
The inclusion index is given Eq. 1.

Figure 15: The distribution of the sizes of the final GAM models.

$$index = \frac{observed\ frequency - expected\ frequency}{expected\ frequency} \qquad (1)$$

where the *expected frequency* is the product of the relative inclusion frequencies of the two covariates in question and the *observed frequency* is the observed relative frequency of these two covariates appearing in the final GAM models together. The inclusion index will have a value greater than one if the combination of the two covariate are more common than expected and vice versa.

In Figure 17 each panel refers to one covariate. The inclusion indices of that covariate together with the other covariates are indicated by +-es. The panels are ordered from bottom left to top right according to the total inclusion frequency for each covariate as are the covariates on the y-axes.

Option 8 will produce a similar plot for the inclusion index of non-linear covariate models and the covariates (Figure 18).

11.4.6 Inclusion index of covariates and individuals

Similar to the inclusion index of covariates we can calculate the inclusion index for covariates/non-linear covariate models and individuals. The calculation of the index is done in the same way as for the covariate index (Eq. 1). Option 9 will produce plots of
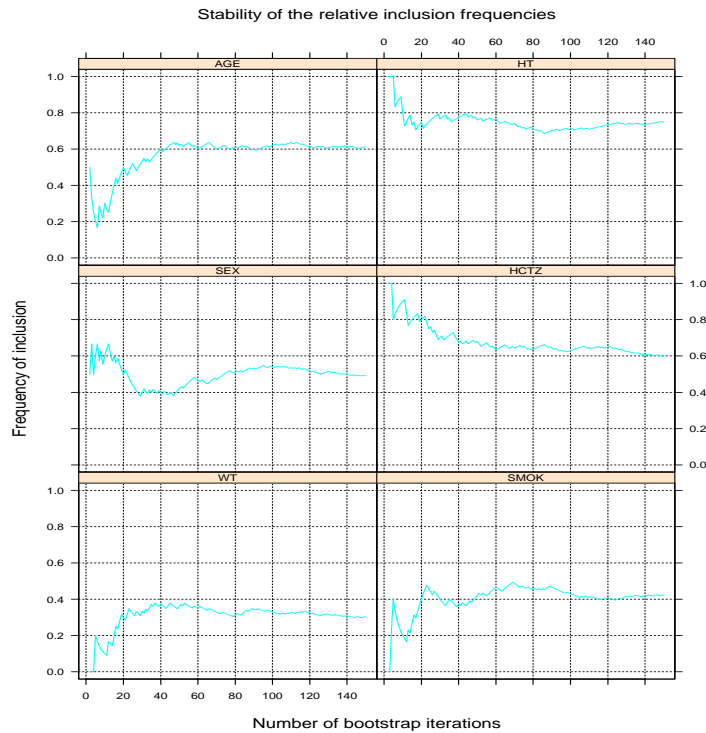
Figure 16: Stability of relative inclusion frequencies. Each panel shows the inclusion frequency vs the number of iterations for the covariate indicated in the strip above it.

the inclusion index of individuals and covariates.

In analogy with the other inclusion indices we can also plot the inclusion index of individuals and non-linear covariate models.

