Casses Software
Version 2.0.0

User Manual

Assistance: casses@cemagref.fr

Website: https://casses.cemagref.fr/

Contents

I Installing Casses

There are two versions of the software **Casses**: **Casses-Mono** and **Casses-Multi**. In this manual, the term **Casses** is used without distinction and applies to both versions.

I.1 *Advised configuration*

Casses uses the virtual Java machine to operate. Therefore, you also need to install Java version 1.5 or higher (software can be freely downloaded at http://www.java.com/fr/).

Memory: 512 Mb minimum, 1024 Mb or more recommended.

For more details, please refer to annex XIII.6.

I.2 *Installation under Windows operating system*

The software is protected by a USB key “Actikey”

Important: Install the software **BEFORE** inserting the key.

Launch by double clicking Cassesaaaa_x.y.z.exe (aaaa meaning Mono or Multi and x.y.z, the number of the software version) and follow the instructions of the installation program:

- Choose the installation language (English or French).
- Accept the licence agreement.
- Choose the installation directory (by default C:\Program Files\Cemagref\Cassesaaaa).
- Choose the Start-up menu directory (by default Cassesaaaa).
- Click on <Install> to confirm the choices and install **Casses** and “Ithea”.
- Click on <Finish>.

Casses can then be launched from the Start-up menu or by double clicking the file lanceur.exe in the chosen installation directory.

The language used by **Casses** depends on the regional and language options selected by the user in the Windows Control Panel:

- If the language option is French or German, **Casses** is installed in French or German respectively.
- If the language option is neither French nor German, **Casses** is installed in English.

II Formatting the data to be imported

II.1 General

The imported files are of text **csv** (Comma-separated values) format with the **semi-colon (;) as the data separator**.

Whatever the version of **Casses**, at least two files are necessary:

- The pipes file
- The breaks file

For **Casses-Multi**, an additional file is needed:

- The networks file

Each of these three files is structured in the same fashion:

- An area for optional comments in the first rows
- Four rows used to describe the data present and their format
- The correctly formatted data (one row per record).

II.2 Pipes File

II.2.i Structure:

#Title, *(free text, suggested to be used for the Project name)*
#Comment 1 *(free text)*
#Comment 2 *(free text)*
#….
#Comment n *(free text)*

IDT	IDR	DDP	DHS	LNG	MAT	CHAR1 *(short name)*	CHAR2 *(short name)*	…	CHARp *(short name)*
Pipe ID	Network ID	Installation date	Removal date	Length	Material	Long name 1	Long name 2	…	Long name p
	QUAL	DATE	DATE	QUAN	QUAL	QUAL	QUAN	…	DATE
		y	d/m/y	m		*(Empty if qualitative)*	*(Unit if quantitative)*	…	*(Format if date)*
wxv987	18barjols	1932	24/10/2006	34.87	CI	Urban	10.23	…	28/06/1996

✓ At the beginning of the file (green area), comments can be added after an initial # character); the text is free but in order to be correctly interpreted it must not contain any semi colons or inverted commas. The first comment row is used as the default value for the Project name.

✓ **The first row without #** (yellow row) contains the short name for each of the data associated with the pipe. It allows the data and characteristics in the file to be identified; therefore there is a uniqueness constraint for each value. It comprises:

- The labels of compulsory or predefined data that are imposed. The order of these data is unimportant:
  - <u>**IDT**, Pipe Identification</u>: Can be numeric or alphanumeric
  - <u>**DDP**, Installation date</u>: This can be either a precise date or a year. In the case of the latter, the 1st of January of that year is used as the installation date.
  - <u>**LNG**, Length</u>: Quantitative data expressed in metres. Decimal value.
  - <u>**MAT**, Material</u>: Qualitative data.
  - <u>**IDR**, Network identification</u>: The network identification is only compulsory in the multi-network version. When it is present in the mono-network version, it is considered as an additional qualitative characteristic.
  - <u>**DHS**, Date removed from service</u>: This usually results from the abandonment of a pipe but may also correspond to an important rehabilitation. This can be either a precise date or a year. In the case of the latter, the 31st of December of that year is used as the removal date. This characteristic is not compulsory; when it is not present, all the pipes are considered to still be in service.
- The labels of additional characteristics. They are unlimited in number but the following constraints must be respected:
  - Upper or lowercase characters from the Latin alphabet are authorised without accents as well as numbers and underscore (_).
  - The first character cannot be a number.
  - No spaces
  - Cannot be the same as any compulsory data label
  - The number of characters for short names is limited to 8 (this label will be use for writing functions allowing covariates to be created from characteristics as well as for column titles in the tables).
  - To be able to be used for the creation of a new covariate involving a mathematical expression, a quantitative covariate must have a short name different from the formula names below:

Functions

sin	cos	tan	asin	acos	atan
sinh	cosh	tanh	asinh	acosh	atanh
ln	log	abs	rand	sqrt	erf
erfc	gamma	exp	cot	log2	

  - The short name DIA is reserved for the diameter. If an additional characteristic uses DIA as a label then it is imperative that the values are quantitative and greater than zero.
  - The short name TCM is reserved for the characteristic “maximum failure rate” and is proposed to be used in future developments of the software.
  - When additional data is a date given by the year, the day taken into account is the 1st of July of that year.

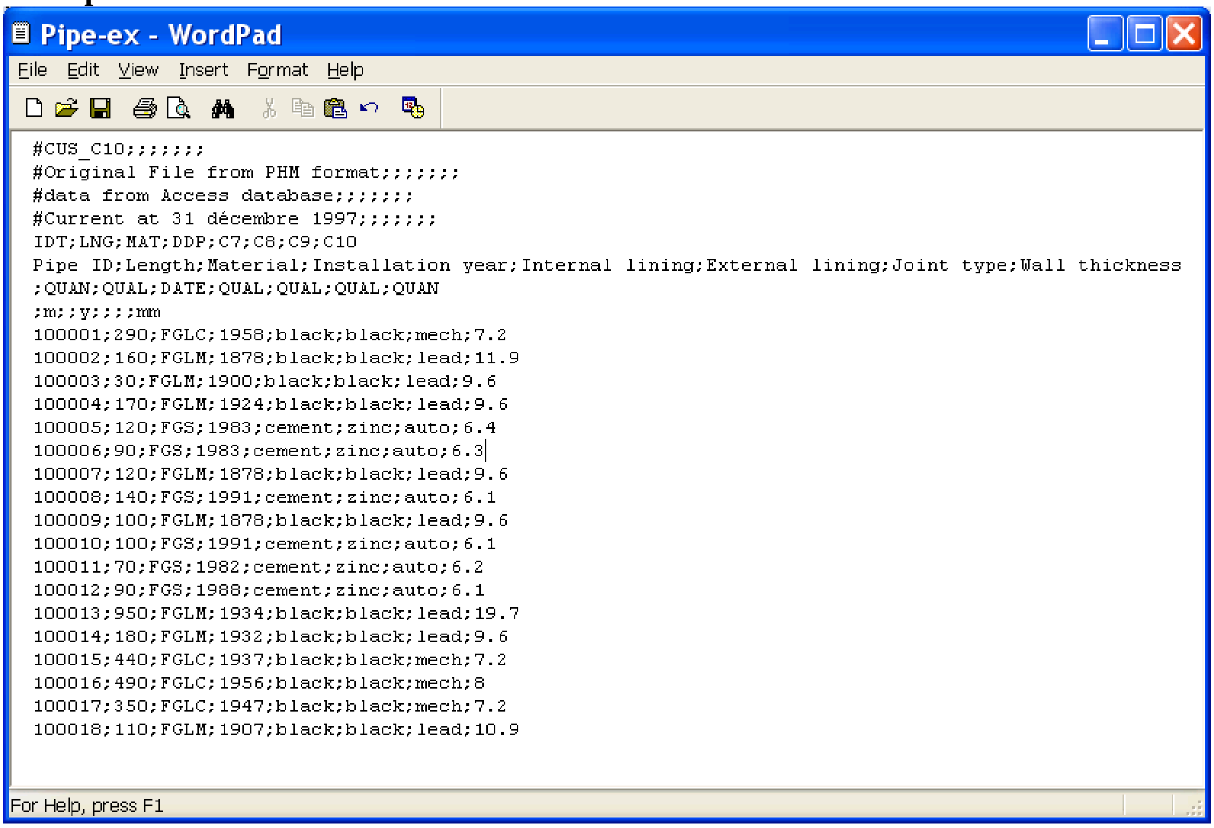
✓ **The second row** (salmon-coloured row) contains a long name associated with the short name given in the preceding row. The long name can be chosen freely with any characters except semi-colons and may contain spaces. It should be intelligible and can be used in the results.

✓ **The third row** (blue row) contains the type of data, chosen from one of three possibilities, QUAL for qualitative, QUAN for quantitative or DATE **(must be uppercase)**. This can be omitted for compulsory data.

✓ **The fourth row** (pink row) specifies the type of data:

- For quantitative data, the unit is given. For numeric data, both the comma and full stop can be used as decimal separators. Spaces between figures are allowed but the presence of two separators or a currency symbol is forbidden.
- For data in the form of a date, it specifies the format of the date from one of the following:
  - "a" or "y", the data is a year with 4 figures.
  - "j/m/a" or "d/m/y", the date is expressed, in order, by day, month and year (in figures) separated by slashes (/). The year has four figures.
  - "m/j/a" or "m/d/y", the date is expressed, in order, by month, day and year (in figures) separated by slashes (/). The year has four figures.
  - "a/m/j" or "y/m/d", the date is expressed, in order, by year, month and day (in figures) separated by slashes (/). The year has four figures.
- For qualitative data, the values are empty. Qualitative data values may contain all characters with the exception of semi-colons, and may even contain exclusively numbers which will be considered as text. Qualitative data can contain up to 20 different values (modalities).

II.2.ii Example:




```
#CUS_C10;;;;;;;
#Original File from PHM format;;;;;;
#data from Access database;;;;;;
#Current at 31 décembre 1997;;;;;;
IDT;LNG;MAT;DDP;C7;C8;C9;C10
Pipe ID;Length;Material;Installation year;Internal lining;External lining;Joint type;Wall thickness
;QUAN;QUAL;DATE;QUAL;QUAL;QUAL;QUAN
;m;;y;;;;mm
100001;290;FGLC;1958;black;black;mech;7.2
100002;160;FGLM;1878;black;black;lead;11.9
100003;30;FGLM;1900;black;black;lead;9.6
100004;170;FGLM;1924;black;black;lead;9.6
100005;120;FGS;1983;cement;zinc;auto;6.4
100006;90;FGS;1983;cement;zinc;auto;6.3
100007;120;FGLM;1878;black;black;lead;9.6
100008;140;FGS;1991;cement;zinc;auto;6.1
100009;100;FGLM;1878;black;black;lead;9.6
100010;100;FGS;1991;cement;zinc;auto;6.1
100011;70;FGS;1982;cement;zinc;auto;6.2
100012;90;FGS;1988;cement;zinc;auto;6.1
100013;950;FGLM;1934;black;black;lead;19.7
100014;180;FGLM;1932;black;black;lead;9.6
100015;440;FGLC;1937;black;black;mech;7.2
100016;490;FGLC;1956;black;black;mech;8
100017;350;FGLC;1947;black;black;mech;7.2
100018;110;FGLM;1907;black;black;lead;10.9
```

II.3 Breaks file

II.3.i Structure:

#Title *(free text)*
#Comment 1 *(free text)*
#Comment 2 *(free text)*
#….
#Comment n *(free text)*

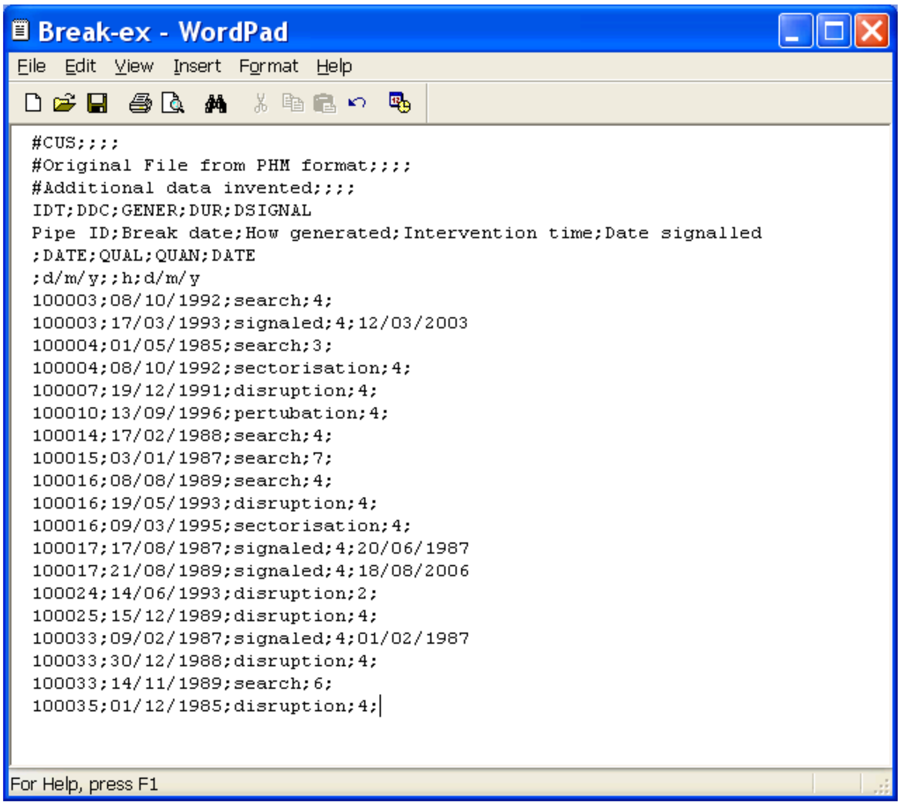
IDT	DDC	DAT1 *(short name)*	DAT2 *(short name)*	…	DATp *(short name)*
Pipe ID	Break date	Long name 1	Long name 2	…	Long name p
	DATE	QUAL	QUAN	…	DATE
	d/m/y	*(Empty if qualitative)*	*(Unit if quantitative)*	…	*(Format if date)*
wxv987	17/12/1998	Signalled	4	…	12/12/1998

The rules are generally the same as for the pipes file. The short names imposed for the two compulsory data fields are:

– **IDT**, <u>Pipe identification</u>: The pipe on which the break occurred.
– **DDC**, <u>Break date</u>: In fact, this is generally the day of the repair. For calculation purposes, no pipe can have more than one break on the same day. If the break date is only given by the year (unadvisable), the 1st of July of that year is used in the calculations.

The breaks file can also include additional quantitative or qualitative data. The constraints of uniqueness and for the short name must be respected.

II.3.ii Example:


Break-ex - WordPad

File Edit View Insert Format Help

```
#CUS;;;;
#Original File from PHM format;;;;
#Additional data invented;;;;
IDT;DDC;GENER;DUR;DSIGNAL
Pipe ID;Break date;How generated;Intervention time;Date signalled
;DATE;QUAL;QUAN;DATE
;d/m/y;;h;d/m/y
100003;08/10/1992;search;4;
100003;17/03/1993;signaled;4;12/03/2003
100004;01/05/1985;search;3;
100004;08/10/1992;sectorisation;4;
100007;19/12/1991;disruption;4;
100010;13/09/1996;pertubation;4;
100014;17/02/1988;search;4;
100015;03/01/1987;search;7;
100016;08/08/1989;search;4;
100016;19/05/1993;disruption;4;
100016;09/03/1995;sectorisation;4;
100017;17/08/1987;signaled;4;20/06/1987
100017;21/08/1989;signaled;4;18/08/2006
100024;14/06/1993;disruption;2;
100025;15/12/1989;disruption;4;
100033;09/02/1987;signaled;4;01/02/1987
100033;30/12/1988;disruption;4;
100033;14/11/1989;search;6;
100035;01/12/1985;disruption;4;
```

For Help, press F1

II.4 Networks file

This file is only necessary for **Casses-Multi**.

II.4.i Structure:

#Title *(free text)*
#Comment 1 *(free text)*
#Comment 2 *(free text)*
#….
#Comment n *(free text)*

IDR	NRE	BRSD	BRED
Network ID	Network name	Break record start date	Break record end date
		DATE	DATE
		d/m/y	d/m/y
Barj2002	Network of Barjols	01/06/2000	31/12/2002

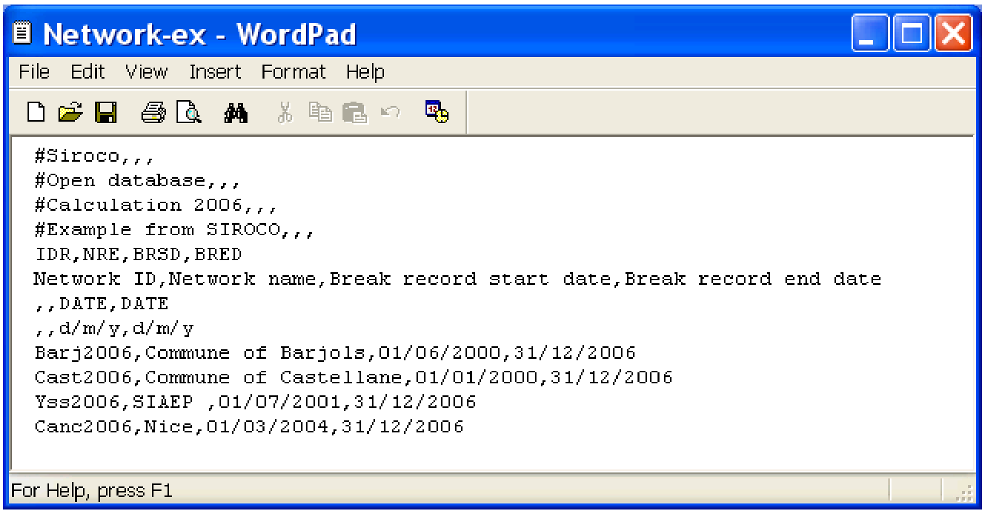
The rules are generally the same as for the pipes and breaks files. The short names imposed for the four compulsory data fields are:

– **IDR**, Network identification
– **NRE**, Name of Network
– **DDE**, Date of start of break records for the network: If the date given is a year, the 1st January will be used.

– **DFE**, Date of end of break records for the network: If the date given is a year, the 31st December will be used. This date must be after the data recording start date.

The networks file shouldn't contain any additional data; when present they will be ignored.

II.4.ii Example:




```
#Siroco,,,
#Open database,,,
#Calculation 2006,,,
#Example from SIROCO,,,
IDR,NRE,BRSD,BRED
Network ID,Network name,Break record start date,Break record end date
,,DATE,DATE
,,d/m/y,d/m/y
Barj2006,Commune of Barjols,01/06/2000,31/12/2006
Cast2006,Commune of Castellane,01/01/2000,31/12/2006
Yss2006,SIAEP ,01/07/2001,31/12/2006
Canc2006,Nice,01/03/2004,31/12/2006
```

II.5 Remarks concerning the creation of csv files

The .csv files are text files and can therefore also be read and modified with software such as: "Notepad", "WordPad", "Microsoft Word", "OpenOffice.org Writer", etc.

Most often, data derive from databases (possibly linked to GIS) and are presented in the form of tables. Several programs allow the creation, opening or modification of .csv files, notably "Microsoft Excel", "OpenOffice.org Calc" and "Microsoft Access" *(Note: OpenOffice.org Base doesn't recognise the .csv format).*

In practice, the creation of an input file may require the creation of an intermediate file of .dbf format (for example for data stored in the ArcView GIS).

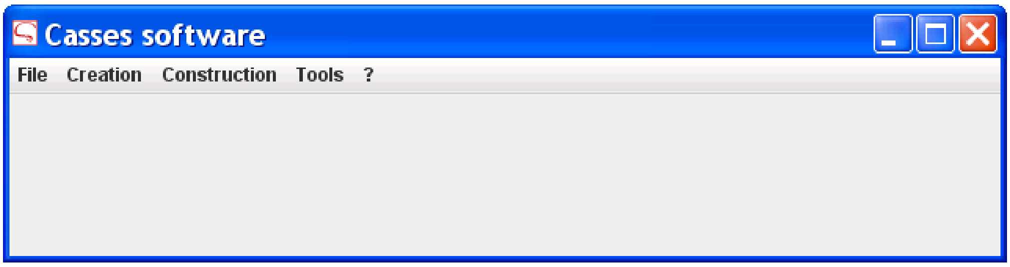
Experience shows that the repeated manipulation of files can lead to formatting errors or the alteration of data. Therefore, it is best to be careful and avoid wherever possible using different software for the same file. Among other reports:

- The opening of a .csv file with Microsoft Excel does not give the same results depending on whether it is opened by double clicking on the file (lines of file as text in the first column of the worksheet) or by opening it from Excel (table with a different field per column). The second method is advised.
- A .csv file created with OpenOffice Calc from a .dbf file doesn't open in the correct format with Excel (the problem can be resolved by inserting a comment row at the beginning of the .csv file).
- If a .dbf file is created in Excel, the field names are truncated at 11 characters.

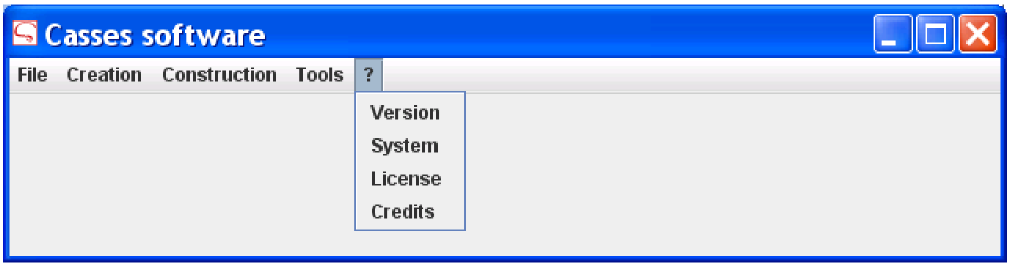
- With OpenOffice Calc, whatever the regional options selected in Windows, if the language option is French, the numeric separator is always the comma (to have the full stop, it is necessary to change to English).
- For the opening or saving of .csv files, OpenOffice Calc systematically shows a dialog box agreeing to select the semi colon as field separator and delete the proposed text separator.
- Regional options of the machine can lead to incorrect reading of .csv file in Excel. For example, in Norway, the date separator is the full stop meaning that certain decimal numbers are interpreted as dates (e.g. 02.10 interpreted as 10th February).

III Configuring Casses

Launching **Casses** results in a window with 5 menu items being displayed.



III.1 Menu "?"



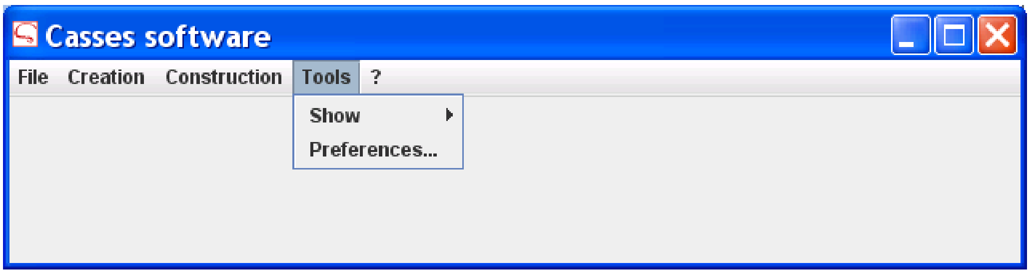
"Version" displays information concerning the version of **Casses** installed on your computer as well as information relevant to the protection key that you are using.