11.5   Settings for the bootstrap of the GAM

There are a number of ways that the user can influence the way the bootstrap of the GAM behaves. This is done in the BOOTSTRAP OF THE GAM SETTINGS MENU, which can be reached by selecting option 5 on the BOOTSTRAP OF THE GAM MENU:

```
BOOTSTRAP OF THE GAM SETTINGS MENU

1: Return to previous menu
2: List current settings
3: Estimate dispersion factor (on/off)
4: Specify starting model
5: Exclude individuals
6: Normalize to median (on/off)
7: Set maximum number of bootstrap iterations
8: Change convergence algorithm
9: Specify iteration to start check convergence at
```
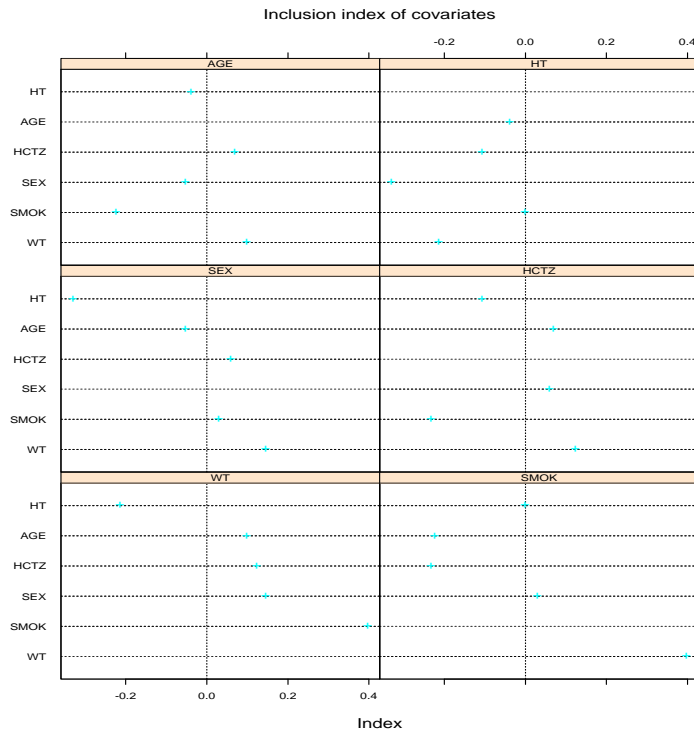
Figure 17: The inclusion index (see text) of covariates. Each panel refer to one covariate as indicated in the strip above it. The dashed horizontal line indicates the expected frequency

```
10: Specify at what interval to check the convergence
11: Set seed number
12: Use weights
Selection:
```

Option 2 is used to list the current settings. Options 3-7 and 13 are equivalent to the same settings in the GAM SETTINGS MENU (see Section 10.5) while options 8-12 are specific to the bootstrap of the GAM. The settings are in effect until S-PLUS is quit.

11.5.1 Specifying the maximum number of iterations

Option 8 sets the maximum number of bootstrap iteration to the value specified by the user. The default is 150. Note that this is not the same as specifying a fixed number of iterations, see Section 11.5.3.

11.5.2 Selecting the convergence algorithm

Option 9 is used to change convergence algorithm and convergence criteria. These algorithms are used by Xpose to determine when the number of iterations are enough
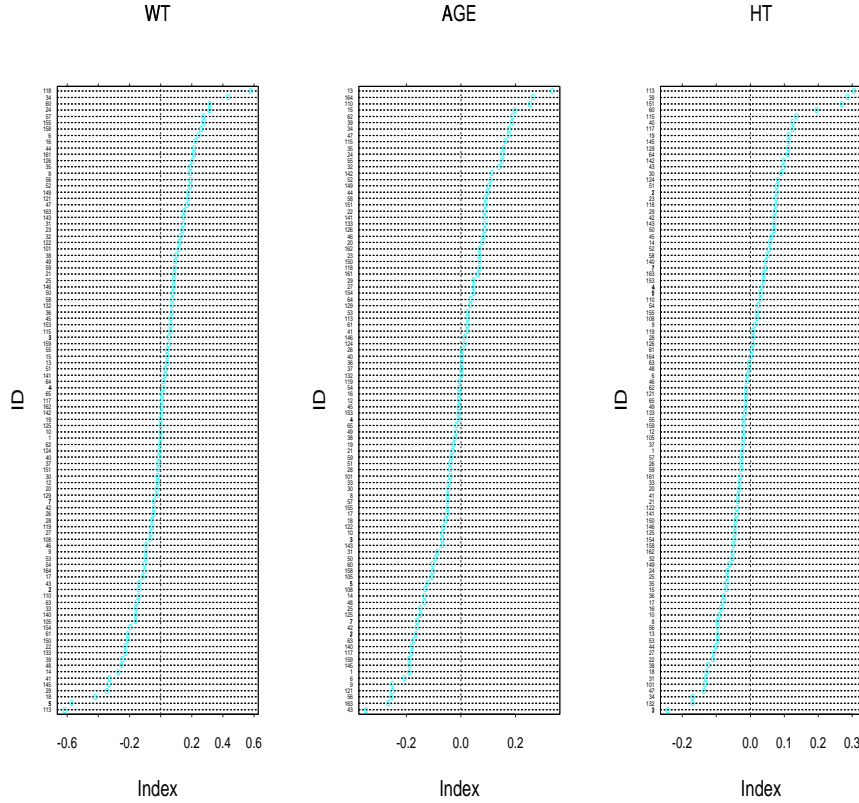
Inclusion index of non-linear models

Figure 18: A plot of the inclusion index of individuals and non-linear models. On the y-axis are the ID numbers and on the x-axis is the index. Each panel refers to one continuous covariate.

to satisfy the convergence criteria. The latter depends on the algorithm used.

The default algorithm in Xpose is the *fluctuation ratio*. This algorithm terminates the bootstrap of the GAM when the fluctuations in the relative inclusion probabilities of the covariates is lower than certain value. The fluctuation of the inclusion frequency for a specific covariate is determined over a certain interval (the default is the last 20 iterations, see Section 11.5.4) and the measure of fluctuation, F, is the ratio of the highest and lowest inclusion frequency in that interval. The overall fluctuation, F_{tot} is obtained by Eq. 2:

$$F_{tot} = \frac{\sum_{i=1}^{n} F_i \cdot P_i}{\sum_{j=1}^{n} P_j} \tag{2}$$

F_i is the fluctuation of the relative inclusion frequency in the interval of the ith covariate and P_i is the relative inclusion frequency of the ith covariate at the end of the interval.

47

The value of F_{tot}, which will be 1 if there are no fluctuations, are compared to the critical value, by default 1.04.

The fluctuation ratio will put more weight to the covariates with a high inclusion frequency, which is desirable if we are only interested in the overall stability of the covariates. If we are interested in the influence that individuals and/or other covariates exerts, it is necessary to make sure that also the covariates with low inclusion frequencies are found by the GAM enough times. This is to make sure that the inclusion indices (e.g. Eq. 1) do not rely on only a few occurrences of the covariate in the final GAM model. The second termination algorithm, *lowest absolute joint inclusion frequency*, makes sure that the individual that has been included (one or more times) in the bootstrapped data sets the least number of times, is included at least a certain number of times (z – lowest absolute joint inclusion frequency) together with a hypothetical covariate with a certain, hypothetical, inclusion frequency (w – lowest important relative inclusion frequency). Let p be the lowest relative inclusion frequency of all individuals at a certain iteration, j. (The relative inclusion frequency of an individual is the number of times an individual has been included in any of the bootstrapped data sets divided by the number of data sets.) The critical value for this algorithm is: $w \cdot p \cdot j$, which is compared to z. The default values of z and w are 25 and 0.2 respectively.

The convergence of the lowest absolute joint inclusion algorithm is only dependent on the random number permutation used in a specific bootstrap of the GAM problem. This means that Xpose could actually calculate the number of iterations needed for convergence, However, the bootstrap of the GAM is not implemented that way and Xpose checks the convergence at each interval as specified by the convergence checking interval (see Section 11.5.4).

11.5.3   Specifying an iteration to start checking convergence

Option 10 is used to specify the first interval at which the convergence should be checked. The default is 30. If the user wish to set the number of bootstrap iterations to a specific number, it can be done by setting the first convergence checking interval to the same as the maximum number of bootstrap iterations.

11.5.4   Specifying the interval between convergence checks

Option 11 is used to specify the interval between convergence checks (after the first check).

11.5.5   Specifying a seed number

The construction of bootstrapped data sets is a random process and, consequently, each bootstrap of the GAM run is unique. Option 12 can be used to specify a seed number for the random number generator. The seed number used for a bootstrap of the GAM run is given in the numerical summary of the results (see Section 11.3), this is true even if the seed number was not supplied by the user. To repeat a run exactly as a previous run, it is necessary to set the seed number to the same as the one used in the old run

12 Tree based modeling

Tree based models are another method to find covariates to include in the NONMEM model [20]. It has so far not been much used in the area of population pharmacokinetic/pharmacodynamic analysis. It offers some, at least theoretical, advantages over the GAM in that the fit is invariant to transformation of the explanatory variables (covariates) and automatically includes the possibility of identifying interactions between covariates. For a more detailed description of tree based models, please refer to *Statistical models in S* by Chambers and Hastie [12] or *Modern applied statistics with S-plus* by Venables and Ripley [21].