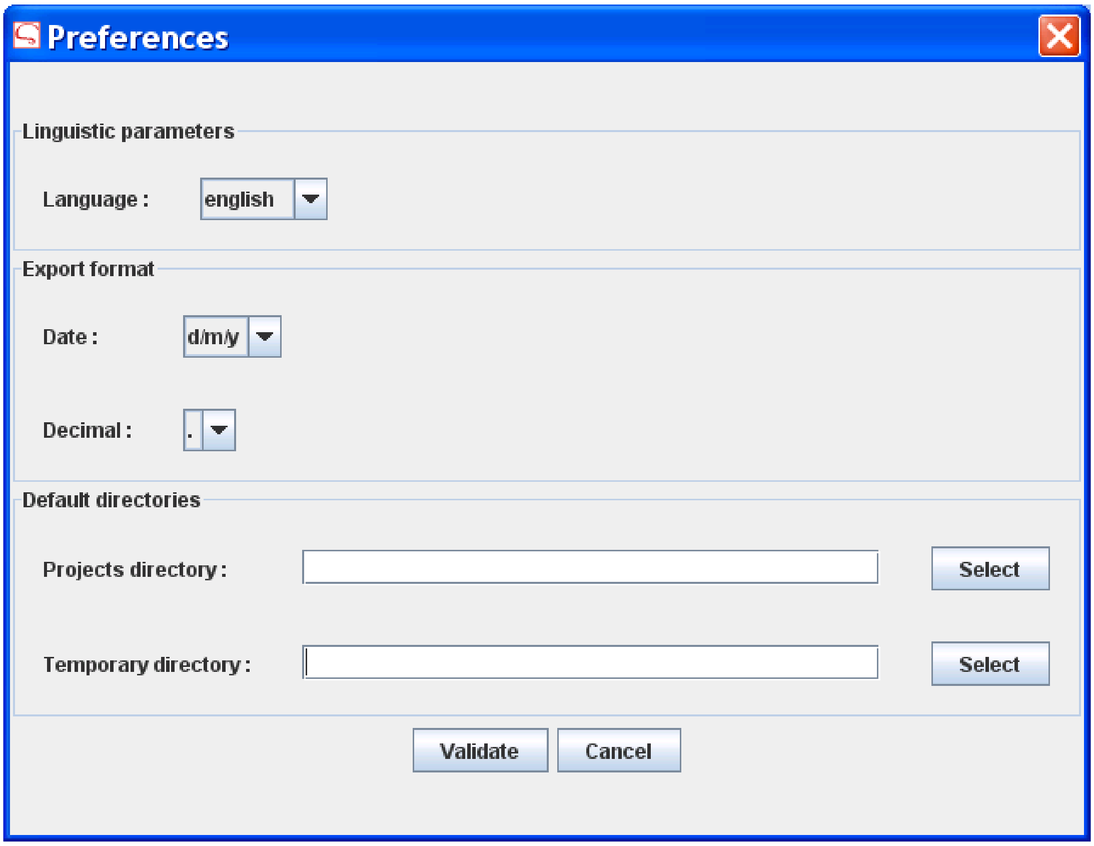
"System" informs you of the version of Java installed on your computer and on the memory allocated to running **Casses**.

"Licence" displays the user licence of **Casses** to which you have agreed.

"Credits" mentions the "freeware" used by the software and their appropriate licences.

III.2 Préférences



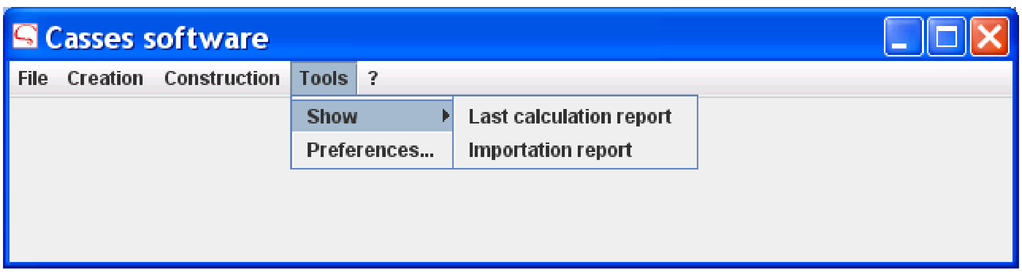


Casses is, by default, in the language of your operating system. However, you have the possibility to select another language. The modification requires restarting **Casses**.

The display of **Casses** uses the date and decimal separator formats defined by the operating system of your computer. However, the input files can use any of the accepted formats so long as they are specified in the file. For exportation, these formats can be chosen in the Preferences dialog.

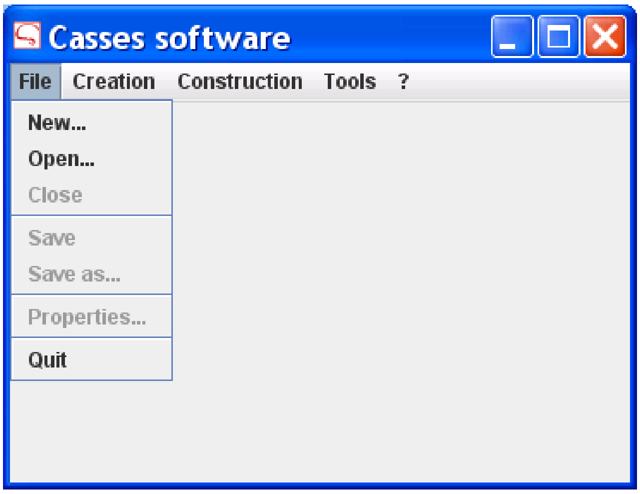
Preferences concerning both the choice of default folder for opening and saving projects and for temporarily storing importation and calculation reports can also be made.

It should be noted that files stored in the temporary directory are deleted and replaced after each new importation and each new calculation. These files can be accessed directly by the Tools menu:



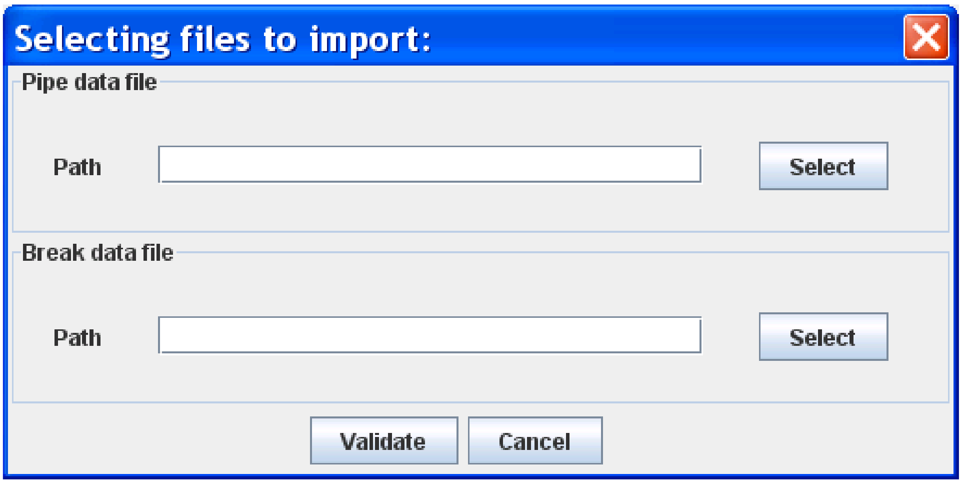
IV Starting a Project

IV.1 Creating a new Project



IV.1.i Casses-Mono

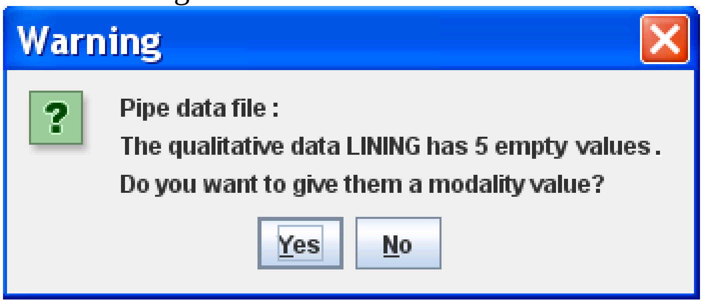
A dialog box asks you to identify the pipes and breaks files that you wish to use as well as their locations.



An initial series of importation tests is performed. If needs be, dialog boxes appear inviting corrective action:

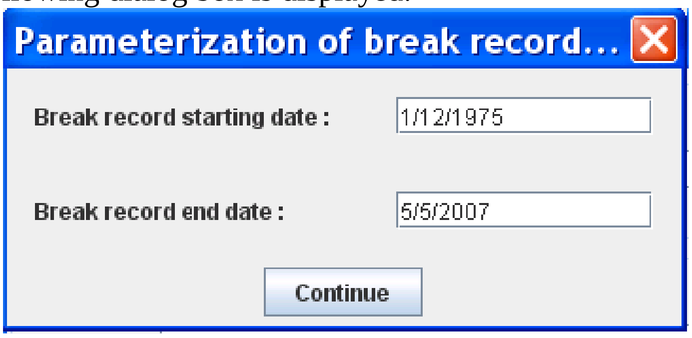
- By selecting "Yes", the blank values are replaced by the Value "Empty".
- By selecting "No", the data will not be used in the Project.

Example of a corrective dialog box:



If the initial series of tests concludes that the files are invalid then the importation is interrupted and the importation report is displayed.

Otherwise the following dialog box is displayed:



You must indicate the period during which the breaks presented in the breaks file have been recorded on the network. By default, the software proposes the day before the first break and the day after the last break in the breaks file.

A new series of tests is carried out.

If no anomalies are detected during the various tests then you will be able to access the main screen.

Otherwise, the importation report is displayed.

Each anomaly detected is displayed in a table with the following column labels: CFT, CAN, LIB, SEV, NCG, IDX, DDC, NLI

CFT is the code of the file(s) treated. It can contain the following values:

- **T** for the pipes file
- **C** for the breaks file
- **TC** for the linking of the pipes and breaks files
- **R** for the networks file (**Casses-Multi** only)
- **RT** for the linking of the pipes and networks files or in the case of mono-network, the coherence of pipes and breaks with the data recording period (RT.1 and RT.2 only concern **Casses-Multi**)

CAN is the code for the anomaly.

LIB is the label for the anomaly,
SEV is the severity of the anomaly with one of two possible values:

- **B** if the anomaly is blocking (critical),
- **I** if the anomaly is not blocking (information).

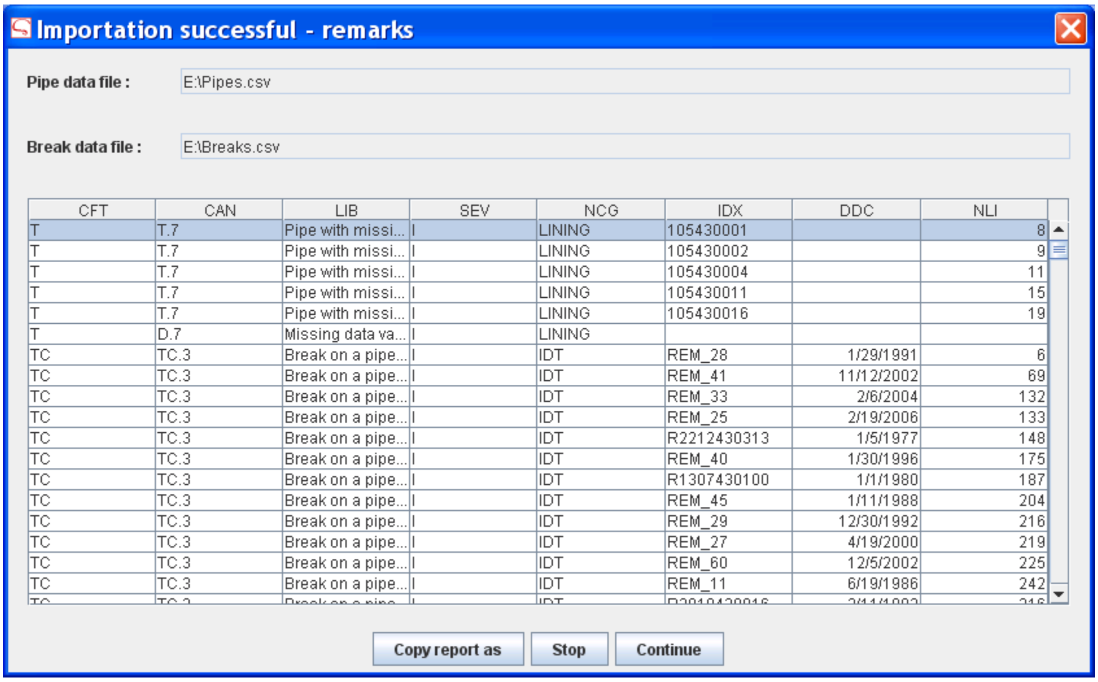
NCG is the short name of the data characteristic concerned (if applicable, otherwise empty).
IDX is the ID of the network if the anomaly concerns a network (Multi-network version only) otherwise it is the ID of the pipe (or empty).
DDC is the date of the break (or empty).
NLI is the row number of the data in the treated file (or empty).

List of anomalies treated

CAN Code anomaly	**LIB** Label of anomaly	**SEV** Severity of anomaly
F.1	Missing file	B
F.2	Unreadable file	B
F.3	Error in data specification area	B
F.4	Error in data area	B
F.5	File contains no records	B
F.6	Unable to match any break with any pipe	B
F.7	All breaks outside the record periods of the networks	B
D.1	Invalid or missing short data name	B if compulsory I if additional
D.2	Predefined mandatory or additional data with non-compliant type	B if compulsory I if additional
D.3	Additional data type not recognized	I
D.4	Missing mandatory data	B
D.5	Quantitative data value invalid	B if compulsory I if additional
D.6	Categorical data with more than 20 modalities	B if compulsory I if additional
D.7	Missing data value	B if compulsory I if additional
D.8	Date format not recognized	B if compulsory I if additional
D.9	Short data name already used in the dataset	B if compulsory I if additional

CAN Code anomaly	**LIB** Label of anomaly	**SEV** Severity of anomaly
T.1	Missing pipe identifier	B
T.2	Pipe identifier not unique	B
T.3	Missing or incorrect installation date	B
T.4	Installation date after removal date	B
T.5	Pipe length not positive	B
T.6	Pipe data invalid	B if compulsory I if additional
T.7	Pipe data missing	B if compulsory I if additional
C.1	Installation date after replacement date	I
C.2	Pipe length not strictly positive	I
C.3	Pipe with invalid data value	B
C.4	Pipe with missing data value	I
C.5	Break without pipe identifier	I
TC.1	Missing or incorrect break date	I
TC.2	Pipe with duplicated break date(s)	I
TC.3	Break with invalid data value	I
R.1	Break with missing data value	B
R.2	Break occurrence before pipe installation	B
R.3	Break occurrence strictly after pipe installation	B
R.4	Break on a pipe unidentified in the pipe file	B
R.5	Unidentified network	B
RT.1	Pipe with network identifier absent from network file	B
RT.2	Network contains no pipes	I
RT.3	Missing pipe observation window	I
RT.4	Break outside of observation period	I

Remark: For **Casses-Multi**, IDR is compulsory and anomalies D.2, D.4, and D.7 apply.



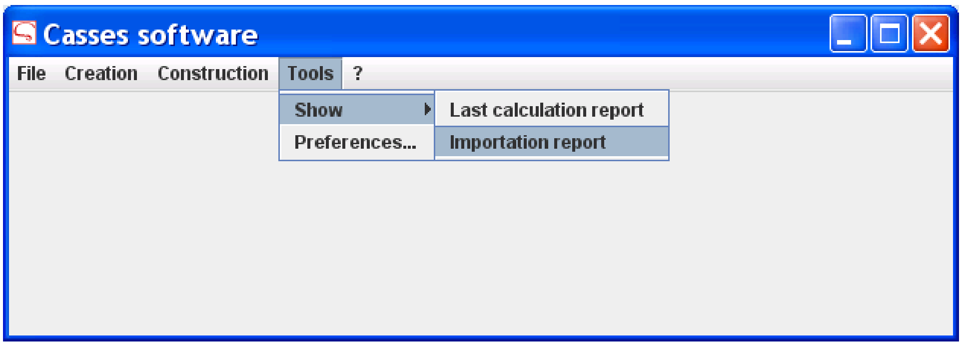
CFT	CAN	LIB	SEV	NCG	IDX	DDC	NLI
T	T.7	Pipe with missi...	I	LINING	105430001		8
T	T.7	Pipe with missi...	I	LINING	105430002		9
T	T.7	Pipe with missi...	I	LINING	105430004		11
T	T.7	Pipe with missi...	I	LINING	105430011		15
T	T.7	Pipe with missi...	I	LINING	105430016		19
T	D.7	Missing data va...	I	LINING			
TC	TC.3	Break on a pipe...	I	IDT	REM_28	1/29/1991	6
TC	TC.3	Break on a pipe...	I	IDT	REM_41	11/12/2002	69
TC	TC.3	Break on a pipe...	I	IDT	REM_33	2/6/2004	132
TC	TC.3	Break on a pipe...	I	IDT	REM_25	2/19/2006	133
TC	TC.3	Break on a pipe...	I	IDT	R2212430313	1/5/1977	148
TC	TC.3	Break on a pipe...	I	IDT	REM_40	1/30/1996	175
TC	TC.3	Break on a pipe...	I	IDT	R1307430100	1/1/1980	187
TC	TC.3	Break on a pipe...	I	IDT	REM_45	1/11/1988	204
TC	TC.3	Break on a pipe...	I	IDT	REM_29	12/30/1992	216
TC	TC.3	Break on a pipe...	I	IDT	REM_27	4/19/2000	219
TC	TC.3	Break on a pipe...	I	IDT	REM_60	12/5/2002	225
TC	TC.3	Break on a pipe...	I	IDT	REM_11	6/19/1986	242

By means of the button “Copy report as” you can save the anomaly report in csv format, the fields being separated by semi-colons (;).

The first row contains the titles:
CFT;CAN;LIB;SEV;NCG;IDX;DDC;NLI

Then one row per anomaly detected, the information being delivered in the same order as the first row and also separated by semi-colons.

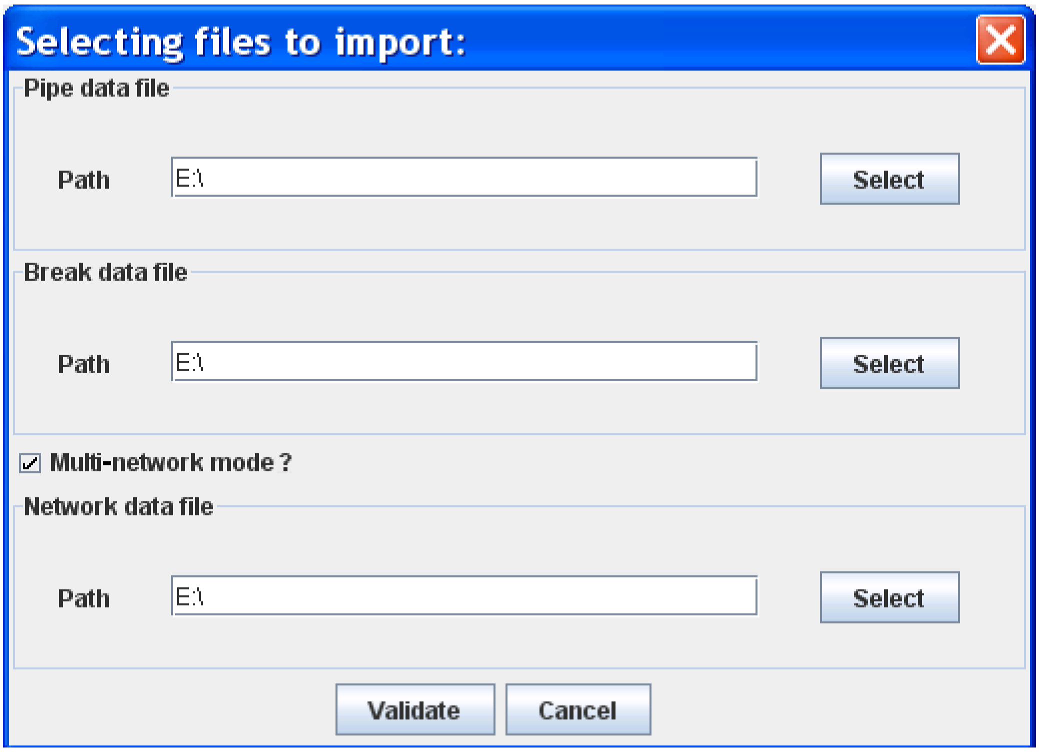
If at least one anomaly is blocking (code SEV= B), the “Continue” button is disabled. Otherwise clicking on this button allows the main screen to be accessed and for the Project to be started. The last importation report produced by **Casses** is accessible by:



It is also available as a text file in the directory specified in the “Preferences” under the name “Rimp.txt”.

IV.1.ii Casses-Multi

The following dialog box appears:

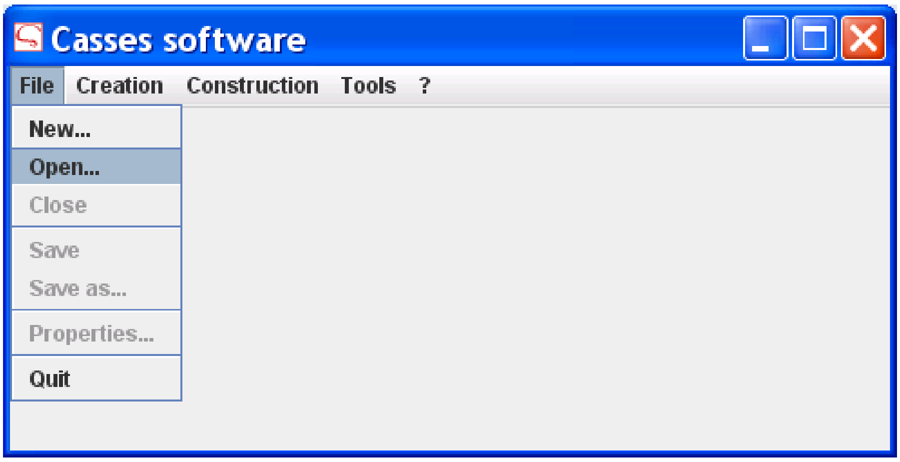


As well as the breaks and pipes files, it is necessary to specify the name and location of the networks file.

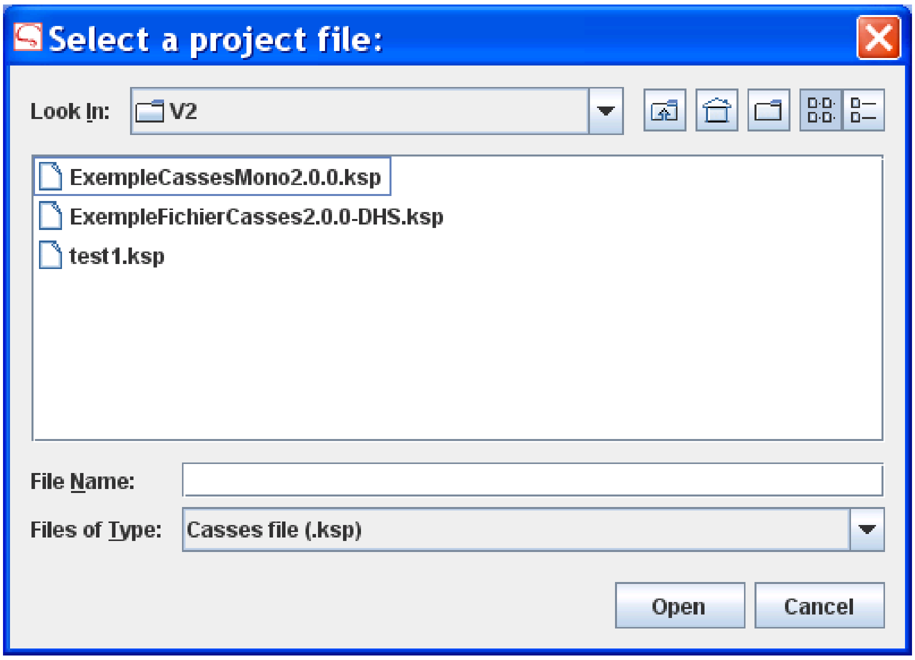
Unchecking the box “Multi-network mode?” reverts to mono-network mode.

Multi-network mode proceeds as with mono-network mode other than the dialog concerning the break recording period, which does not appear.

IV.2 Open a Project



A dialog box asks you to identify the Project file that you wish to open and its location. The files have the extension .ksp.



Casses-Multi allows files that were created with **Casses-Mono** to be opened. The inverse is not possible.