To fit a tree model in Xpose it is necessary to have covariates and at least one parameter defined in the current data base. Select option 6 from the `MAIN MENU` followed by option 6 in the `COVARIATE MODEL MENU`. This will lead to the `TREE MENU`:

```
TREE MENU

1: Return to previous menu
2: Fit tree
3: Plot tree
4: Find optimal tree size
Selection:
```

12.1 Fitting a tree model

To fit a tree model, select option 2 from the `TREE MENU`. If there is more than one parameter defined in the current data base, Xpose will prompt the user for a parameter name.

12.2 Plotting a tree

To plot the fitted tree select option 3 from the `TREE MENU`. Xpose saves the fitted trees so it is also possible to specify an old tree fit. The Xpose tree objects are named with a combination of `tree.xpose.`, the parameter and the run number, e.g. `tree.xpose.CL.1`. The user will be prompted for the size of the tree to be plotted (see the next section). Pressing return as the answer to this prompt will produce a tree similar to Figure 19.

12.3 Pruning trees

The tree in Figure 19 is an unpruned tree, i.e. it is a result of growing the tree until it is no longer possible to make any more splits in the terminal branches, a point which is reached when there are less than 10 points in each of the terminal branches. This is likely to induce over fit and it is usually necessary to prune the tree[21], i.e. the user has to specify which size of the tree that is relevant (hence the prompt when plotting a

---

[21]In Xpose 1.1 it was not at all possible to prune the fitted trees making the tree module in that version almost useless.
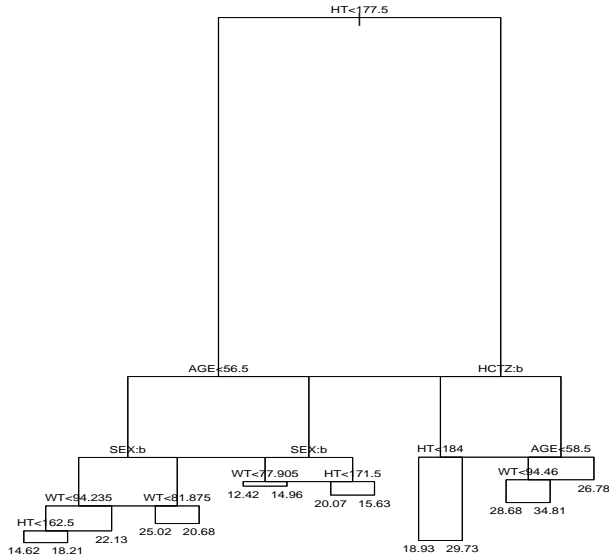
Example fit of a tree model

Figure 19: Plot of a tree fit. The length of the vertical lines are proportional to the importance of the split

fitted tree). Some guidance to the importance of a specific split is given by the length of the branches in the tree plot but this can, however, not be used as an indicator of the relevant size of a tree. One approach to find the relevant tree size is to use a cross validation procedure. First, the data set is divided into x number of groups, second, x number of trees are grown, each with one of the x groups omitted from the growth process and third, the fitted trees are used to predict the group of data not used to grow the tree. The improvement in the fit (measured by the deviance), at each node of the tree, is averaged over the x groups and the results are plotted. The key point here is that the improvement in fit will only continue up to a certain point after which the deviance will increase, i.e. the plot of the averaged deviance vs tree size will often exhibit a minimum at some point, which is the suggested optimal tree size. This procedure will not always work but since it a random process it is possible to repeat it until a satisfactory minimum is found. The last sentence implies a degree of subjectivity in the determination of the optimal tree size. To minimize this, Xpose will repeat the cross-validation six times and display the results. An example of cross-validation plots are shown in Figure 20 and the resulting pruned tree (size 6) are shown in Figure 21.

Admittedly, this is not satisfactory but is better than using the unpruned trees. With this implementation it seems as if the the tree models are most useful for exploratory analysis.
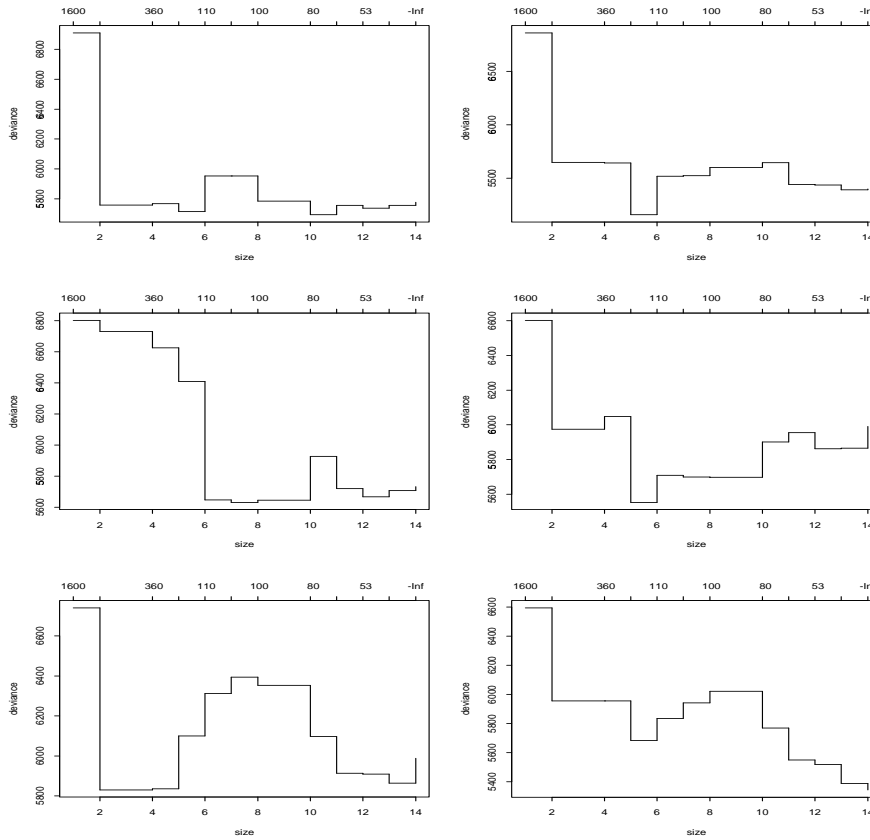
Figure 20: Results from exploring the optimal tree size using cross-validation. On the y-axis is the deviance and on the x-axis are the corresponding tree sizes.

13   Printing and exporting graphics

The plots in Xpose are always displayed on screen[22] (except the run summary and run record). When the plot has been drawn, the user is asked if he or she wants to print the plot or export to a file:

```
Do you want to print/export the graph? n(p=print,e=export)
```

Pressing n or return accepts the default – not to print or export the plot. Pressing p will send the plot to the printer and pressing e will export the plot to a file.

When a plot is printed or exported the user has the possibility to customize the plot title and/or the axis labels (the latter is not available when there are multiple plots per page). To omit the plot title completely, type n at the prompt.

---

[22]In Xpose 1.1 the user had select if the plots should be displayed on screen or sent to the printer before making the plot.
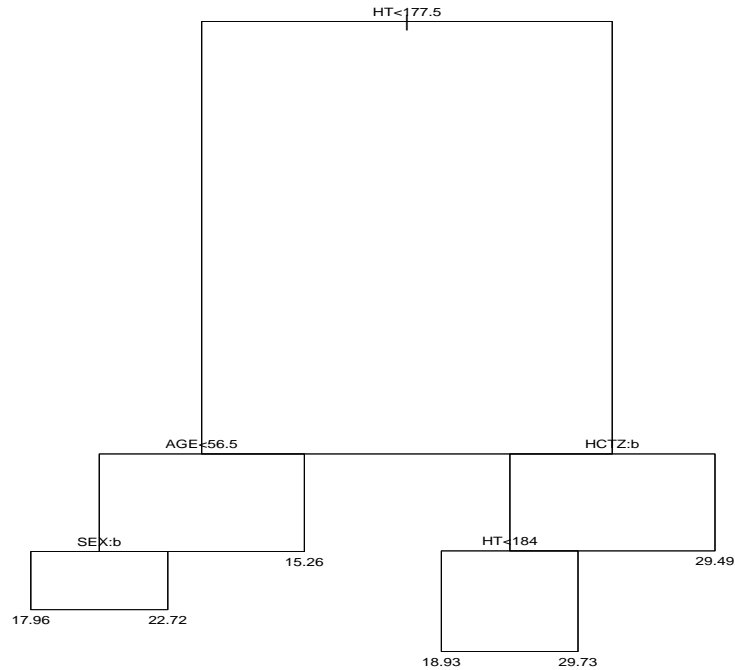
Example of a pruned tree