V Exploring the data

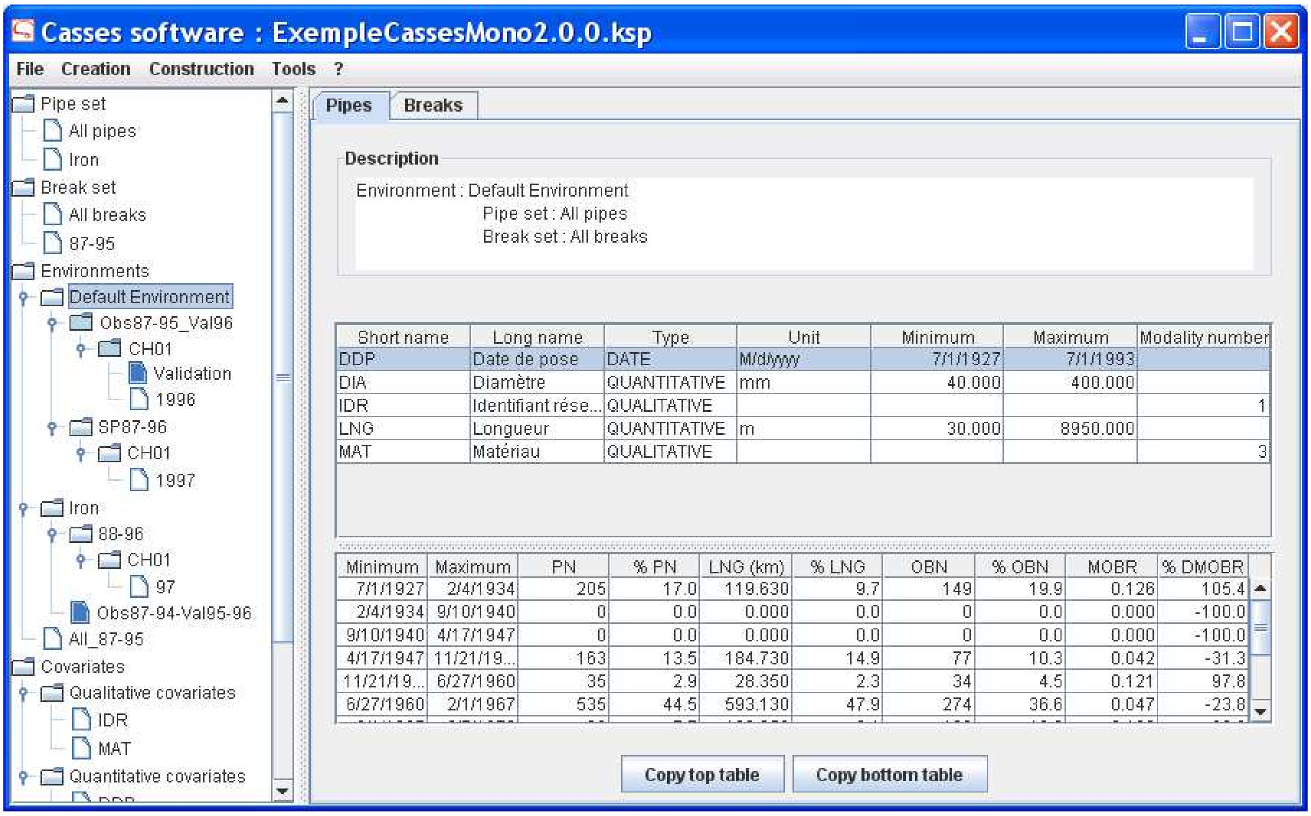
Casses comprises two modes, “Exploration” that allows visualisation of the Project data and calculations and “Construction” that allows the calculations to be performed and predictions to be made.

V.1 *Organisation of the explorer window*

The “Exploration” window is divided as follows:

- On the left side, a “navigation tree” that acts as a file manager,
- On the right side, one or more pages accessible by tabs at the top; the contents of these pages is adapted to the object selected on the left side.

In many cases, the right window is itself divided into top and bottom sections. The information in the bottom section depends on the selection made in the top section.



The families of objects present in the navigation tree on the left are:

- The pipe sets,
- The break sets,
- The Environments,
- The covariates,
- The break data,
- The networks (only in the multi-network version).

V.2 Some useful functionality

A certain number of ergonomic rules are applicable for the exploration windows:

- The size of the different sections is adjustable by moving the separating borders;
- The columns in the tables are adjustable in size and moveable by actions in the header row (shaded grey);
- The tables can be sorted according to each column by simply clicking the column header according to the sequence “sort ascending”, “sort descending”, “no sort”;
- All or part of each table can be copied to the clipboard by selecting the cells and pressing <Ctrl> + C. The copy includes the header row and the exact data values;
- At the bottom of the right section, the buttons allow the entire table to be copied and if necessary exported in .csv format;
- Right clicking on an element of the navigation tree accesses the menu of possible operations for that object.

VI Creating pipe sets or break sets

VI.1 What is a set?

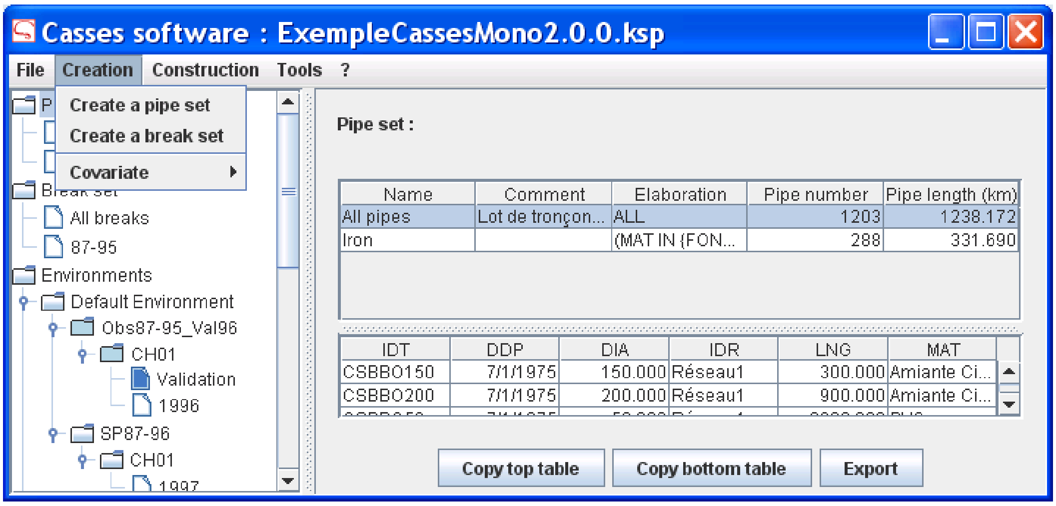
A pipe set is a collection of pipes selected from those present in a Project.

In the same way, a break set is a collection of breaks selected from those present in a Project.

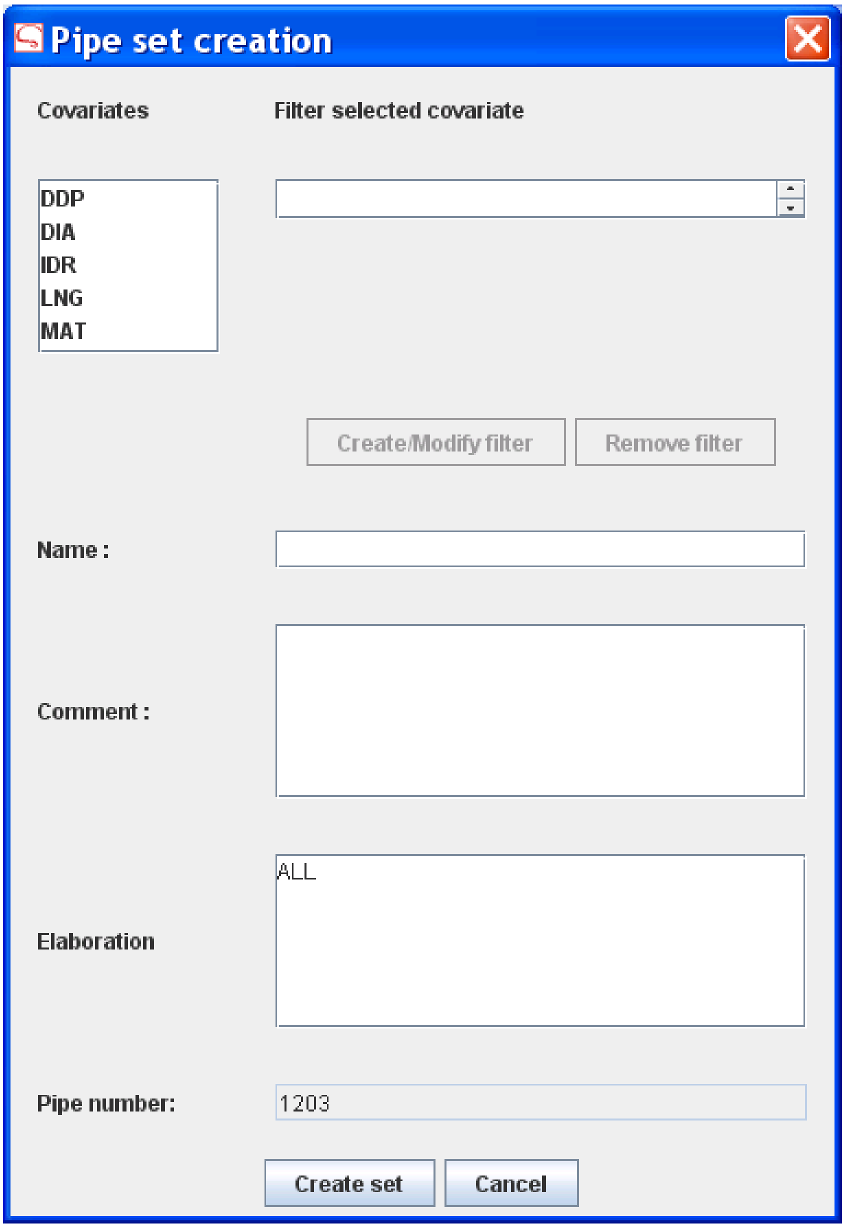
After importing the data **Casses** automatically creates a pipe set containing all the pipes and a break set containing all the valid breaks.

You have the possibility of creating other sets. This is particularly useful if you only wish to study one category of pipes, or if you wish to be selective in terms of which breaks to consider.

The functionality is accessible by the menu “Creation” or by right clicking in the navigation tree on one of the objects “Pipe set” or “Break set”.



The creation of a set is made in a new window.



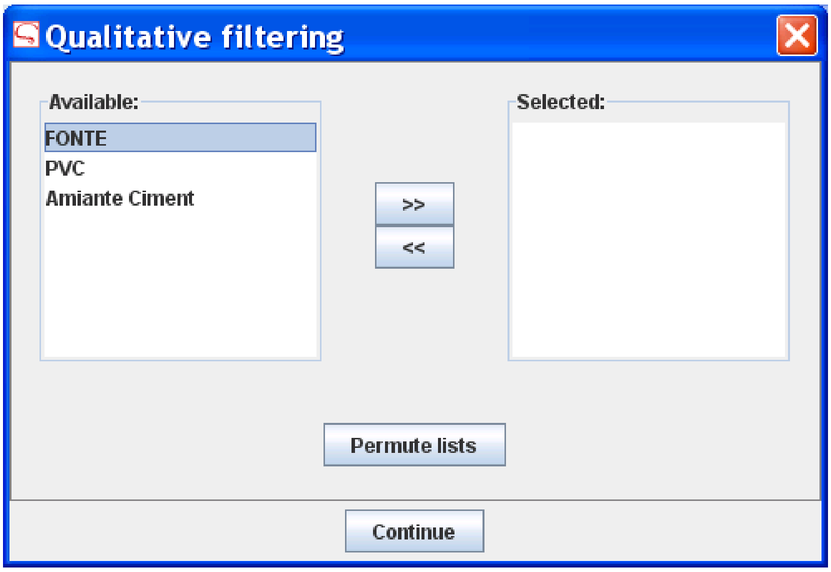
The set must be named and a space for optional comments is provided to allow a more precise description of the set.

VI.2 The functioning of filters

Sets are created by filtering. You must therefore select a covariate from which the filter is performed and then click on the button “Create/modify filter”.

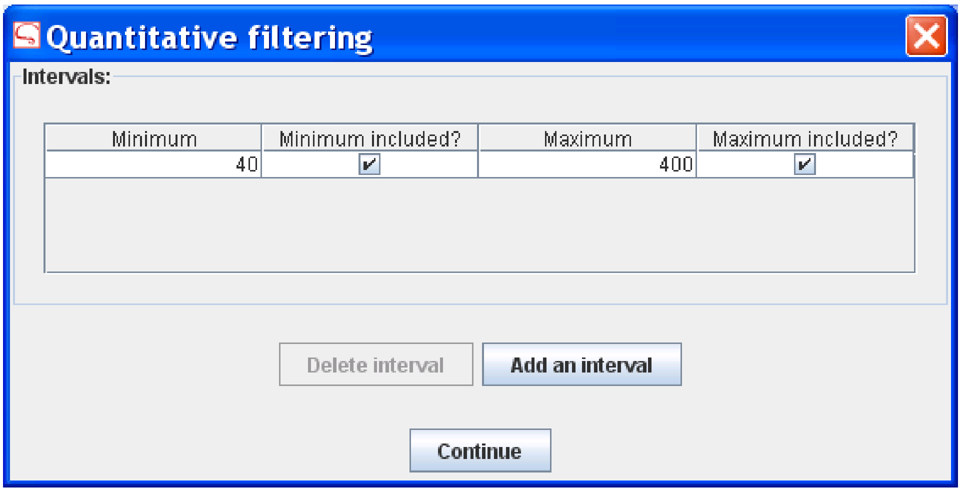
VI.2.i Qualitative filtering

If the data are qualitative, the filtering consists of selecting the modalities to be kept with the aid of the arrows or by double clicking. Only the pipes (or breaks) possessing one of the chosen values will be included in the set.



VI.2.ii Quantitative filtering

If the data are quantitative, the filtering consists of defining intervals. Only the pipes (or breaks) whose value lies between the intervals will be included in the set.



VI.2.iii Combination of many filters

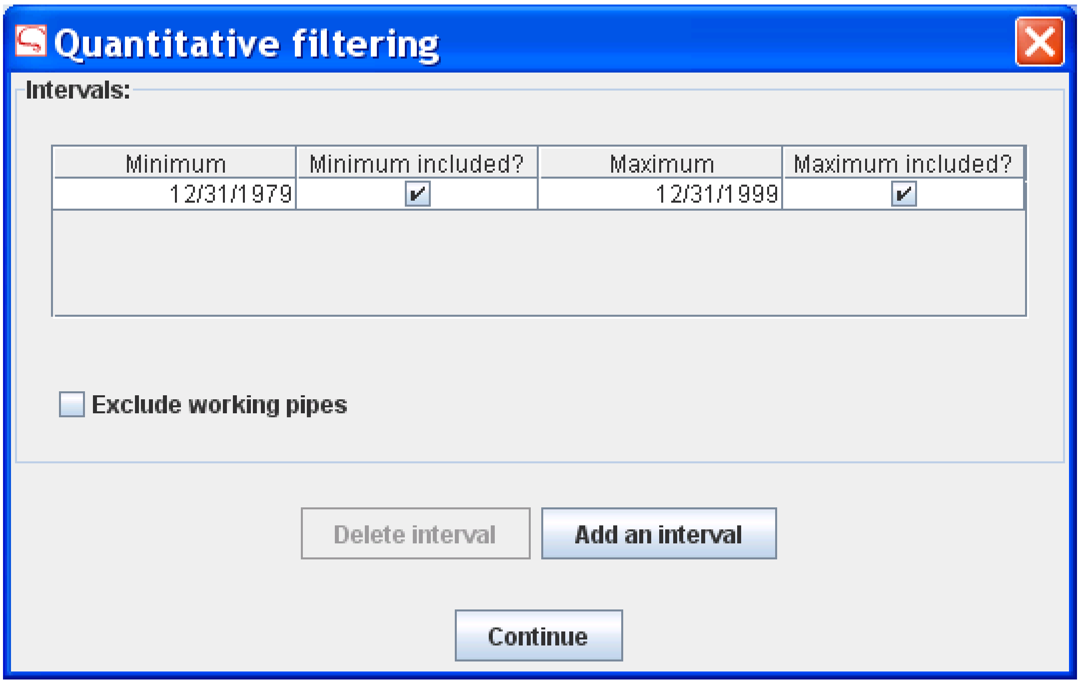
You can define many filters; they are taken into account in a cumulative fashion (logical “AND”).

The set is characterised by a logical "signature" that is indicated in the "Elaboration" field. For example, for iron pipes greater or equal to 100m and less than 1000m, the signature of the lot is: (LNG IN [100;1000[)^(MAT IN {IRON}).

The number of records in the set is indicated at the bottom of the creation window.

VI.2.iv Particular case of filtering by DHS

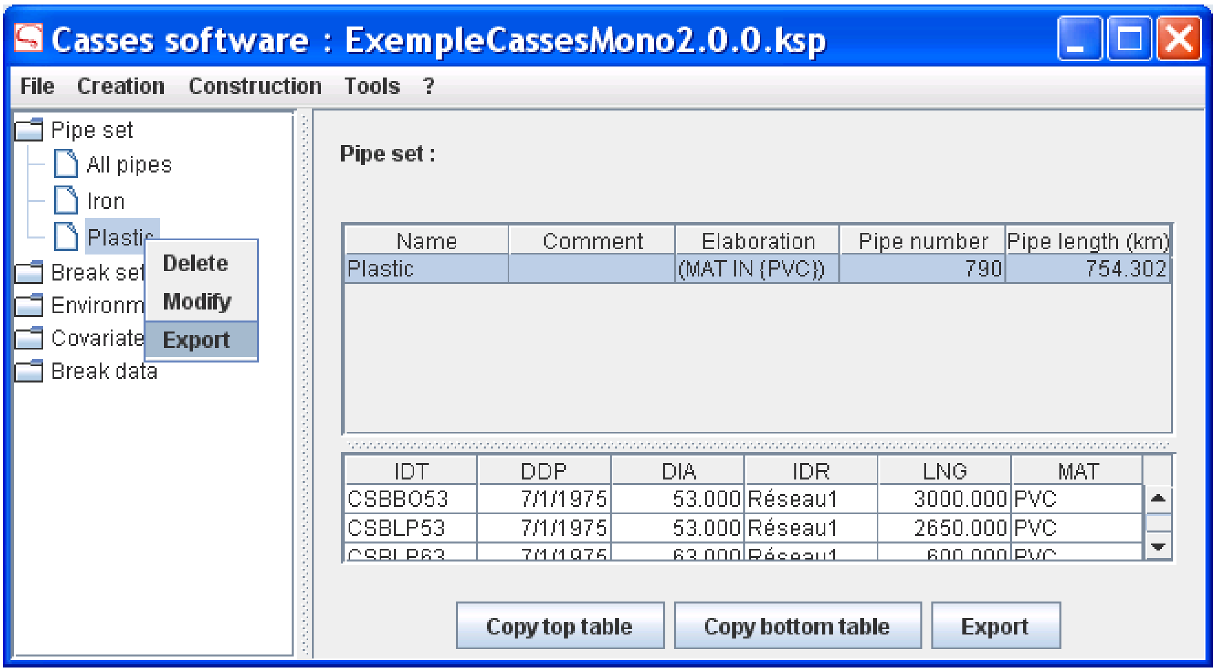
The data DHS, date removed from service, has the particularity of being the only data for which blank values are allowed in the importation (pipes in service). To take into account this particularity, the quantitative filter includes a checkbox "Exclude working pipes".



In practice, **Casses** attributes a distant removal date (01/01/2999) to pipes in service thus checking the box excludes the interval]1/1/2998 ;1/1/2999].

VI.3 Exporting sets

Sets can be exported as .csv files by clicking on the set and on the button "Export" at the bottom of the screen or by right clicking on the set.



The sets exported in this way are in the **Casses** format thus they can be imported into a new Project without modification.

VII Create, modify or delete covariates

Covariates are the data attributed to pipes and can be distinguished as quantitative or qualitative.

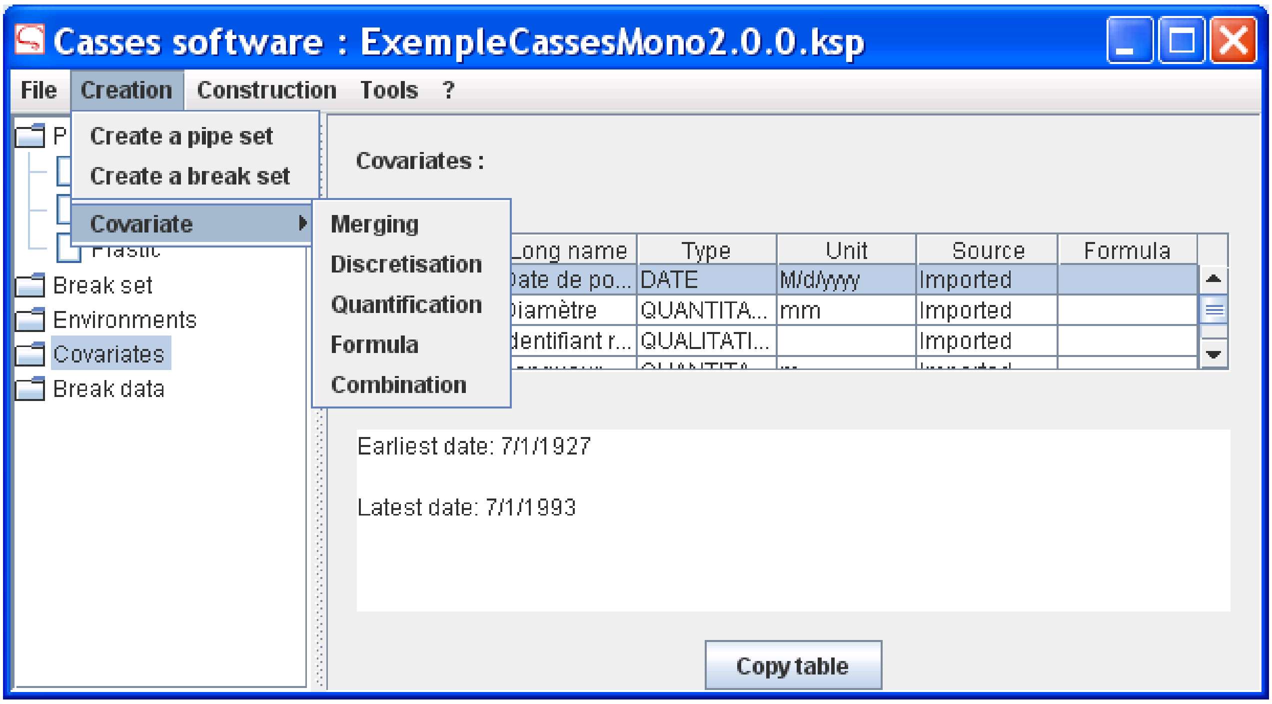
Quantitative covariates represent a quantity (numeric) or a date.

Qualitative covariates have a limited number of values (maximum 20 modalities).

After the importation of the data, all the covariates present and useable in the pipes file are available to be used in a Project. In the Exploration window, they are regrouped as an object called "Covariates" with a "Quantitative Covariates" branch and a "Qualitative Covariates" branch.

It is possible to create new covariates from existing ones.

This functionality is accessible from the "Creation" menu or by right clicking on the "Covariates" object.

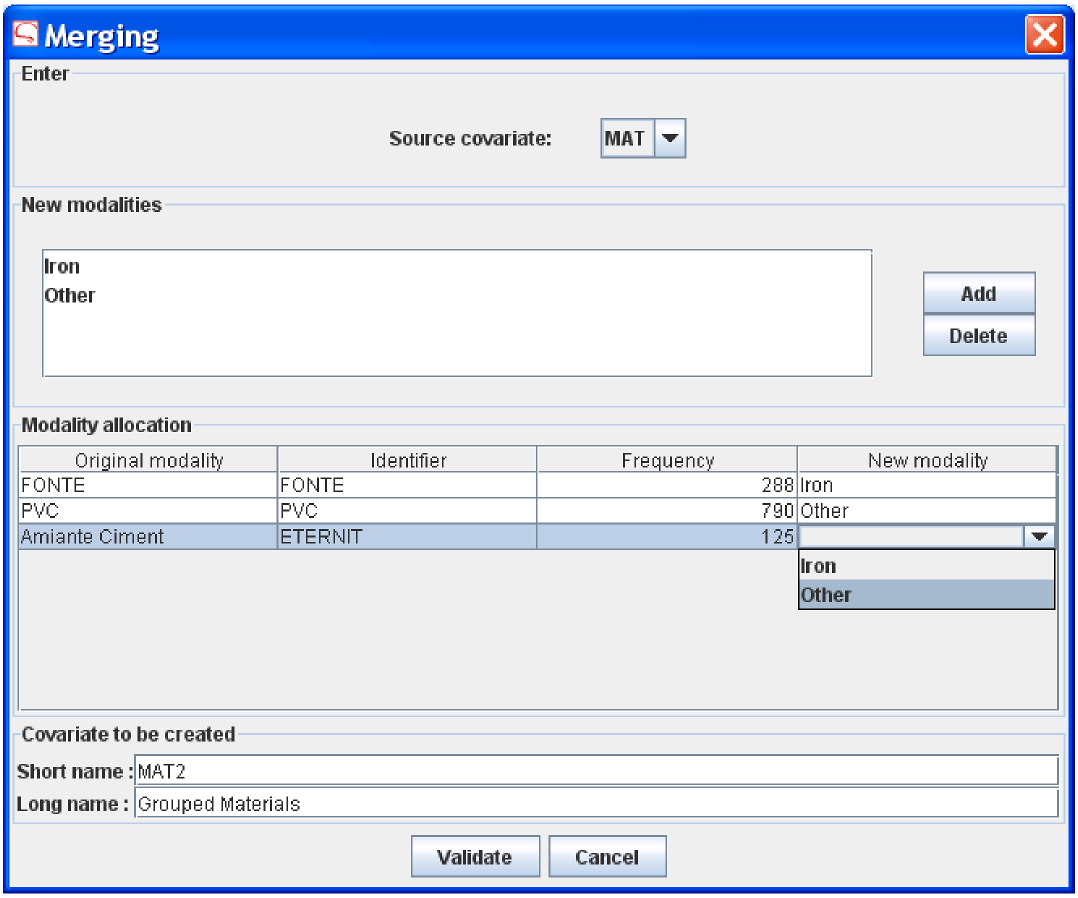


Five methods are available allowing covariates to be created:

- Merging,
- Discretisation,
- Quantification,
- Formula,
- Combination.

VII.1 Covariates creation: “Merging”

Merging allows a new qualitative covariate to be created from an existing one by regrouping existing modalities into fewer groups.



Firstly, use the scrolling list to select the source for the merging operation from the existing qualitative covariates.

You then must create the new modalities of the new covariate with the aid of the “Add” button.

The next stage involves associating each of the source covariate modalities with the new covariate modalities using the dropdown list.

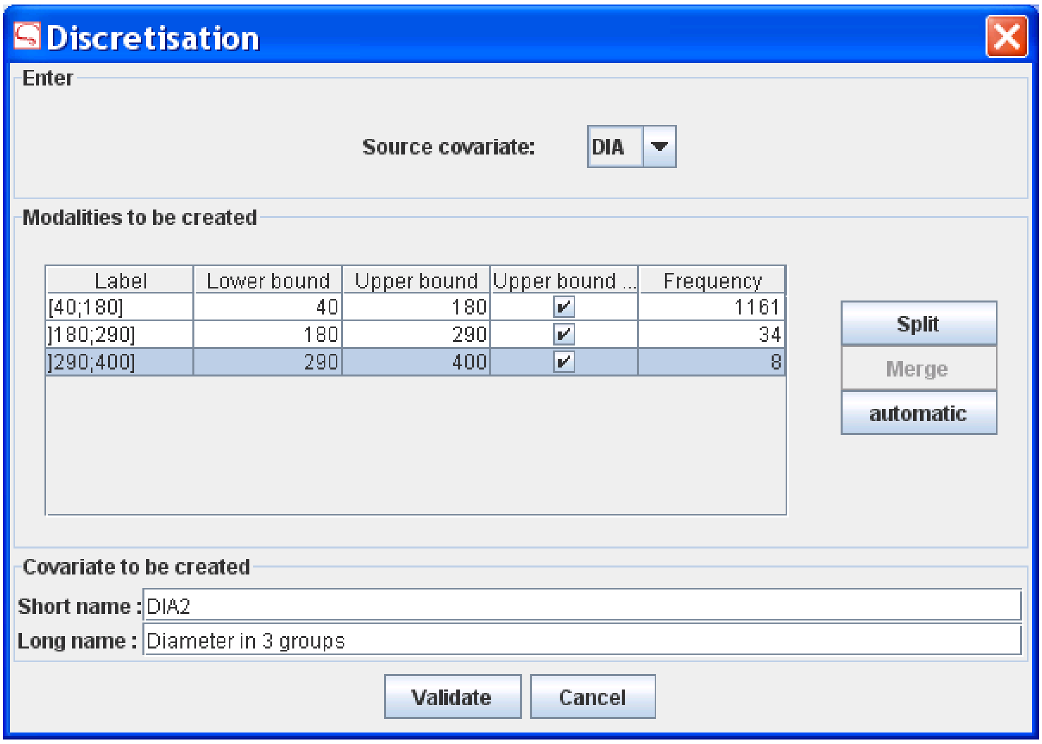
The new covariate must be identified with a short name. This must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).

An optional long name may be used to further identify the covariate.

The description of how the new covariate is created is memorised; this information is accessible on the right part of the screen when the covariate is selected in the explorer window.

VII.2 Covariates creation: "Discretisation"

Discretisation allows a new qualitative covariate to be created from an existing quantitative one. The modalities of the new covariate are intervals defined by the values of the source covariate.



Firstly, use the scrolling list to select the source for the discretisation operation from the existing quantitative covariates.

You can then create the modalities of the new covariate:

- "Split": After you have selected one row in the table, this button splits the selected interval into two halves.
- "Merge": After you have selected several adjacent rows in the table, this button merges the intervals into one, covering the whole range of the rows selected.
- "Automatic": This button opens a dialog box in which you must specify the number of intervals (between 2 and 20) to create. The intervals created have the same size range and include the upper bound.

For each interval created, you can modify its upper bound so long as it remains consistent with the adjacent intervals. The inclusion or not of the upper bound of the interval must be chosen. The lower bounds are deduced from the upper bounds.

A default label describes the interval and serves as the name for the modality of the created covariate. This label is modifiable.

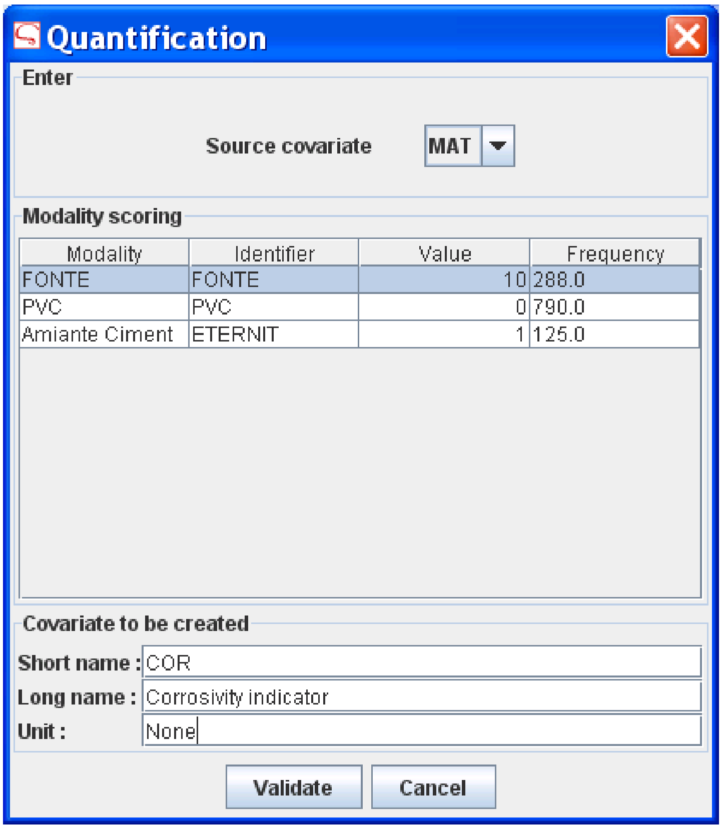
The new covariate must be identified with a short name. This must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).

An optional long name may be used to further identify the covariate.

The description of how the new covariate is created is memorised; this information is accessible on the right part of the screen when the covariate is selected in the Exploration window.

VII.3 Covariates creation: “Quantification”

Quantification allows a new quantitative covariate to be created from an existing qualitative one. A numeric value is attributed to each of the source covariate modalities.



Firstly, use the scrolling list to select the source for the quantification operation from the existing qualitative covariates.

The next stage consists of attributing a numeric value to each of the source covariate modalities.

The new covariate must be identified with a short name. This must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).

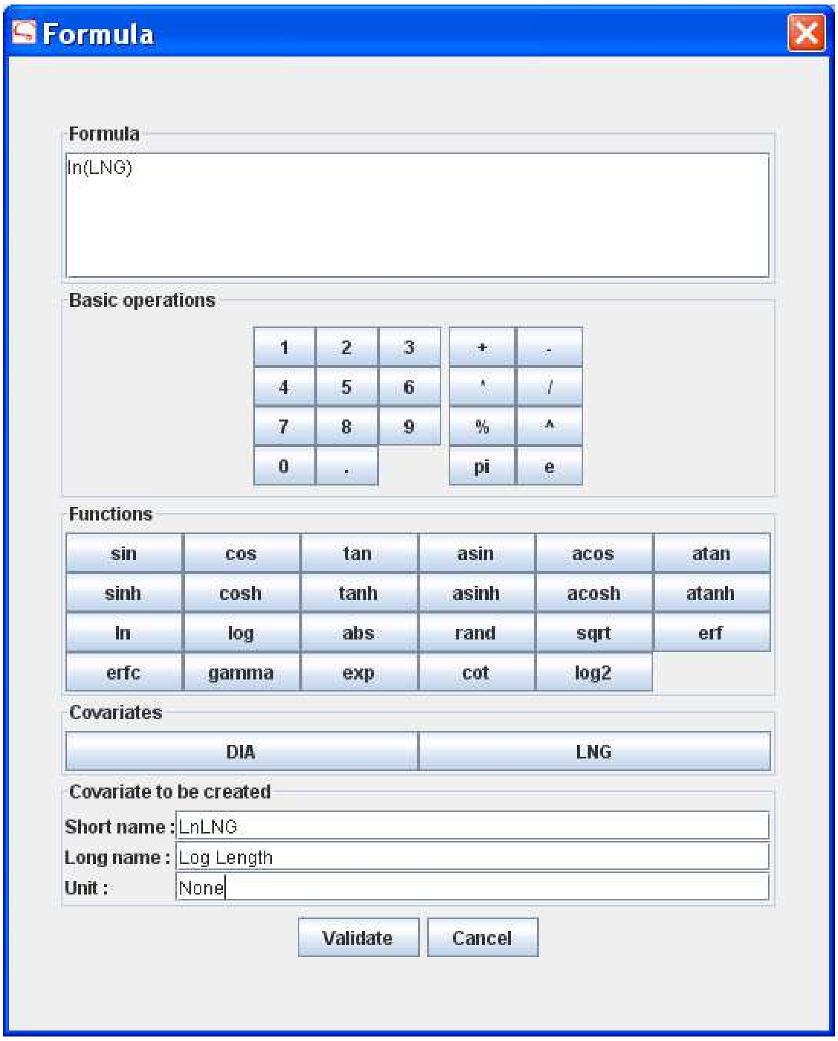
An optional long name may be used to further identify the covariate.

The units applicable to the new values can also be specified (optional).

The description of how the new covariate is created is memorised; this information is accessible on the right part of the screen when the covariate is selected in the explorer window.

VII.4 Covariates creation: "Formula"

The formula tool allows a new quantitative covariate to be created by applying a mathematical formula involving one or more existing quantitative covariates.



The formula for calculating the new covariate is obtained by clicking on the various mathematical functions and operators and the eligible covariates. The formula can also be entered directly in the area intended for this purpose.

The new covariate must be identified with a short name. This must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).

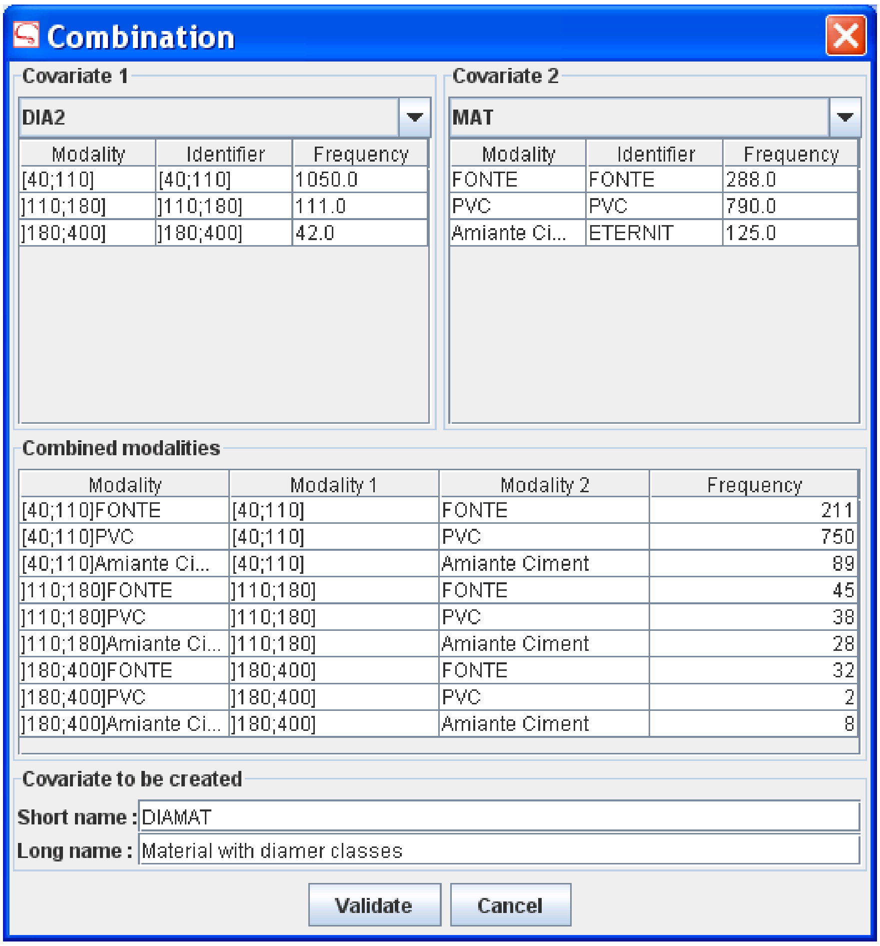
An optional long name may be used to further identify the covariate.

The units applicable to the new values can also be specified (optional).

The description of how the new covariate is created is memorised; this information is accessible on the right part of the screen when the covariate is selected in the explorer window.

VII.5 Covariates creation: “Combination”

Combination allows a new qualitative covariate to be created from two existing qualitative covariates. The modalities of the new covariate result from the combination of those of the source covariates. Only modalities applying to at least one pipe will be created.



Using the scrolling menus, select two existing qualitative covariates to serve as the sources of the new covariate.

If the combined number of modalities is greater than 20, the covariate cannot be created.

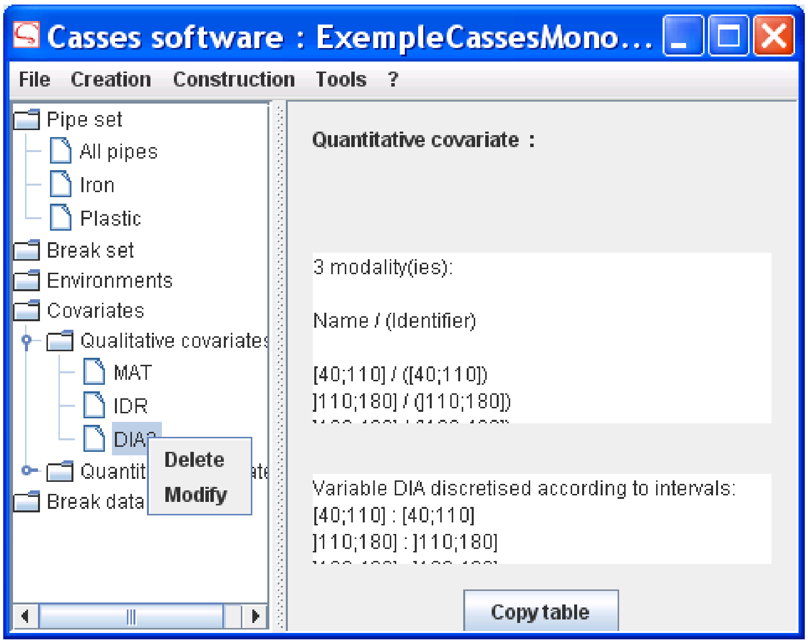
The new covariate must be identified with a short name. This must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).

An optional long name may be used to further identify the covariate.

The description of how the new covariate is created is memorised; this information is accessible on the right part of the screen when the covariate is selected in the explorer window.

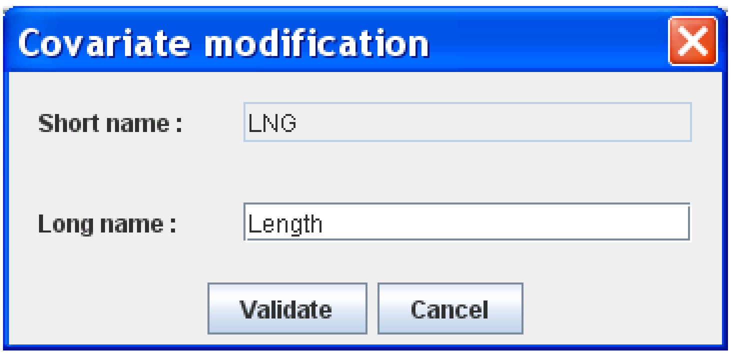
VII.6 Modify or Delete Covariates

It is possible to modify or delete covariates. These functionalities are accessible by right clicking the appropriate covariate in the exploration window.

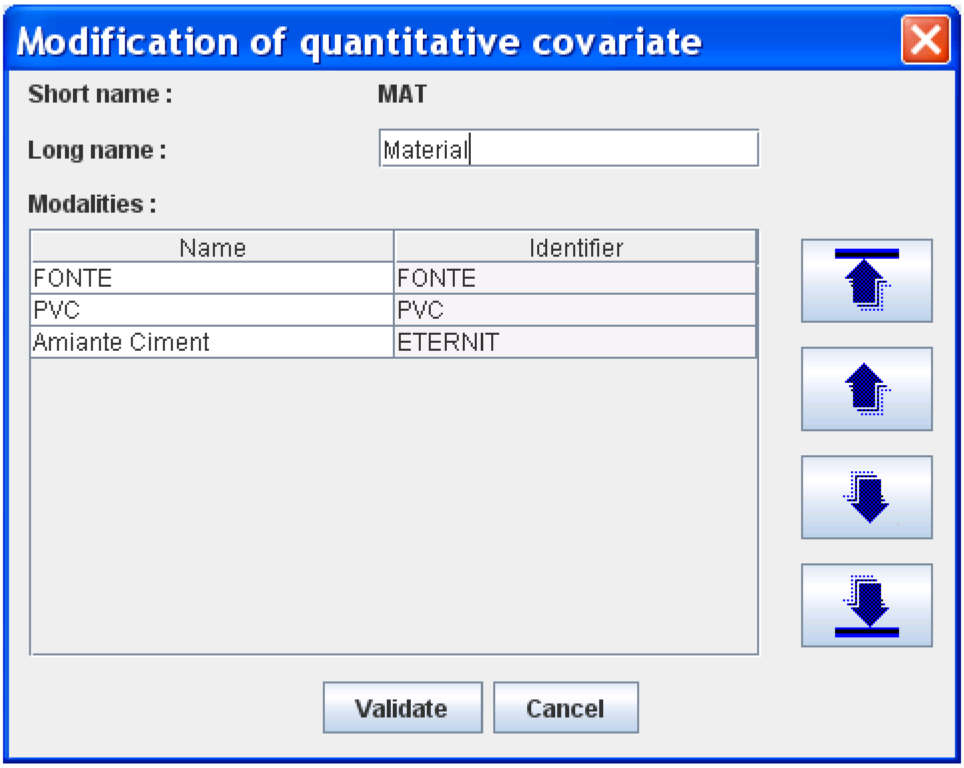


VII.6.i Modify a covariate

For quantitative covariates only the long name can be modified.



For qualitative covariates it is also possible to change the modality labels and the order in which they appear.



Two distinct modalities of the same covariate cannot share the same name.

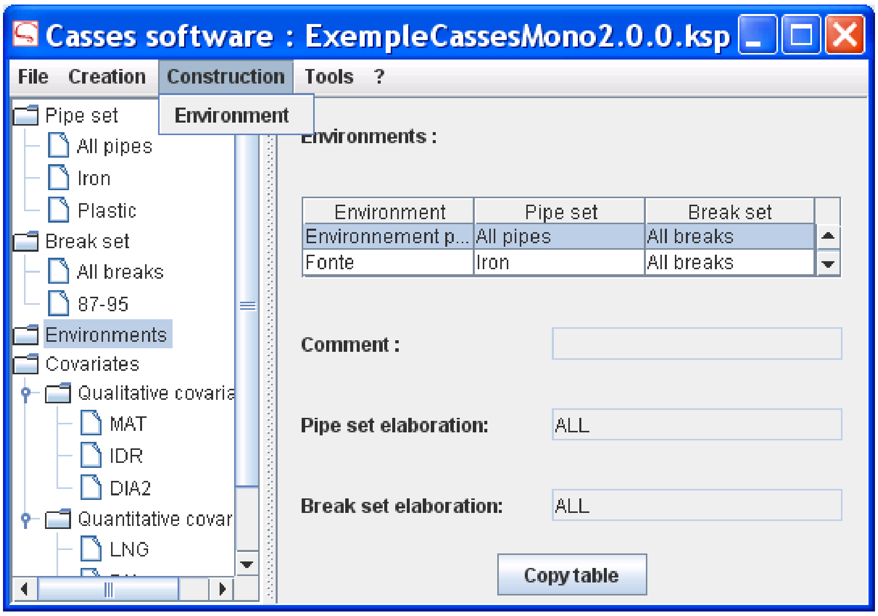
VII.6.ii Delete a covariate

Covariates that are not compulsory and that are not used in any Progression can be deleted. This action is irreversible and does not affect the pipe sets nor covariates previously created from the deleted covariate.

VIII Constructing a prediction

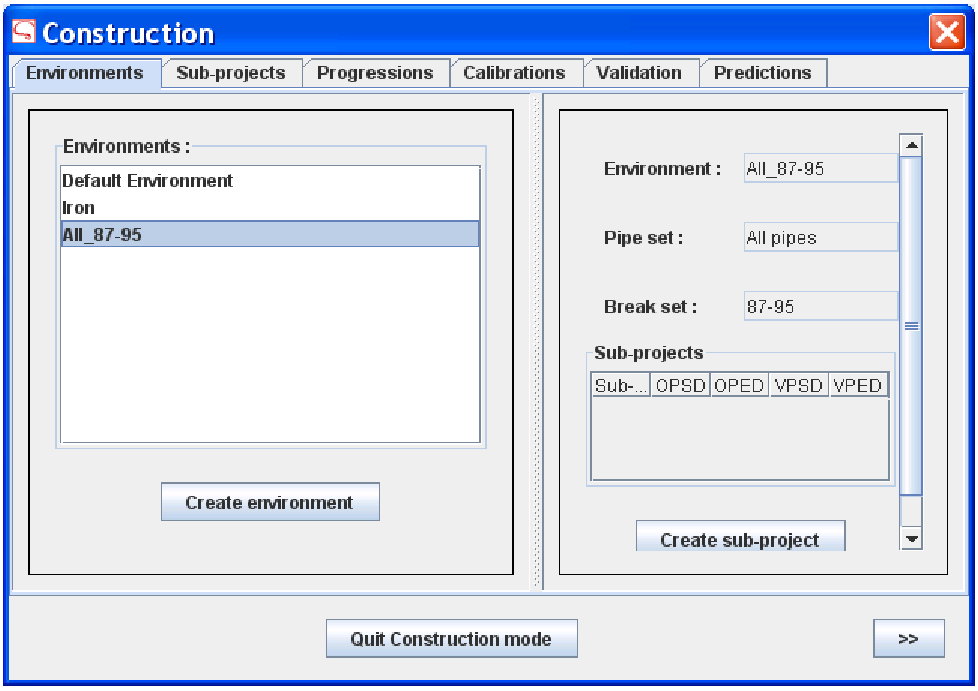
VIII.1 The assistant for constructing a prediction

From the Exploration window, the prediction constructing assistant is accessible from the “Construction” menu or by right-clicking on the “Environment” object or any Environments within it.