Figure 21: Pruned tree of size 6.

```
Do you want to print/export the graph? n(p=print,e=export)p
Type new title (return for default):
Final model
Type new x-axis title (return for default):
Time (h)
Type new y-axis title (return for default):
Weighted residuals
```

When printing a plot it is sent to the systems default printer. To change that behavior on a UNIX system it is necessary to use a command similar to `ps.options(command="lp -d my.printer")`, please refer to the S-PLUS manuals for further details. On a PC the user can specify the printer in the usual Windows way.

The format of an exported plot depends on the platform. On UNIX machines the default format is postscript and the plot will end up in a file called `xp.ps` in the current directory. On PCs the plots are exported to the `clipboard` in windows metafile format, i.e. when a plot has been exported it is necessary to switch to another Windows

application, e.g. Microsoft Word, and paste the plot in a document. Note, on a PC it is generally not possible to export multi-page plots.

As per default plots are exported and printed in black and white. To use color, go to the DATABASE MANAGEMENT MENU and select option 26. This option will toggle color printing and exporting on and off.

If the user wishes to change the behavior of the printing and/or exportation it can be done in the Xpose functions `ask.print` and `set.xp.print.dev`.

14   Altering the the way Xpose draws plots

Section 2.5 describes how the Xpose data variables can be re-defined. For some variables it is also possible to undefine them. To undefine a variable, select the relevant option on the MANAGE DATABASES MENU and type NULL at the prompt (note the capitalization). Only some data variables can successfully be undefined: *idlab*, *wres*, *iwres*, *pred*, *ipred* and *dv*.

Undefining *idlab* will suppress the usage of ID numbers as plotting symbols.

Undefining either or both of *pred* and *ipred* will suppress the plotting of the corresponding curve in the `Individual plots` option in the GOODNESS OF FIT PLOTS MENU. This also works in the `Predictions vs dependent variable` option in the same menu.

Undefining either or two of *pred*, *ipred* and *dv* will suppress the corresponding plot(s) in the `Predictions vs independent variable` in the STRUCTURAL MODEL DIAGNOSTICS MENU.

Undefining either of `iwres` or `wres` will suppress the corresponding plot in options 5 and 6 on the RESIDUAL ERROR MODEL DIAGNOSTICS MENU

15   Making plots of a subset of the data

Sometimes it is desirable to concentrate on a subset of the data, for example the pharmacodynamic data from a fit of a simultaneous pharmacokinetic/ pharmacodynamic model . This can be done in Xpose by specifying two Xpose data variables: *flag* and *curflag* (current value of flag). The *flag* variable can be specified in the DATABASE MANAGEMENT MENU or set to a default value by the use of a `mutab` (see Section 2.1.4). The *flag* variable determines which data item to be used to define the subsets of the data. The *flag* variable can, however be set without making Xpose produce plots for a certain subgroup[23]. To use a certain subgroup it is also necessary to specify the *current value of the flag variable*. (See below for exceptions.) This done with option 14 in the DATA BASE MANAGEMENT MENU:

```
Type the value to be used as the current flag value. Possible values
are: 2 1 (q to leave it as is, n to unspecify):
```

---

[23]Using a mutab in Xpose 1.1 made it possible to select plots from a certain menu with plots that knew how to handle multiple response variables, i.e. it was enough to specify the "flag" variable in the mutab .

It is necessary to specify the *flag* variable before setting the current value of the flag. In the above example the *flag* variable has been set to SEX. When selecting the current value of the flag, Xpose shows the possible values the current value can have, in this case 1 and 2, and gives the user the options to leave it as is, that is, not to change the current value of the flag, to unspecify it, i.e. to go back to look at all data at the same time, and to specify any of the possible values. Note! The *flag* variable will be treated as a factor variable, meaning that it is not a good idea to set it to a continuous variable like AGE.