The Construction mode comprises six screens: “Environments”, “Sub-projects”, “Progressions”, “Calibrations”, “Validation” and “Predictions”.

These screens are accessible sequentially using the “>>” and “<<” buttons or by clicking on the tabs.



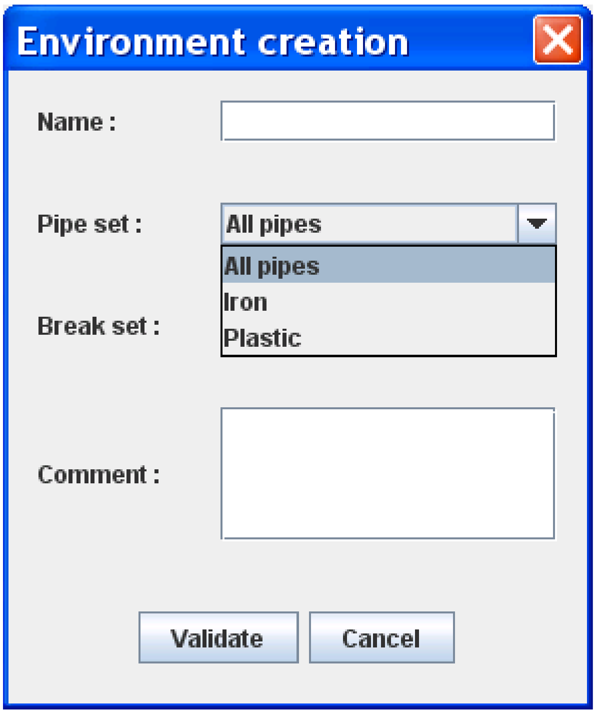
To return to the Exploration window, click on the button “Quit Construction mode”.

Each screen is separated in two parts by a vertical movable bar. The display on the right section depends on the object selected in the left section.

VIII.2 Create an “Environment”

An “Environment” is an association of a pipe set and a break set. The default Environment is a pipe set with all the pipes and a break set with all the breaks and is created automatically.

The creation of a new Environment is made from the “Environments” screen in Construction mode by clicking on the button “Create an Environment”.



Firstly, the Environment must be named. The name can be chosen freely, the only constraint being that two Environments in the same Project cannot share the same name.

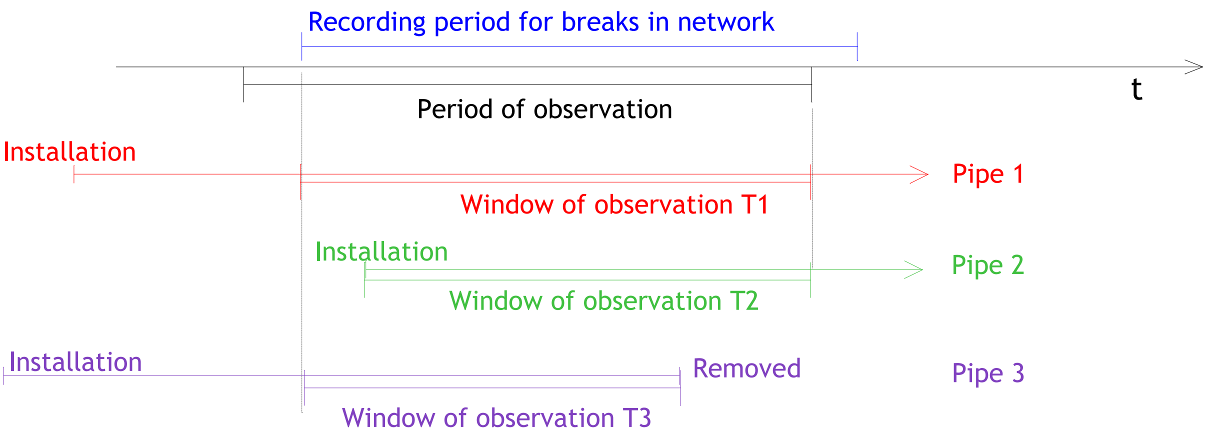
Next, use the scrolling lists to select a pipe set and a break set from those available in the Project.

A text box for optional additional comments is available for specifying the nature of the Environment.

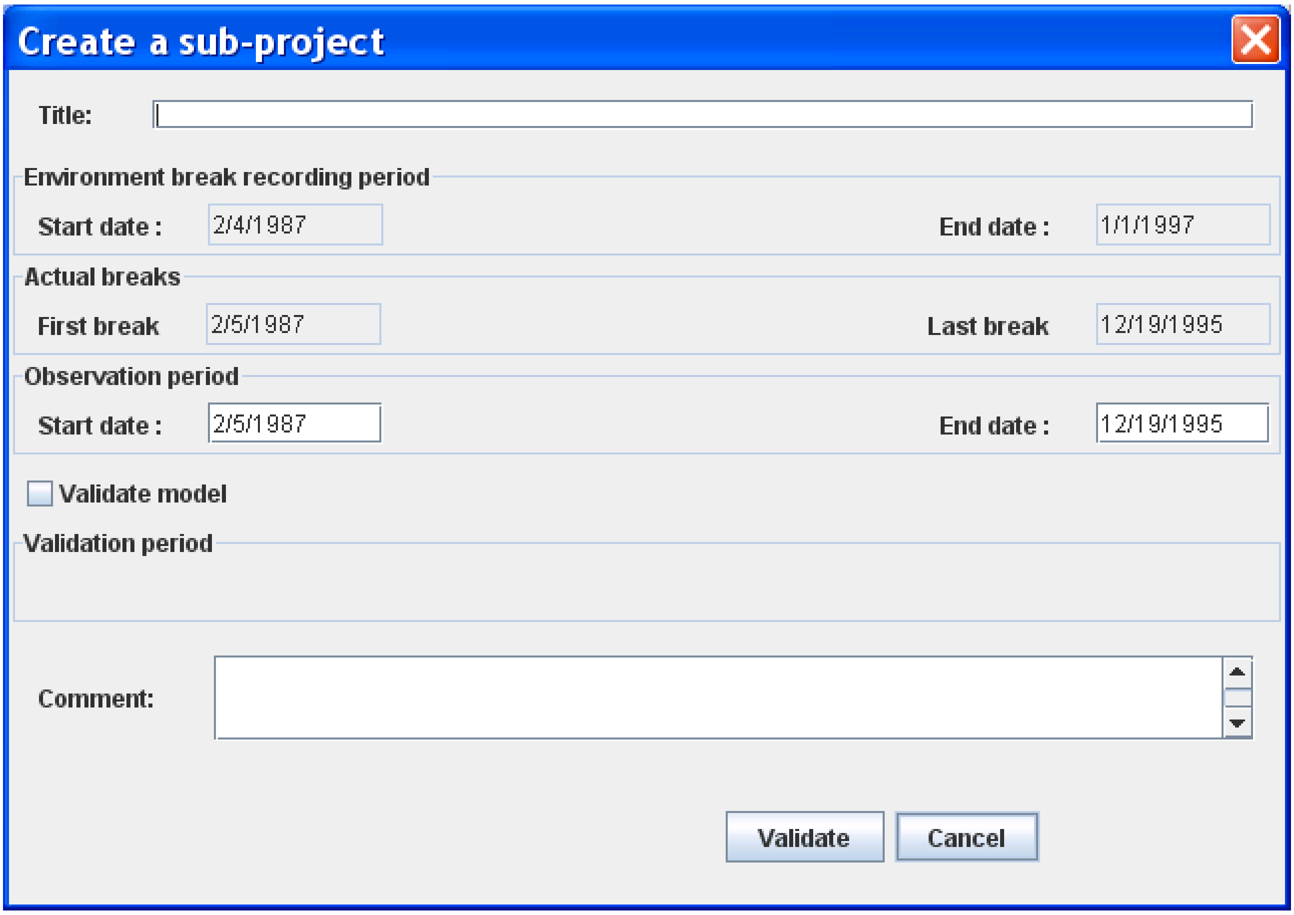
VIII.3 Create a “Sub-project”

A “Sub-project” is the association of an Environment and a period of observation. Only the breaks occurring during this period are taken into account.

By combination of the recording period for breaks in the network, the period of observation of the Sub-project, the pipe installation dates and removal dates (if applicable) the software determines the window of observation for each pipe.



Having selected an Environment, a Sub-project is created in the Construction mode either on the right side of the Environments screen or the left side of the Sub-projects screen by clicking on the button "Create a Sub-project".



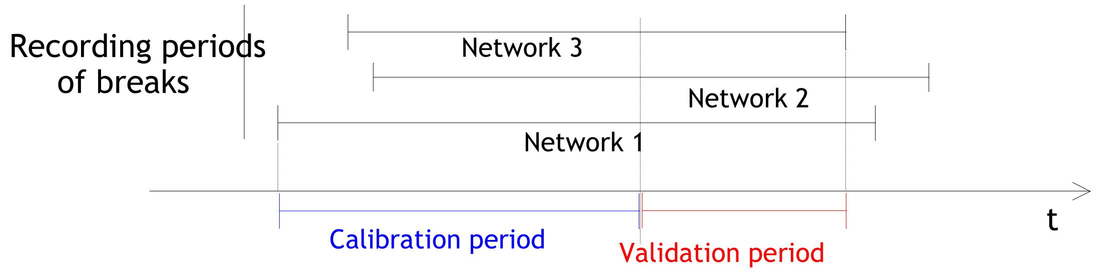
Firstly, the Sub-project must be named. The name can be chosen freely, the only constraint being that two Sub-projects in the same Environment cannot share the same name.

Next, choose the dates for the beginning and end of the observation period. By default, the dates of the first and last breaks in the Environment are proposed.

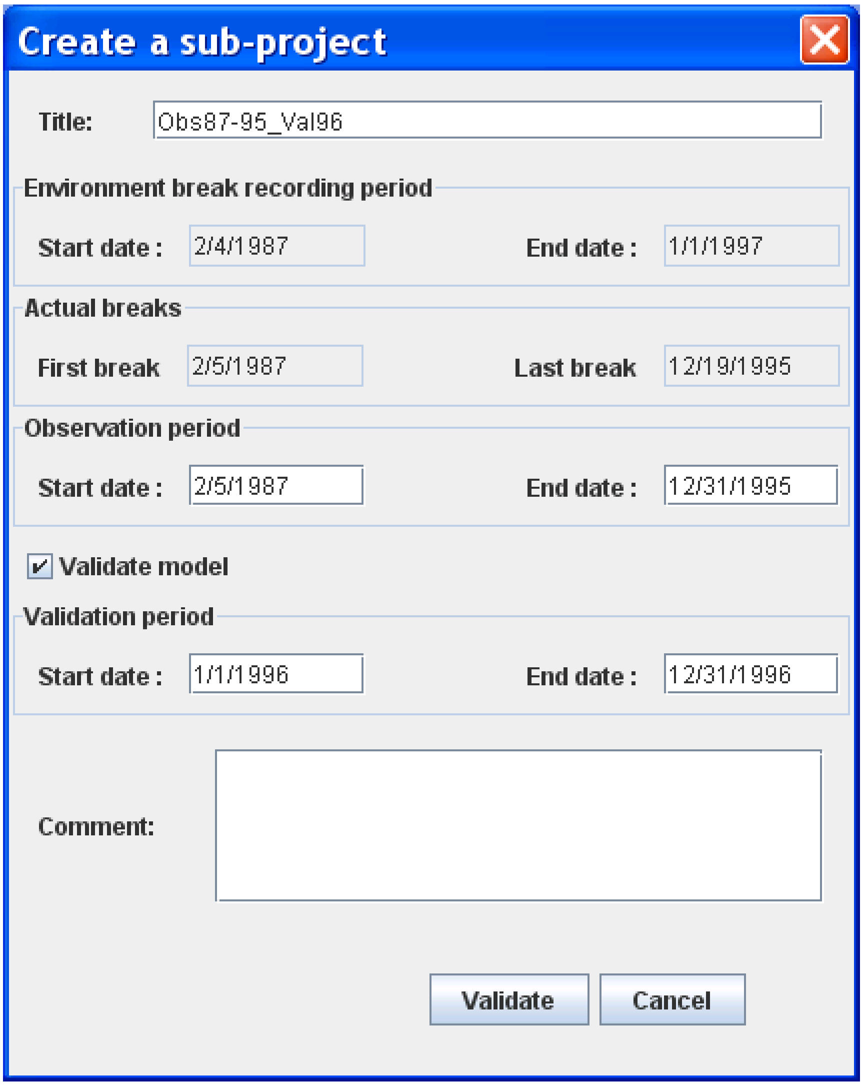
A text box for optional additional comments is available for specifying the nature of the Sub-project.

At this stage of the Sub-project the user can choose to carry out validation calculations by clicking the box “Validate Model”.

The general principle of the Validation is to compare the predictions of the model with the actual breaks observed (cf. XIII.5).



To perform a Validation it is therefore necessary to choose a date for the end of the period of observations that is before that of the recording period for the Environment.

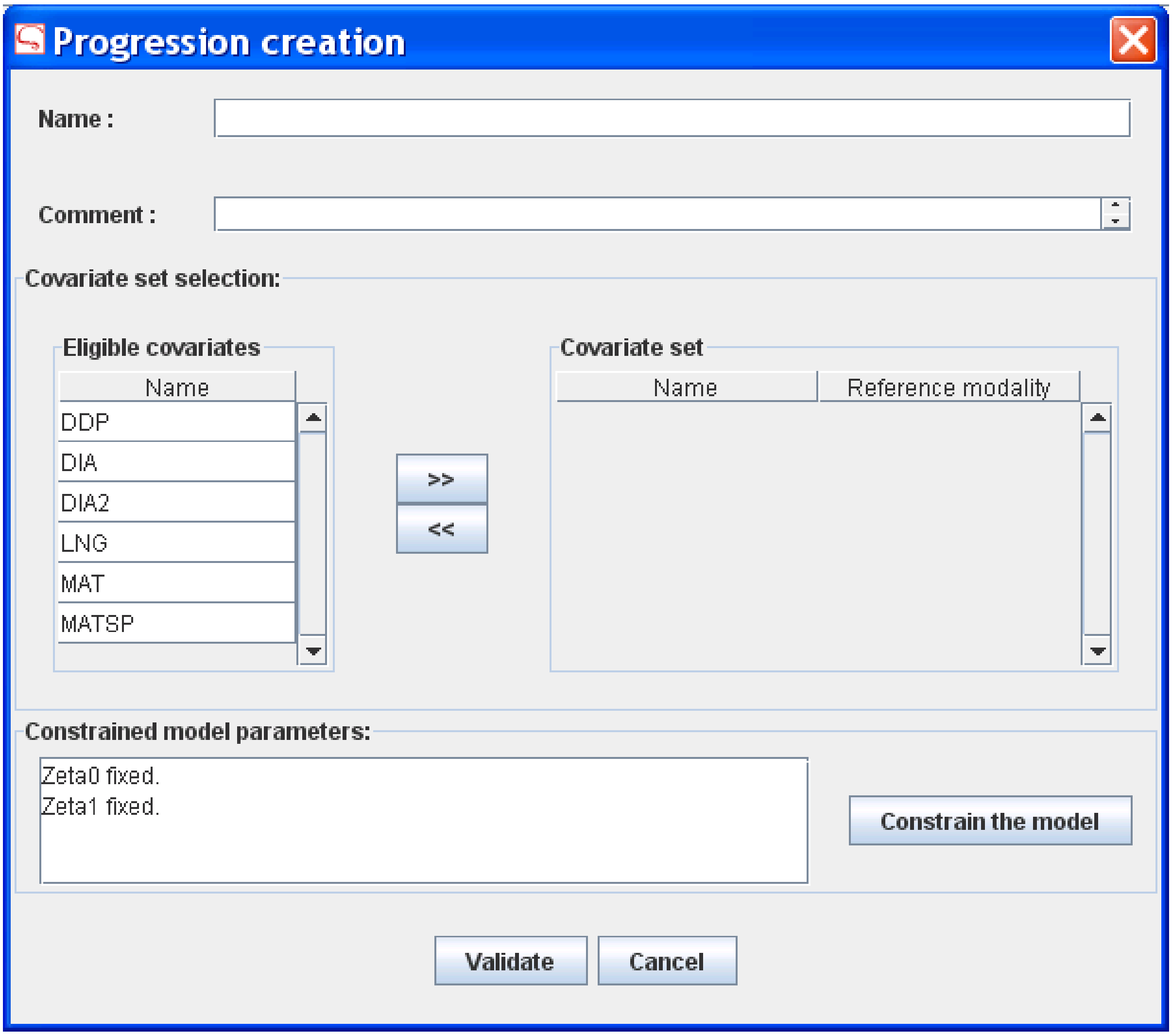


Select the dates for the start and end of the validation period. By default, the Validation start date is the day after the end of the period of observations and the end date, the end of the recording period for the Environment.

VIII.4 Create a "Progression"

A "Progression" is a selection of covariates and a list of constraints applied at the Sub-project level. The Progression contains all the information necessary to obtain a model by successive calibrations.

Having selected a Sub-project, a Progression is created in the Construction mode either on the right side of the Sub-projects screen or the left side of the Progressions screen by clicking on the button "Create a Progression".



Firstly, the Progression must be named. The name can be chosen freely, the only constraint being that two Progressions in the same Sub-project cannot share the same name.

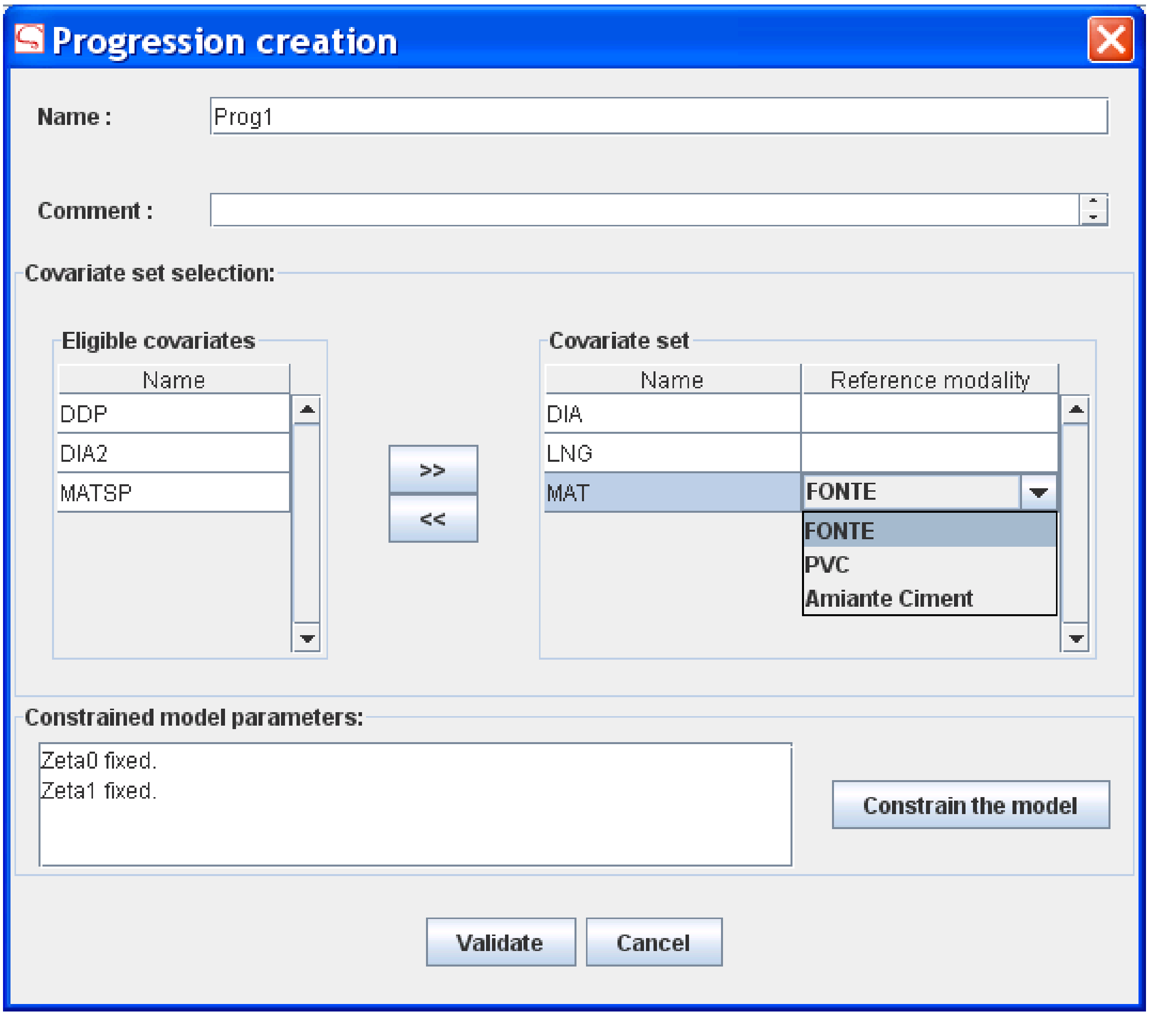
A text box for optional additional comments is available for specifying the nature of the Progression.

The next stage consists of selecting covariates to be included using the arrows or by double clicking. Only covariates that don’t have a unique value for every pipe in the Sub-project are available.

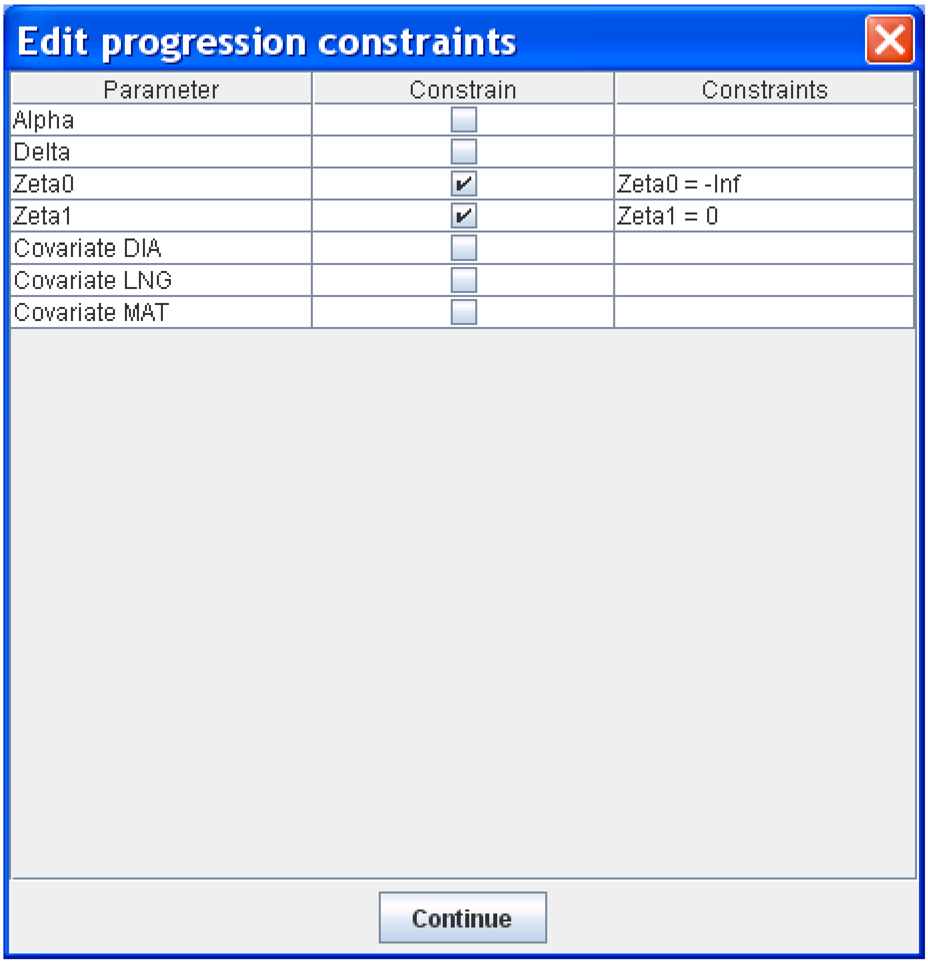
To be treated by the calculation kernel a qualitative covariate with “n” modalities represented by the pipes in the Sub-project is represented by “n-1” indicators.

One of the modalities is considered as the reference modality. Each of the other modalities is represented by a quantitative covariate (the indicator) that has the value "1" for the pipes with this modality or otherwise the value "0".

The reference modality is chosen with the help of a scrolling list to the side of the list of selected covariates.



If required, constraints for the calculations can be fixed by clicking on the button "Constrain the model". The following dialog box is displayed:



The following constraints are possible:

- Constrain Alpha: The influence of previous breaks is not considered (unchecked by default);
- Constrain Delta: The influence of ageing is not considered (unchecked by default);
- Constrain Zeta0: No selective survival bias (checked by default);
- Constrain Zeta1: No correction of the time-dependent part of the selective survival bias (checked by default);
- Constrain a covariate: The covariate is forced, i.e. it will be kept in the model even if it is not significant (unchecked by default).

VIII.5 Calculate a model

VIII.5.i Launching a calculation

Having selected a Progression, the calculation of a model can be made in the Construction mode by clicking on one of the buttons “Automatic computations” or “Semi-automatic computations situated on the right side of the “Progressions” screen.

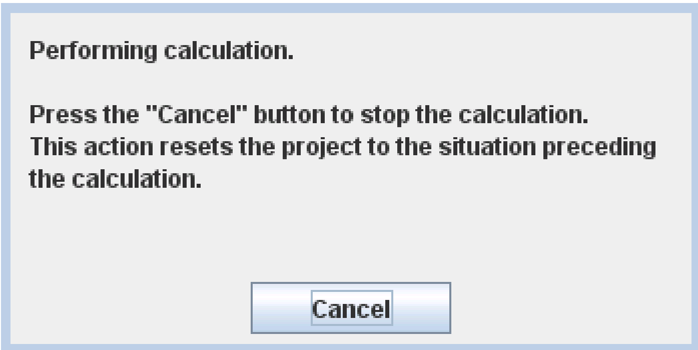
The calculation kernel **LEYP** integrated in **Casses** performs the model calibration (cf. **Erreur ! Source du renvoi introuvable.**). Other than calculate the parameters of the model, a test is performed on each of them to evaluate their significance.

From these test results, an advice module integrated in the software indicates if the obtained model is satisfactory or otherwise proposes a modification of the data entered with a view to a new calibration

In “Automatic computation” mode, the software carries out a succession of calibrations without intervention from the user until a model with all the parameters significant is created.

In “Semi-automatic computation” mode, the software carries out one calibration at a time allowing the user to intervene.

When a calibration is being made, a window allowing the interruption of the calculation is displayed.



VIII.5.ii Consulting the results

The results of the calculations are shown in the “Calibrations” screen in the Construction mode.

On the left side of the screen, a scrolling list allows the selection of the Progression whose results are to be visualised.

The right side of the screen indicates the results of the selected calibration from the list on the left side.

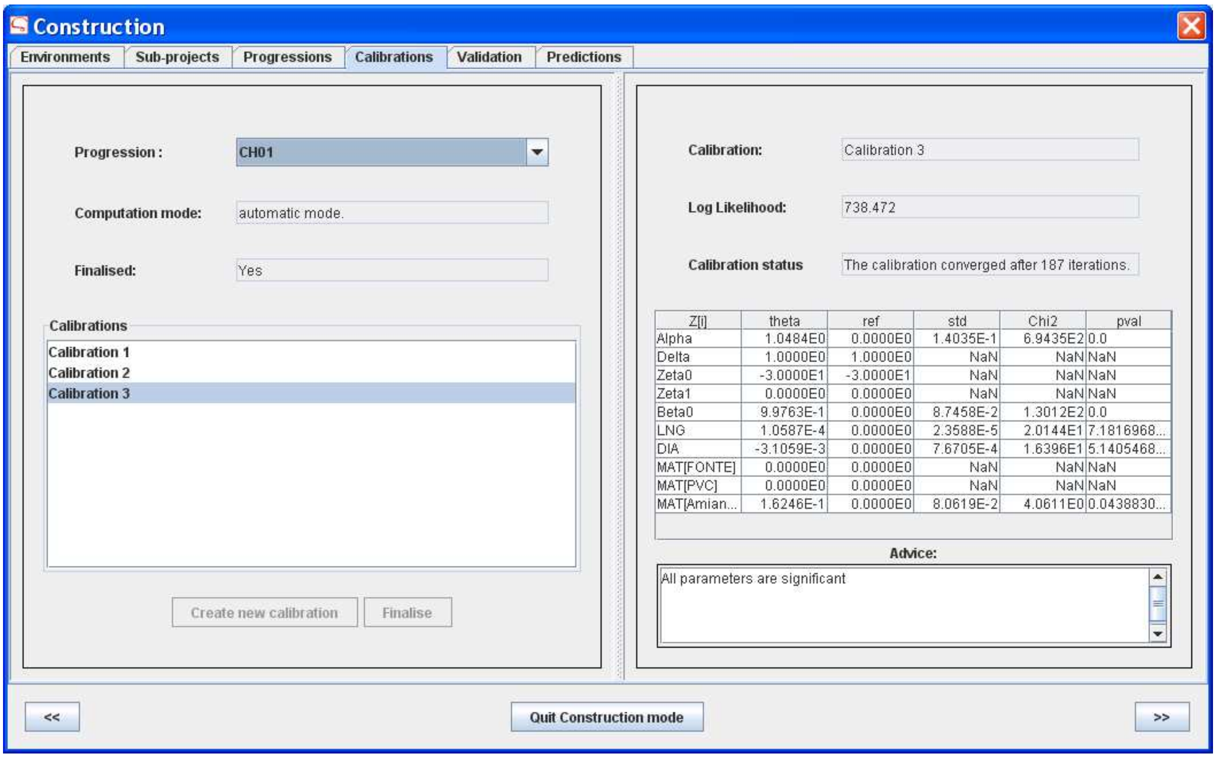
Log Likelihood is a global statistical calibration indicator. The quality of the model is better the higher its value is.

“Calibration status” indicates if the model converged and if so, the number of iterations necessary for the calibration.

A table displays the main results of the calculations. It contains the following columns:

- Z(i): Abbreviated name of parameter
- Theta: Value of parameter
- Ref: Initial value of parameter

- Std: Standard deviation
- Chi2: Chi2 value of the parameter
- Pval: p-value of the parameter



The rows of the table are as follows:

- Alpha, parameter that takes into account the influence of previous breaks; when it is not significant its value is 0
- Delta, parameter that takes into account the influence of ageing; when it is not significant its value is 1
- Zeta0, parameter which corrects the fixed part of the selective survival bias; when it is not significant its value is -∞ (in practice -30)
- Zeta1, parameter which corrects the time-dependent part of the selective survival bias; when it is not significant its value is 0
- A row for each quantitative covariate
- A row for each modality of qualitative covariates.

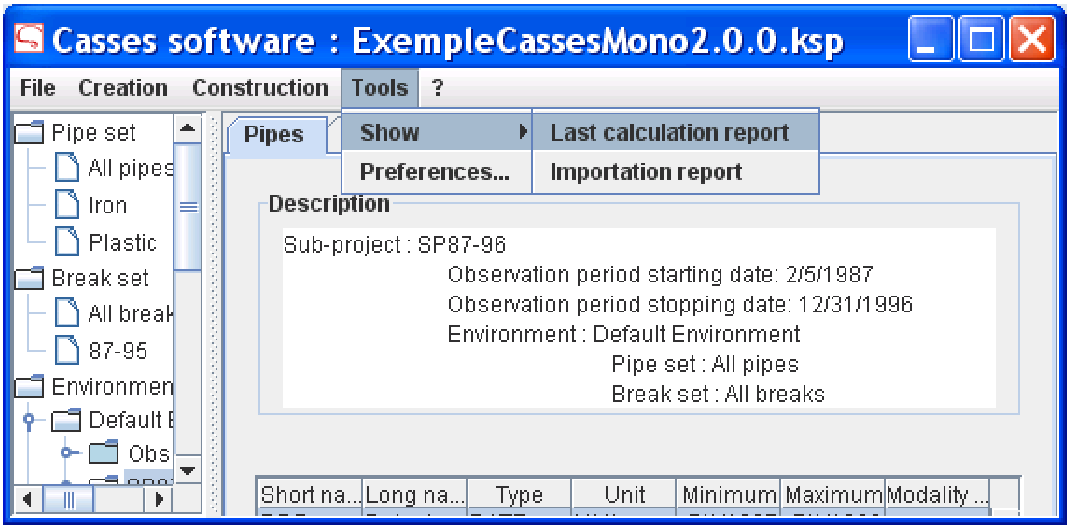
The reference modality and the modalities identified as being non-significant by the advice process have the values 0; 0; NaN; NaN; NaN.

The advice given is based on the “probability value” (Pval). A parameter is considered significant if Pval is less than 0.05.

It is a “null-hypothesis” test: Less than 5% chance of error by rejecting the hypothesis that the parameter has no effect.

It is possible that a calibration doesn’t converge (notably if there is a linear relationship between covariates). In such cases, it is necessary to create a different Progression.

It is possible to access the detailed results of the last calibration carried out by **LEYP.**



The corresponding text file is accessible in the directory specified in "Preferences" under the name "Rcal.txt".

VIII.5.iii Finalise a Progression

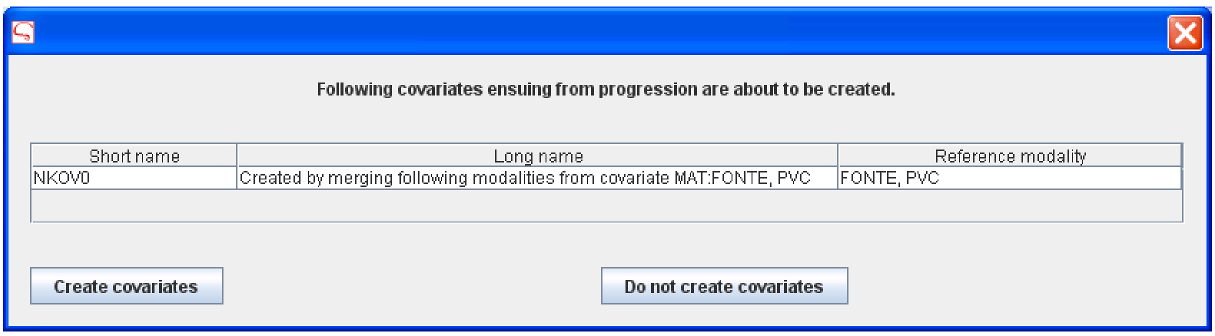
Once the calibrations of a Progression are finished, the Progression needs to be "finalised" in order to be able to make predictions.

In the case of an "Automatic computation" all the calibrations occur in succession until all the parameters are significant or they lead to non-convergence.

In the case of a "Semi-automatic computation", as long as advice can be applied, a new calibration can be performed by clicking on the button "Create new calibration" situated on the right side of the "Calibration" screen in Construction mode.

To finalise the Progression, click on the button "Finalise" situated on the right side of the "Calibration" screen in Construction mode.

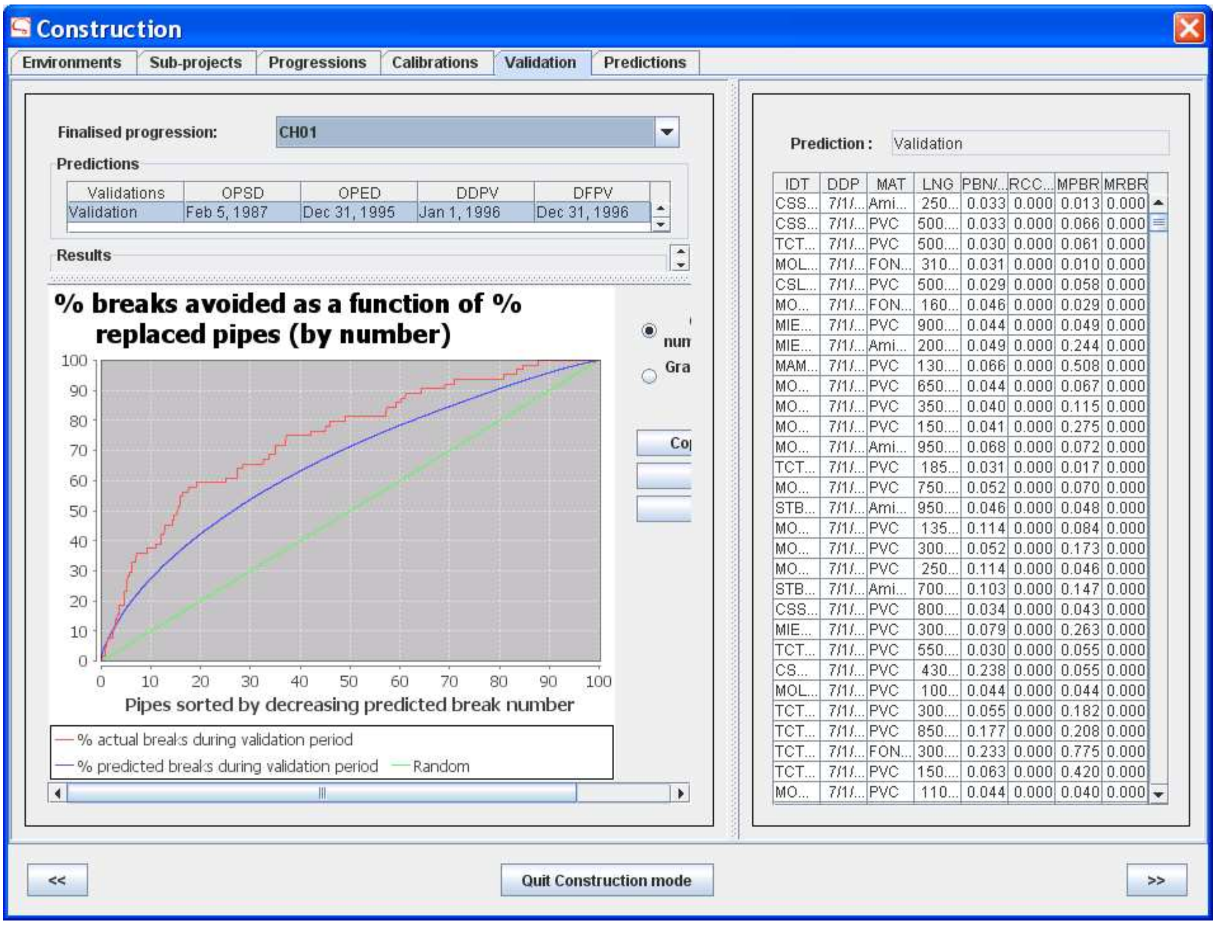
If necessary, a dialog box appears indicating the creation of new covariates that you can rename. The short name must contain no more than eight characters and respect the constraints for short names described in the data importation section (cf. II).



A new covariate is created when certain modalities of a qualitative covariate are significant whilst others are not. The non-significant modalities are merged with the reference modality. The new covariate created is therefore only useable in the Sub-Project in which the particular Progression belongs.

VIII.6 Consult a validation

When a Sub-project includes a validation period, the finalisation of the Progression launches the Validation calculations and the results are accessible from the "Validation" screen (cf. XIII.5).



VIII.6.i Left part of the "Validation" screen

The section to the top left of the "Validation" screen presents the validation indicators.

Les variables used are as follows:

- OPSD: Observation period start date
- OPED: Observation period end date
- VPSD: Validation period start date
- VPED: Validation period end date
- An: Area under the red validation curve according to number of pipes
- Al: Area under the red validation curve according to network length
- PBN: Predicted break number during the validation period
- ABN: Actual break number during the validation period

– Rn: Ratio PBN/ABN

– Cxn: % of breaks at the point on the red curve corresponding to x% of pipes (by number)

– Cxl: % of breaks at the point on the red curve corresponding to x% of pipes (by length)

– Rxn Ratio between the number of breaks predicted and the actual number of breaks for the first x% of pipes (sorted by descending annual number of breaks)

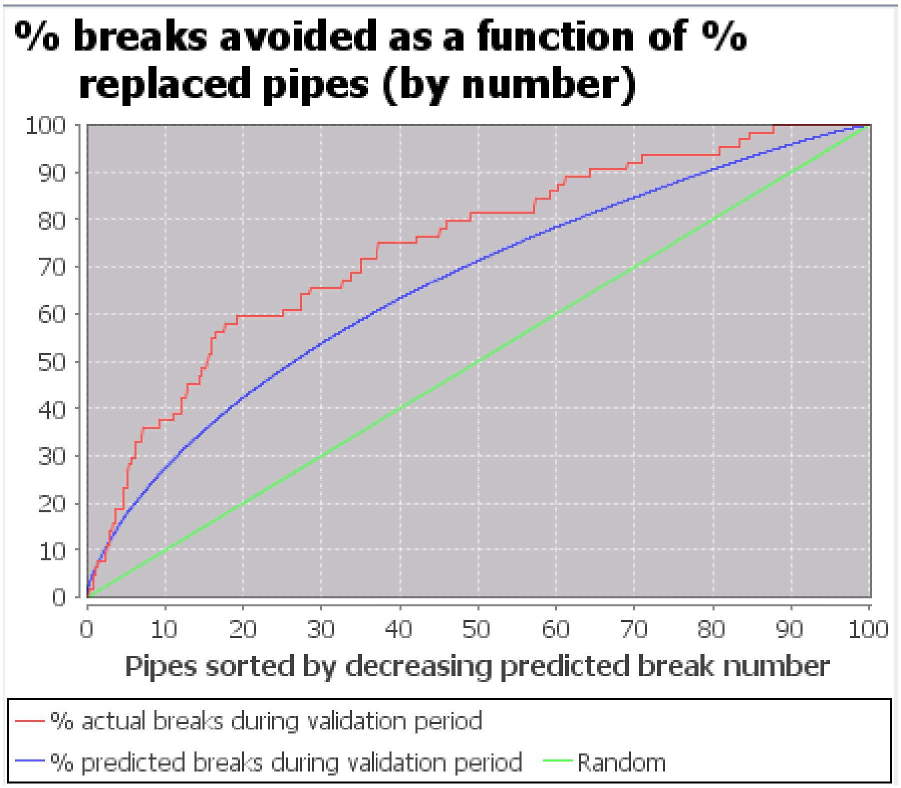
– Rxl Ratio between the number of breaks predicted and the actual number of breaks for the first x% of pipes (sorted by descending annual break rate)

The last row of the table corresponds to a value of x chosen by the user. The values of the indicators affected by the choice of x are updated in the table by clicking on the button "Calculate".

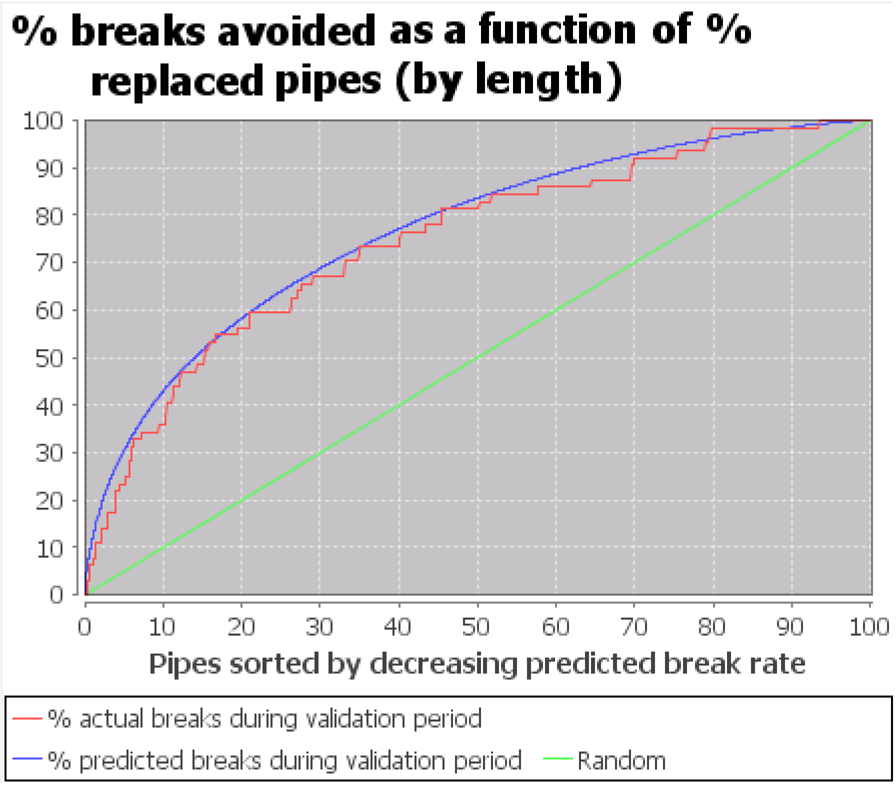
The tables are not modifiable. They can be copied to the clipboard by means of the buttons at the side.

The area at the bottom left of the screen shows a graphic visualisation of the Validation. A button allows the user to select:

– "Graph by number of pipes" for which the X-axis represents the percentage of pipes sorted by descending predicted annual break number.



– “Graph by length of pipes” for which the X-axis represents the percentage of cumulative pipe length sorted by descending predicted annual breaks rate.



For each of the graphs three curves are represented:

– The red curve (that allows the indicators to be calculated) represents the percentage of actual breaks as a function of x

– The blue curve represents the percentage of predicted breaks as a function of x

– The green curve y=x (simulates random behaviour)

Each graph can be copied, saved (in .png format) or printed. It is possible to zoom in to a part of the graph by clicking then dragging towards the bottom right. To zoom out click and drag to the top left.

VIII.6.ii Right part of the “Validation” screen

The right section of the “Validation” screen consists of a table which displays for each pipe concerned:

– The compulsory data (IDT, DDP, MAT, LNG).

– PBN/yr, the annual number of breaks predicted by the model over the validation period.

– ABN/yr, the annual number of actual breaks during the validation period.

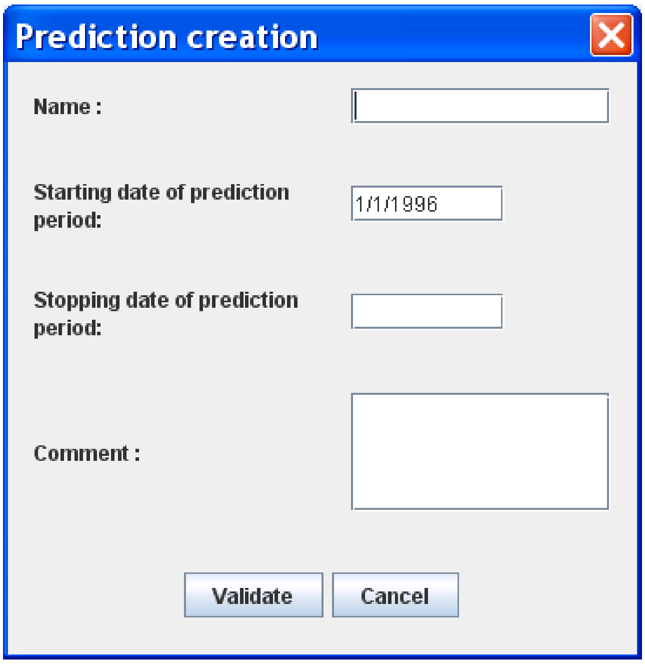
– PBR and ABR, the predicted and actual break rates in breaks per km per year.

The table can be sorted by clicking on the column headers. It can be partially copied (select with the mouse and then <CTRL> + C) or copied in its entirety (click on the table and then <CTRL> + A and then <CTRL> + C).

VIII.7 Create a “Prediction”

To create a “Prediction” a finalised “Progression” has to be associated with a prediction period. The prediction period is a period of any length that occurs after the observation period of the Sub-project.

The creation of a new Prediction is carried out from the left of the “Predictions” screen in the Construction mode. A scrolling list allows the selection of the finalised Progression of the model on which the prediction is based. Click on the button “Create new prediction”.



Firstly, the Prediction must be named. The name can be chosen freely, the only constraint being that two Predictions in the same Progression cannot share the same name.

Next, the dates for the start and end of the prediction period must be defined. By default the start date is the day after the end of the observation period. The start date must be after the observation period. The end date must be after the start date.

A text box for optional additional comments is available for specifying the nature of the Prediction.

.

The right part of the “Predictions” screen consists of a table displaying the results for each pipe present in the prediction period selected in the left section.

PBN is the number of breaks predicted in the prediction period.