When the *flag* and *curflag* variables are set, all plots and analyses will be made based on the subset of the data specified by these variables.

If the *flag* is set and not the *curflag* there are a few plots that offers to use that *flag* variable as the conditioning variable. Specifically, this is the coplots on the STRUCTURAL MODEL DIAGNOSTICS MENU and the RESIDUAL ERROR MODEL DIAGNOSTICS MENU (see Fig. 1) that corresponds to the plots in the Basic goodness of fit plots and Additional goodness of fit plots that uses the covariates as the conditioning variable. These options will offer to use the *flag* variable as the conditioning variable instead of the data items defined by the covariate variables.

References

[1] L. Aarons. Sparse data analysis. *Eur. J. Metab. Pharmacokin.*, 18:97–100, 1993.

[2] T. H. Grasela, E. J. Antal, R. J. Townsend, and R. B. Smith. An evaluation of population pharmacokinetics in therapeutic trials. Part I. Comparison of methodologies. *Clin. Pharmacol. Ther.*, 39(6):606–612, 1986.

[3] L. B. Sheiner and S. L. Beal. Evaluation of methods for estimating population pharmacokinetic parameters II. Biexponential model; experimental pharmacokinetic data. *J. Pharmacokin. Biopharm.*, 9(4), 1983.

[4] M. O. Karlsson, E. N. Jonsson, C. Wiltse, and J. R. Wade. Assumption testing in population pharmacokinetic models: Illustrated within an analysis of moxonidine pk data. Submitted.

[5] StatSci, a division of Mathsoft, Inc, Seattle. *S-PLUS guide to statistical and mathematical analysis, Version 3.3*, 1995.

[6] W. S. Cleveland. *Visualizing data*. Hobart press, Summit, New Jersey, USA, 1993.

[7] W. S. Cleveland. *The elements of graphing data*. Hobart press, Summit, New Jersey, USA, 1985.

[8] J. W. Tukey. *Exploratory data analysis*. Addison-Wsley, Reading, Massachusetts, USA, 1977.

[9] R. A. Becker and W. S. Cleveland. *S-PLUS Trellis Graphics user's manual*. Seattle:MathSoft, Inc, Murray Hill: Bell Labs, 1996.

[10] J. W. Mandema, D. Verotta, and L. B. Sheiner. Building population pharmacokinetic-pharmacodynamic models. I. models for covariate effects. *J. Pharmacokin. Biopharm.*, 20(5):511–528, 1992.

[11] J.W. Mandema, D. Verotta, and L. B. Sheiner. Building population pharmakokinetic-pharmacodynamic models. In D. Z. D'Argenio, editor, *Advanced methods of pharmacokinetic and pharmacodynamic systems analysis, Volume 2*. Plenum Press, 1993.

[12] J. M. Chambers and T. J. Hastie, editors. *Statistical models in S*. Chapman & Hall, London, 1993.

[13] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*. Chapman and Hall, New York, 1990.

[14] E. N. Jonsson amd M. O. Karlsson. Identification of factors that influences the importance of covariate relationships in population pharmacokinetic/pharmacodynamic models. *Submitted*, 1997.

[15] A. C. Davison and E. J. Snell. Residuals and diagnostics. In D. V. Hinkley, N. Reid, and E. J. Snell, editors, *Statistical theory and modelling: In honor of Sir David Cox*. Chapman & Hall, London, 1991.

[16] B. Efron and R. J. Tibsgirani. *An introduction to the bootstrap*. Chapman and Hall, New York, 1993.

[17] R. Mick and M. J Ratain. Bootstrap validation of pharmacodynamic models defined via stepwise linear regression. *Clin. Pharmacol. Ther.*, 56(2):217–222, 1994.

[18] W. Sauberei and M. Schumacher. A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistics in medicine*, 11:2093–2109, 1992.

[19] E. I. Ette. Stability and performance of a population pharmacokinetic model. *J. Clin. Pharmacol.*, 37:486–495, 1997.

[20] E. I. Ette and T. M. Ludden. Population pharmacokinetic modelling: The importance of informative graphics. *Pharm. Res.*, 12(12):1845–1855, 1995.

[21] W. N. Venables and B. D. Ripley. *Modern applied statistics with S-PLUS*. Springer-Verlag, New York, 1994.