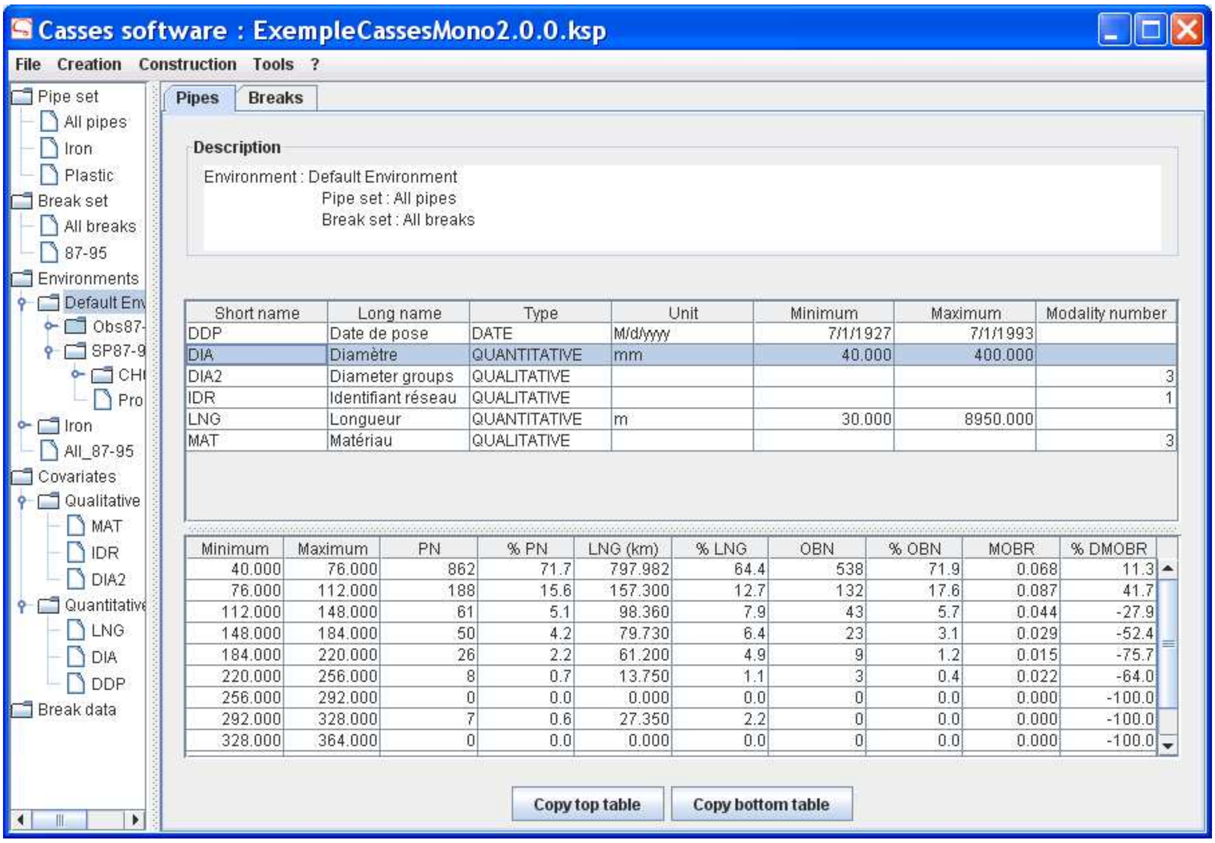
MPBR is the mean predicted break rate, in breaks per km per year.

The table can be copied by clicking on the column headers.

It can be partially copied (select with the mouse and then <CTRL> + C) or copied in its entirety (click on the table and then <CTRL> + A and then <CTRL> + C).

IX Exploring the results

IX.1 Exploring an Environment



When you select an Environment in the Exploration window, you have access to two pages on the right, accessed by the tabs “Pipes” and “Breaks”.

At the top of each of these pages is a section describing composition of the Environment.

A second section, the top table, displays a list and description of all the covariates for the “Pipes” tab and all the data for the “Breaks” tab.

The bottom table describes the element selected in the top table.

The first row of the table contains the column labels.

For qualitative data, the first column displays the modalities.

For quantitative data other than the date removed from service (DHS) or break date (DDC), the range of values are divided into ten equal intervals and the first two columns display the upper and lower bounds for each interval; the intervals include the upper bound value.

For DHS, the data are also subdivided into ten intervals but an extra interval]01/01/2998 ; 01/01/2999] is added regrouping all the pipes in service (DHS imported blank).

For DDC, the first column contains the break year and the table has one row per year.

The last row is the row "TOTAL" which concerns the whole Environment.

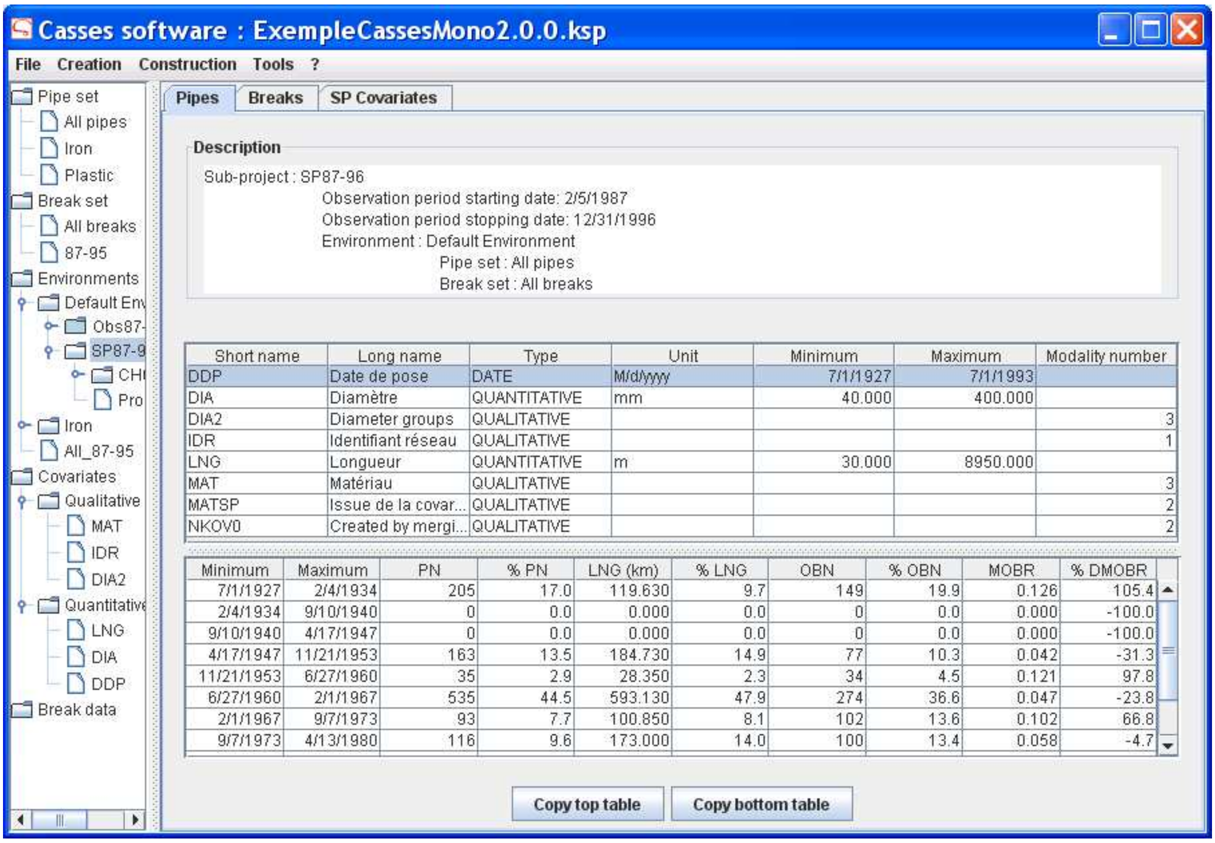
The signification of the column titles of the lower table is as follows:

PN	Number of pipes
% PN	Percentage of the total number of pipes
LNG (km)	Length of the pipes concerned in kilometres
% LNG	Percentage of the total pipe length
OBN	Number of breaks observed on the pipes concerned
% OBN	Percentage of the total number of breaks observed
MOBR	Mean observed break rate in breaks per kilometre per year
% DMOBR	Percentage difference from mean overall observed break rate

IX.2 Exploring a Sub-project

The Exploration screens for Sub-projects differ depending on whether there is a Validation or not. In the navigation tree, the icon symbolising a Sub-project with Validation is shaded grey.

IX.2.i Sub-project without Validation



Selecting a Sub-project in the Exploration window gives access to three pages on the right under the tabs “Pipes”, “Breaks” and “SP Covariates”.

The “Pipes” and Breaks” pages are organised in the same way as with the Environments.

It is important to note that the pipes and breaks of a Sub-project and those of the Environment in which it belongs are not necessarily the same. The Sub-project doesn’t consider:

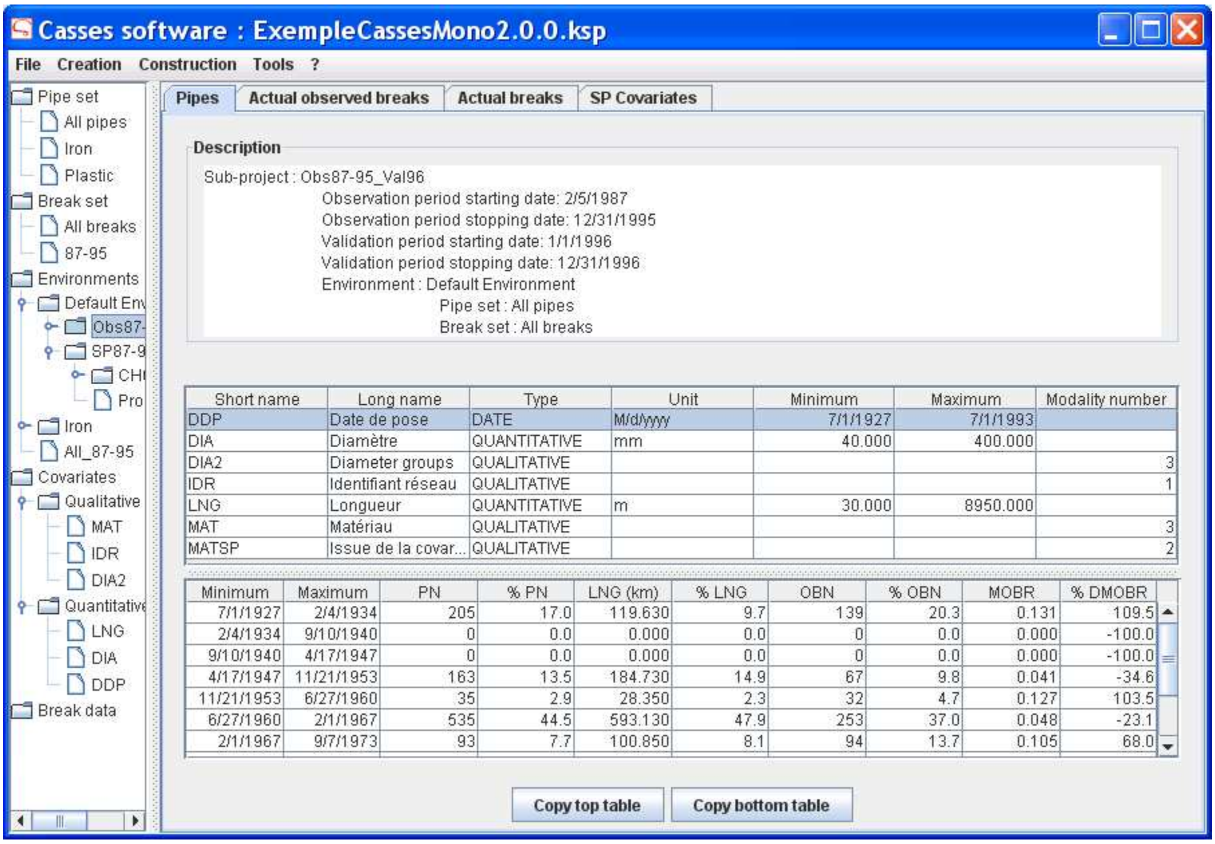
- Breaks outside the observation period,
- Pipes for which the window of observation is empty.

The “SP Covariates” page only displays covariates that were present at the finalisation of one or more Progressions in the Sub-project.

The page displays from top to bottom:

- The description of the Sub-project
- A table of the covariates involved
- A section indicating the values of the selected covariate
- A section indicating how the selected covariate was created

IX.2.ii Sub-project with Validation



Selecting a Sub-project in the Exploration window gives access to four pages on the right under the tabs “Pipes”, “Observed breaks”, “Actual breaks” and “SP Covariates”.

The “Pipes” and “SP Covariates” pages are the same as for a Sub-project without Validation and the “Observed breaks” page is similar to the “Breaks” page.

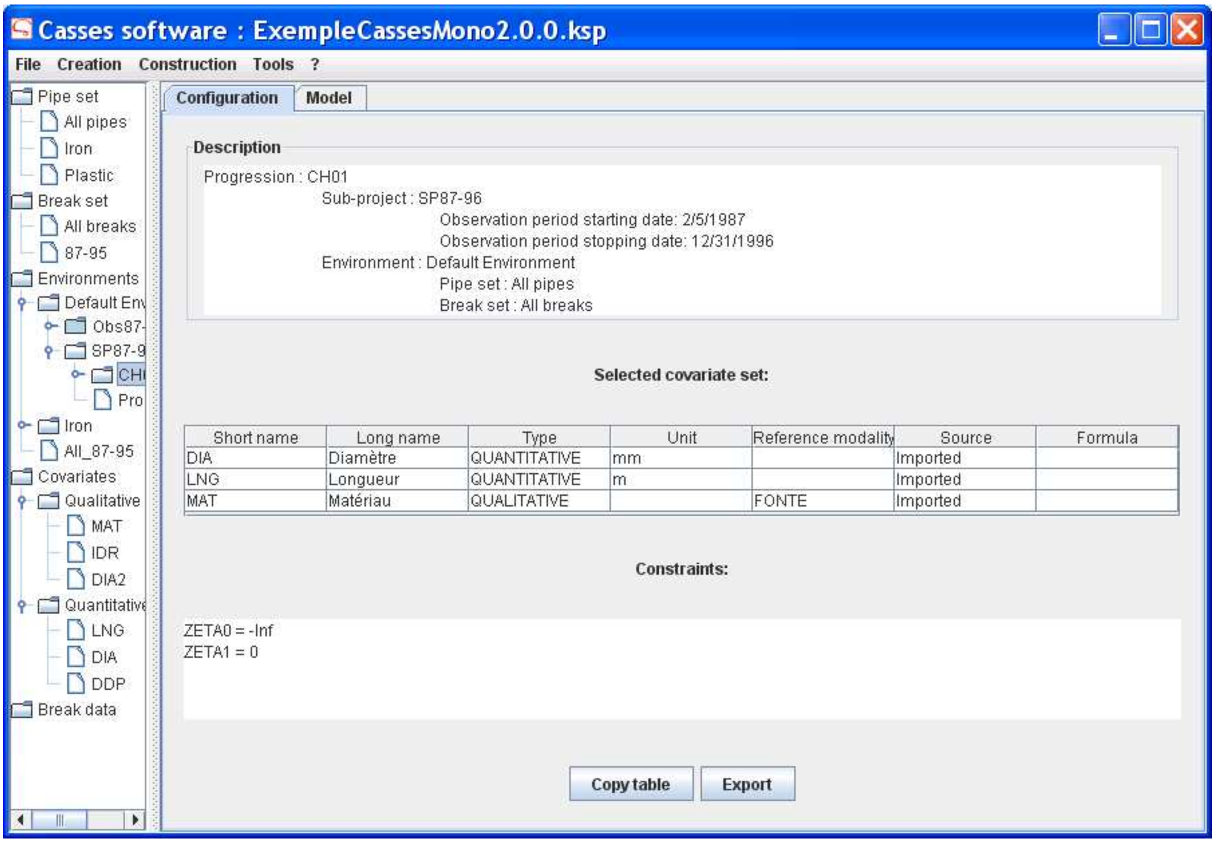
The “Actual breaks” page presents information concerning breaks occurring during the validation period and is in the same form as the “Observed breaks”.

IX.3 Exploring a Progression

Selecting a Progression in the Exploration window gives access to two pages on the right under the tabs “Configuration” and “Model”.

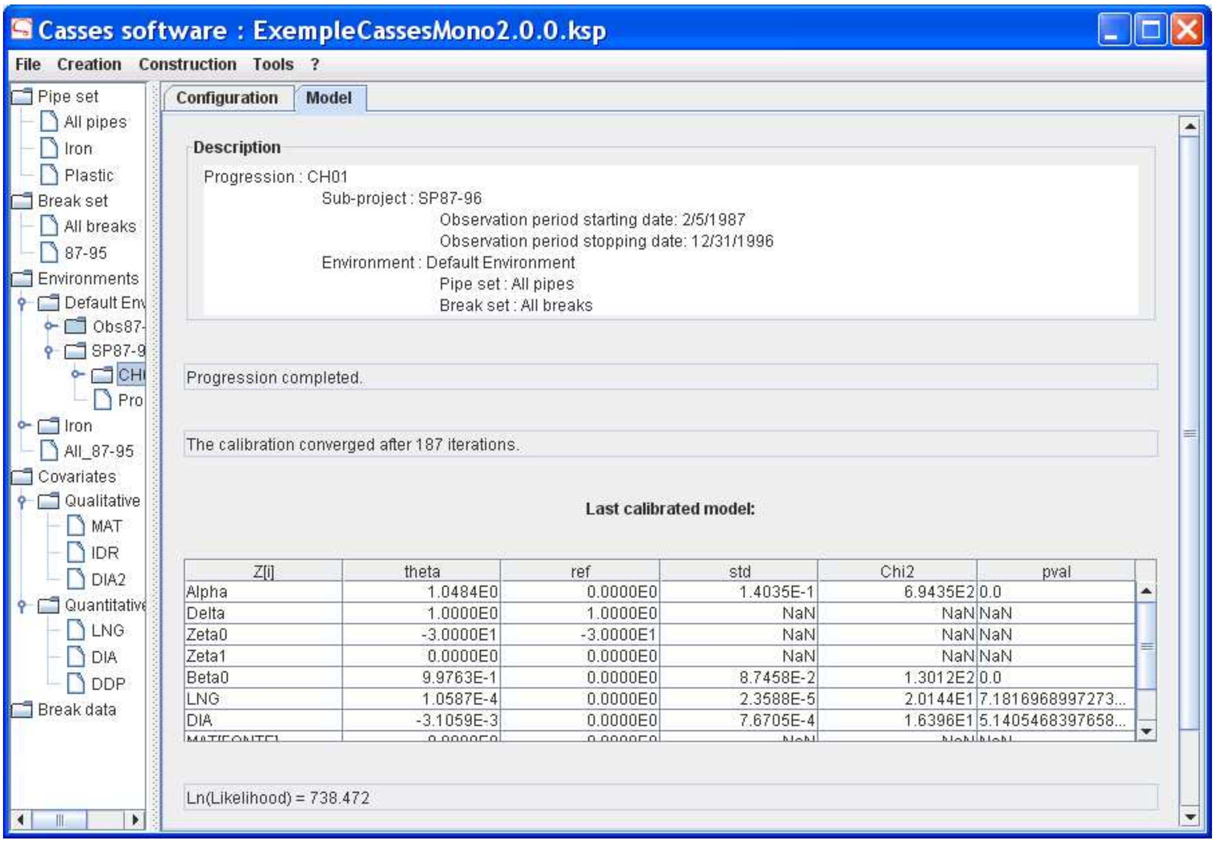
IX.3.i “Configuration” page

The “Configuration” page describes the different components of the Progression (cf. VIII.4).

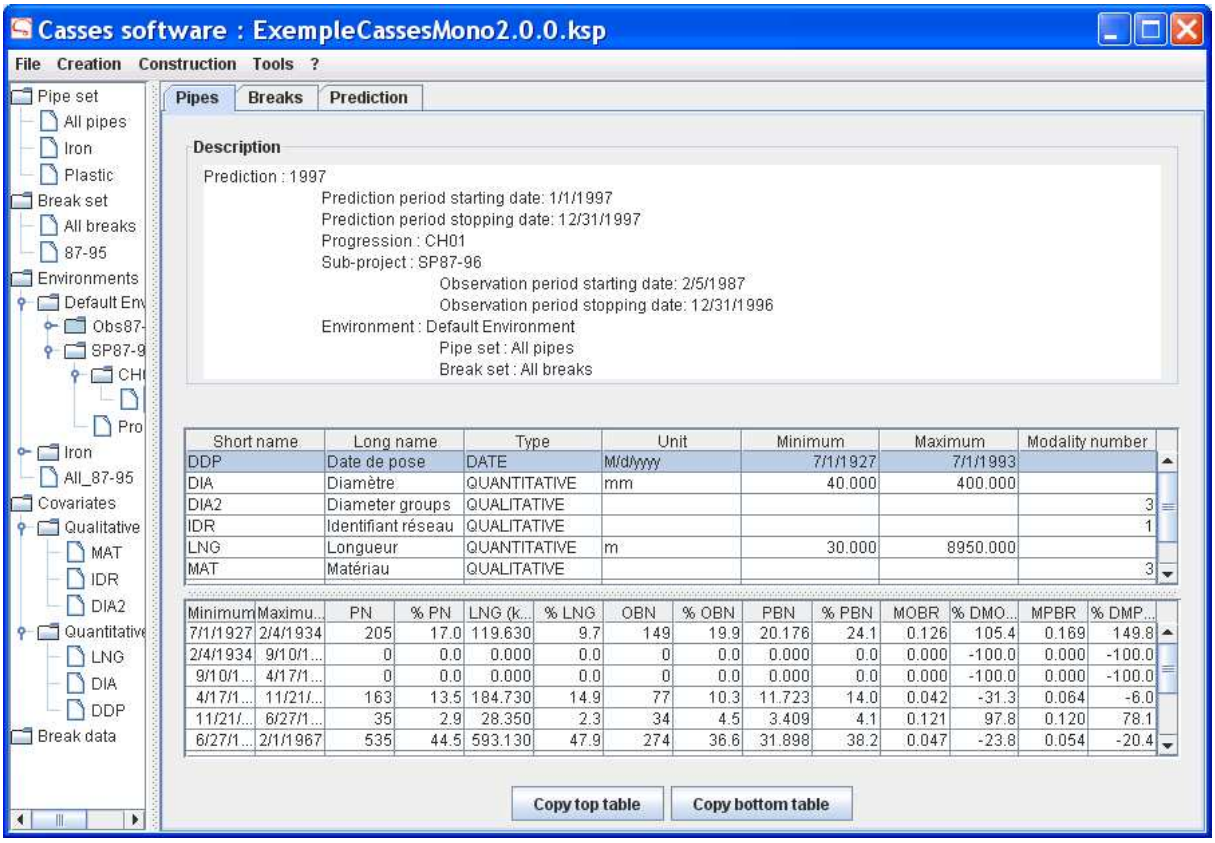


IX.3.ii “Model” page

The “Model” page describes the different components of the model (results of the last calibration, cf. **Erreur ! Source du renvoi introuvable.**) after which the Progression was finalised.



IX.4 Exploring a prediction



Selecting a Prediction in the Exploration window gives access to three pages on the right under the tabs “Pipes”, “Breaks” and “Prediction”.

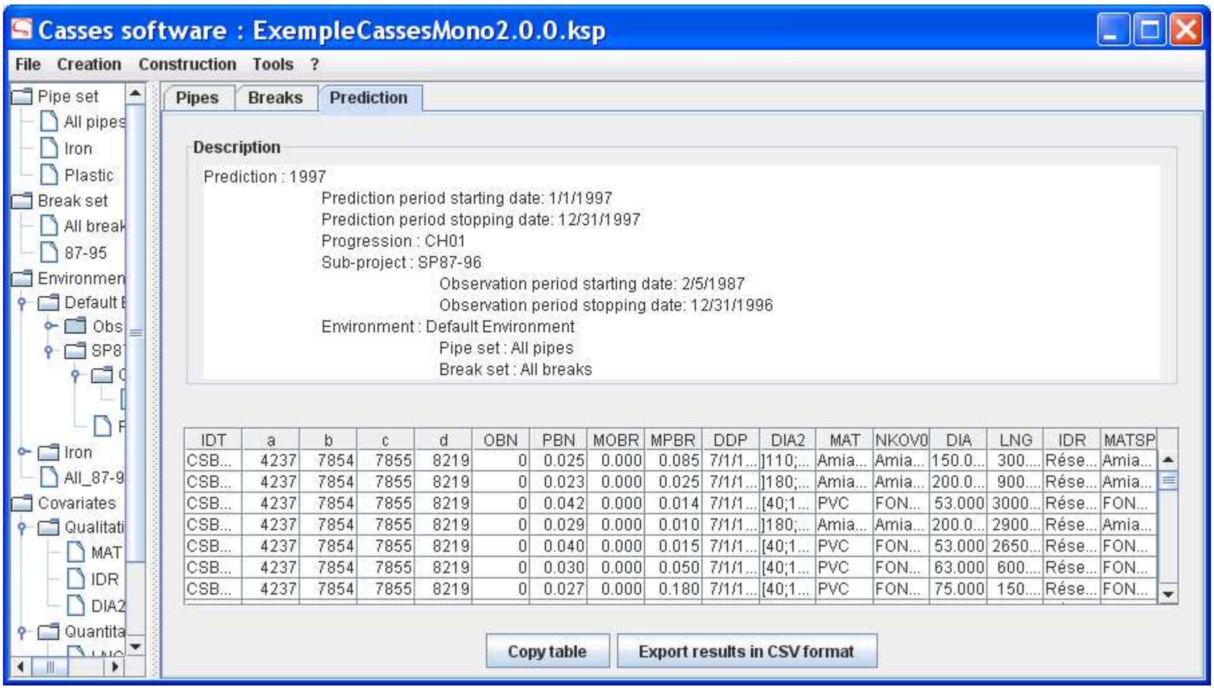
The “Pipes” and “Breaks” pages are generally organised in the same way as for the Environment or Sub-project but contain additional information concerning the predicted breaks. The significance of the abbreviated titles in the tables is as follows:

PN	Number of pipes
% PN	Percentage of the total number of pipes
LNG (km)	Length of the pipes concerned in kilometres
% LNG	Percentage of the total pipe length
OBN	Number of breaks observed on the pipes concerned
% OBN	Percentage of the total number of breaks observed
PBN	Number of breaks predicted on the pipes concerned
% PBN	Percentage of the total number of breaks predicted
MOBR	Mean observed break rate in breaks per kilometre per year
% DMOBR	Percentage difference from mean overall observed break rate
MPBR	Mean predicted break rate in breaks per kilometre per year
% DMPBR	Percentage difference from mean overall predicted break rate

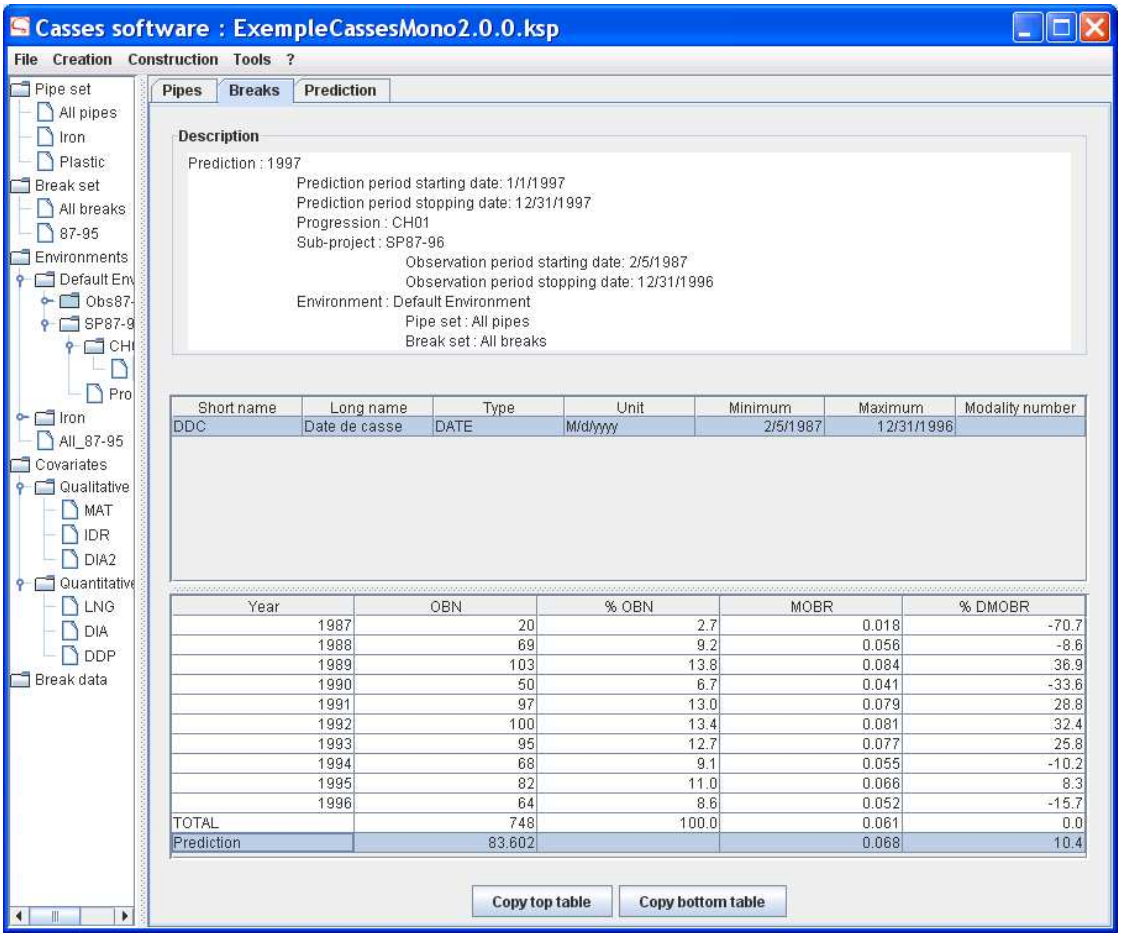
It is important to note that the pipes and breaks of a Prediction and those of the Sub-project in which it belongs are not necessarily the same. Pipes with an empty prediction window and the breaks associated with them are not considered.

The “Prediction” page contains a top section describing the Prediction and its filiations. The bottom section displays a table showing for each pipe in the Prediction, its ID, the covariate values, the number and rate of breaks observed and predicted and the values of a, b, c, defined thus:

- a, age of the pipe in days at the start of the observation period
- b, age of the pipe in days at the end of the observation period
- c, age of the pipe in days at the start of the prediction period
- d, age of the pipe in days at the end of the prediction period

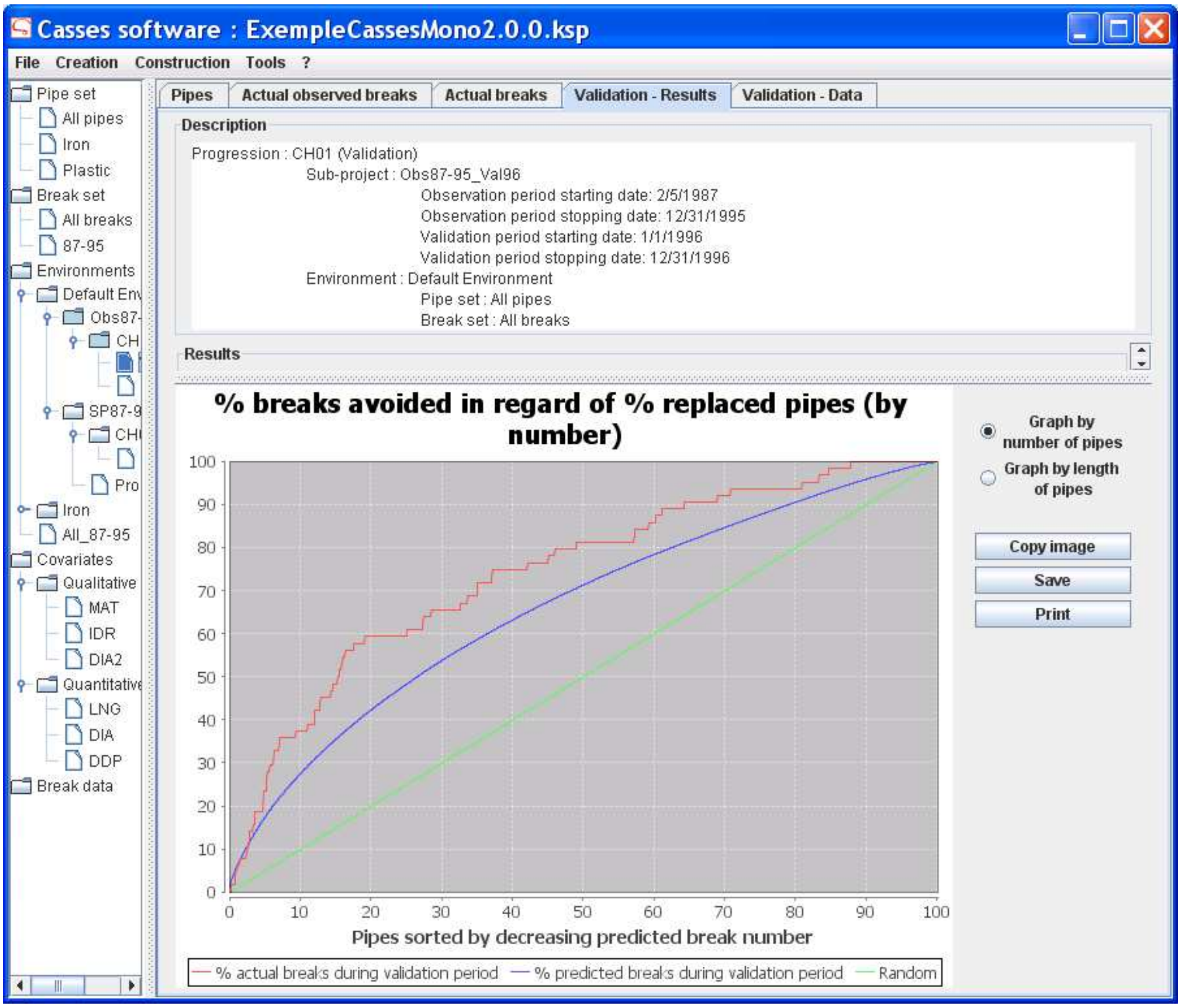


On the "Breaks" page, when selecting "Break date" (DDC), the table at the bottom includes an extra row entitled "Prediction" containing the number and rate of breaks **PREDICTED**.



IX.5 Exploring a Validation

Validations are located at the same level as Predictions in the Exploration window and are distinguished by a grey icon.



Selecting a Validation in the Exploration window gives access to five pages on the right under the tabs “Pipes”, “Observed Breaks”, “Actual breaks”, “Validation – Results” and “Validation – Data”.

The “Pipes” page is identical to that for a “Prediction”.

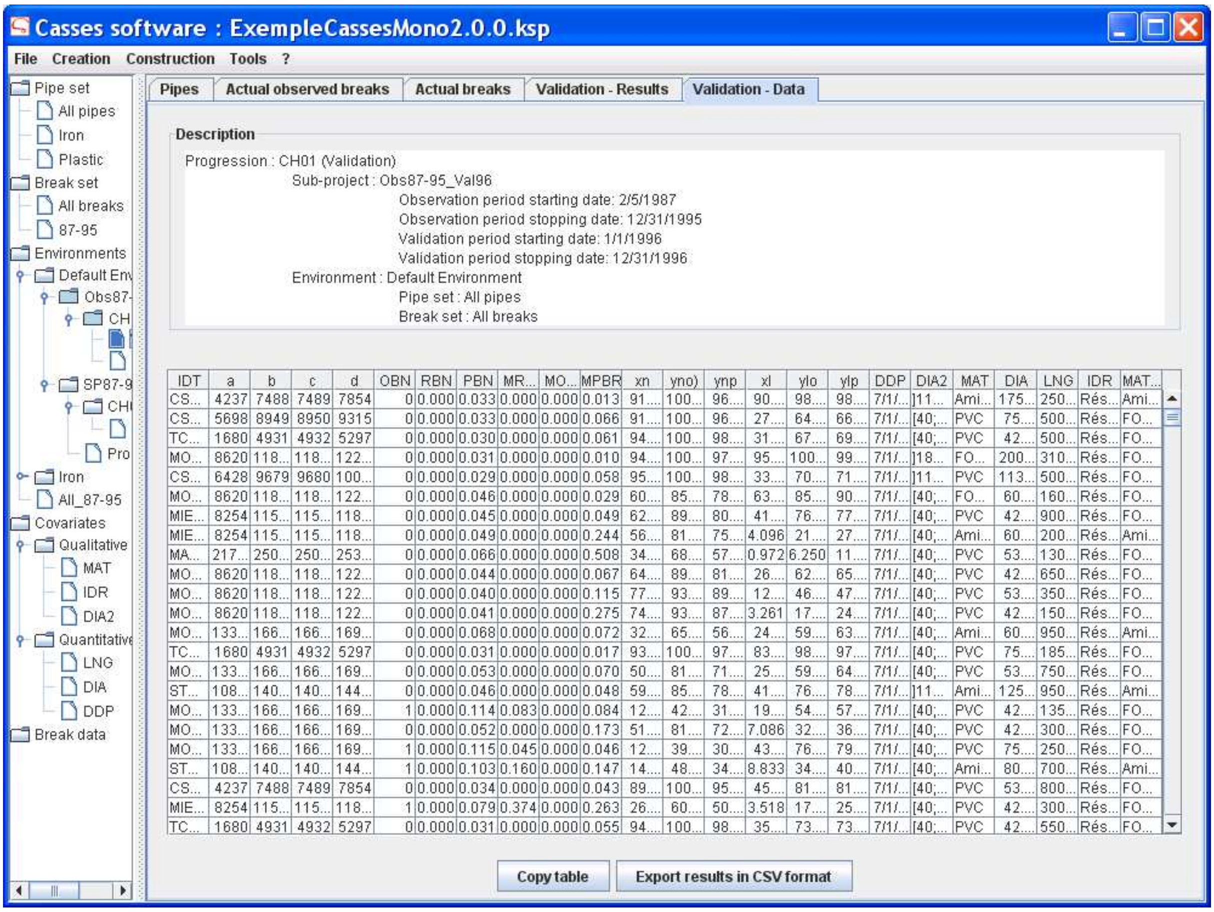
The “Observed breaks” page is similar to that for a “Sub-project”.

The “Actual breaks” page is also similar to that for a “Sub-project” but also contains information relating to predicted breaks when DDC is selected in the top table.

The “Validation – Results” page presents the indicators and graphics of the Validation in an identical fashion as the left part of the “Validation” tab in Construction mode (cf. VIII.6.i).

The “Validation – Data” page is equivalent to the “Prediction” page associated with a Prediction but contains, in addition, the following information:

- ABN, actual number of breaks during the validation period
- ABR, actual break rate during the validation period
- xn, yno and ynp, the x and y co-ordinates (o for observed and p for predicted) permitting the construction of the graph according to pipe number
- xl, ylo and ylp, the x and y co-ordinates (o for observed and p for predicted) permitting the construction of the graph according to pipe length.



IDT	a	b	c	d	OBN	RBN	PBN	MR...	MO...	MPBR	xn	yno)	ynp	xl	ylo	ylp	DDP	DIA2	MAT	DIA	LNG	IDR	MAT...
CS...	4237	7488	7489	7854	0	0.000	0.033	0.000	0.000	0.013	91....	100...	96....	90....	98....	98....	7/1/...	]11...	Ami...	175...	250...	Rés...	Ami...
CS...	5698	8949	8950	9315	0	0.000	0.033	0.000	0.000	0.066	91....	100...	96....	27....	64....	66....	7/1/...	[40;...	PVC	75....	500...	Rés...	FO...
TC...	1680	4931	4932	5297	0	0.000	0.030	0.000	0.000	0.061	94....	100...	98....	31....	67....	69....	7/1/...	[40;...	PVC	42....	500...	Rés...	FO...
MO...	8620	118...	118...	122...	0	0.000	0.031	0.000	0.000	0.010	94....	100...	97....	95....	100...	99....	7/1/...	]18...	FO...	200...	310...	Rés...	FO...
CS...	6428	9679	9680	100...	0	0.000	0.029	0.000	0.000	0.058	95....	100...	98....	33....	70....	71....	7/1/...	]11...	PVC	113...	500...	Rés...	FO...
MO...	8620	118...	118...	122...	0	0.000	0.046	0.000	0.000	0.029	60....	85....	78....	63....	85....	90....	7/1/...	[40;...	FO...	60....	160...	Rés...	FO...
MIE...	8254	115...	115...	118...	0	0.000	0.045	0.000	0.000	0.049	62....	89....	80....	41....	76....	77....	7/1/...	[40;...	PVC	42....	900...	Rés...	FO...
MIE...	8254	115...	115...	118...	0	0.000	0.049	0.000	0.000	0.244	56....	81....	75....	4.096	21....	27....	7/1/...	[40;...	Ami...	60....	200...	Rés...	Ami...
MA...	217...	250...	250...	253...	0	0.000	0.066	0.000	0.000	0.508	34....	68....	57....	0.972	6.250	11....	7/1/...	[40;...	PVC	53....	130...	Rés...	FO...
MO...	8620	118...	118...	122...	0	0.000	0.044	0.000	0.000	0.067	64....	89....	81....	26....	62....	65....	7/1/...	[40;...	PVC	42....	650...	Rés...	FO...
MO...	8620	118...	118...	122...	0	0.000	0.040	0.000	0.000	0.115	77....	93....	89....	12....	46....	47....	7/1/...	[40;...	PVC	53....	350...	Rés...	FO...
MO...	8620	118...	118...	122...	0	0.000	0.041	0.000	0.000	0.275	74....	93....	87....	3.261	17....	24....	7/1/...	[40;...	PVC	42....	150...	Rés...	FO...
MO...	133...	166...	166...	169...	0	0.000	0.068	0.000	0.000	0.072	32....	65....	56....	24....	59....	63....	7/1/...	[40;...	Ami...	60....	950...	Rés...	Ami...
TC...	1680	4931	4932	5297	0	0.000	0.031	0.000	0.000	0.017	93....	100...	97....	83....	98....	97....	7/1/...	[40;...	PVC	75....	185...	Rés...	FO...
MO...	133...	166...	166...	169...	0	0.000	0.053	0.000	0.000	0.070	50....	81....	71....	25....	59....	64....	7/1/...	[40;...	PVC	53....	750...	Rés...	FO...
ST...	108...	140...	140...	144...	0	0.000	0.046	0.000	0.000	0.048	59....	85....	78....	41....	76....	78....	7/1/...	]11...	Ami...	125...	950...	Rés...	Ami...
MO...	133...	166...	166...	169...	1	0.000	0.114	0.083	0.000	0.084	12....	42....	31....	19....	54....	57....	7/1/...	[40;...	PVC	42....	135...	Rés...	FO...
MO...	133...	166...	166...	169...	0	0.000	0.052	0.000	0.000	0.173	51....	81....	72....	7.086	32....	36....	7/1/...	[40;...	PVC	42....	300...	Rés...	FO...
MO...	133...	166...	166...	169...	1	0.000	0.115	0.045	0.000	0.046	12....	39....	30....	43....	76....	79....	7/1/...	[40;...	PVC	75....	250...	Rés...	FO...
ST...	108...	140...	140...	144...	1	0.000	0.103	0.160	0.000	0.147	14....	48....	34....	8.833	34....	40....	7/1/...	[40;...	Ami...	80....	700...	Rés...	Ami...
CS...	4237	7488	7489	7854	0	0.000	0.034	0.000	0.000	0.043	89....	100...	95....	45....	81....	81....	7/1/...	[40;...	PVC	53....	800...	Rés...	FO...
MIE...	8254	115...	115...	118...	1	0.000	0.079	0.374	0.000	0.263	26....	60....	50....	3.518	17....	25....	7/1/...	[40;...	PVC	42....	300...	Rés...	FO...
TC...	1680	4931	4932	5297	0	0.000	0.031	0.000	0.000	0.055	94....	100...	98....	35....	73....	73....	7/1/...	[40;...	PVC	42....	550...	Rés...	FO...

X Exporting the results

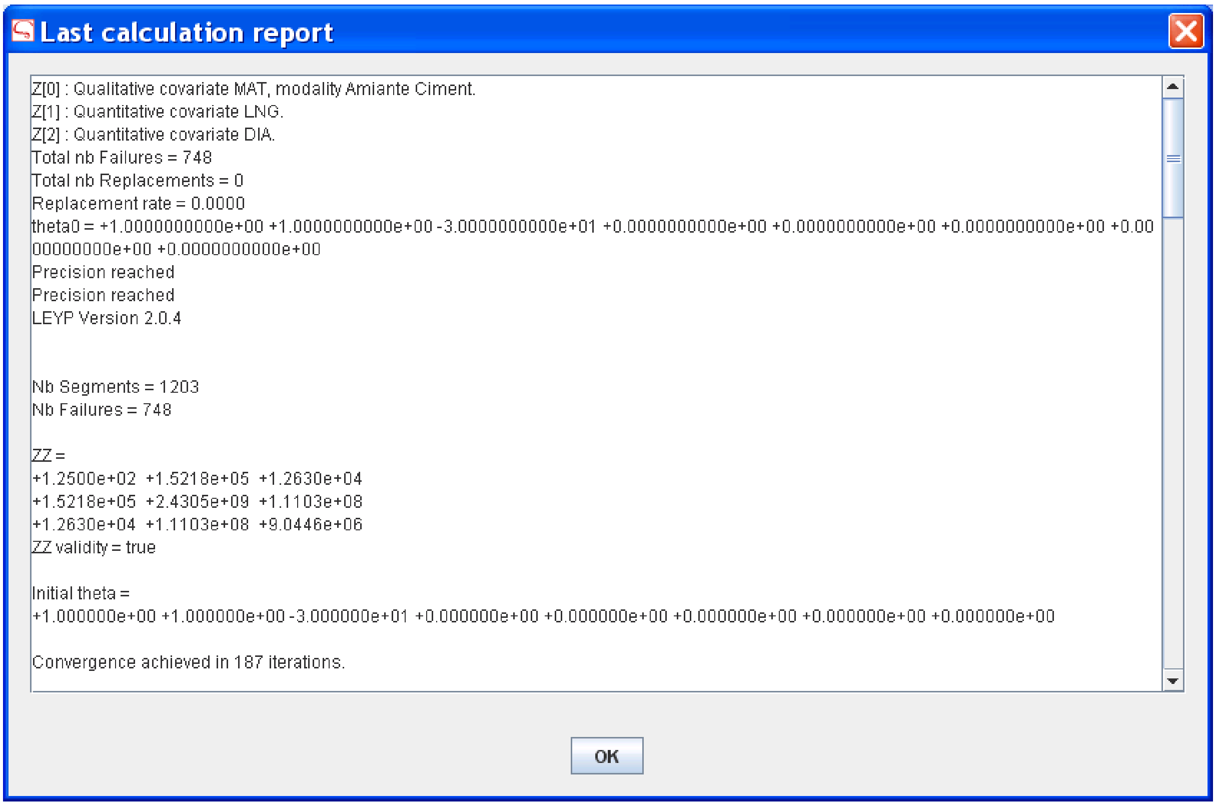
X.1 Exporting intermediate results

All the tables in the software can be partially or totally copied by selecting the required cells and the <CTRL> + C or using the contextual buttons intended for this purpose.

The importation report can be saved as a .csv file with the help of the button shown during the importation phase. It is also accessible via the menu “Tools/Show/Importation report”.

Pipe sets and Break sets can be exported in the **Casses** format by selecting them in the Exploration window and then right clicking “Export”

The detail of the last calibration is accessible via the menu “Tools/Show/Last calculation report”.



X.2 Exporting the predictions

Each Prediction made with the help of **Casses** can be exported in .csv format:

- By selecting it in the Exploration window and then right clicking "Export",
- By selecting the Prediction in the Exploration window and then selecting the "Prediction" tab and clicking on the button "Export results in CSV format".

The filename is chosen by the user: "UserName.csv"

The first lines are as follows:

#Project name
#Environment name
#Sub-project name
#Progression name
#Prediction name
#Casses filename.ksp
#Pipes filename.csv
#Breaks filename.csv
#Pipe set name
#Break set name
#Recording period start date
#Recording period end date
#Observation period start date
#Observation period end date
#Prediction period start date
#Prediction period end date
IDT;PBN;PBR

Dates are in the format d/m/y.

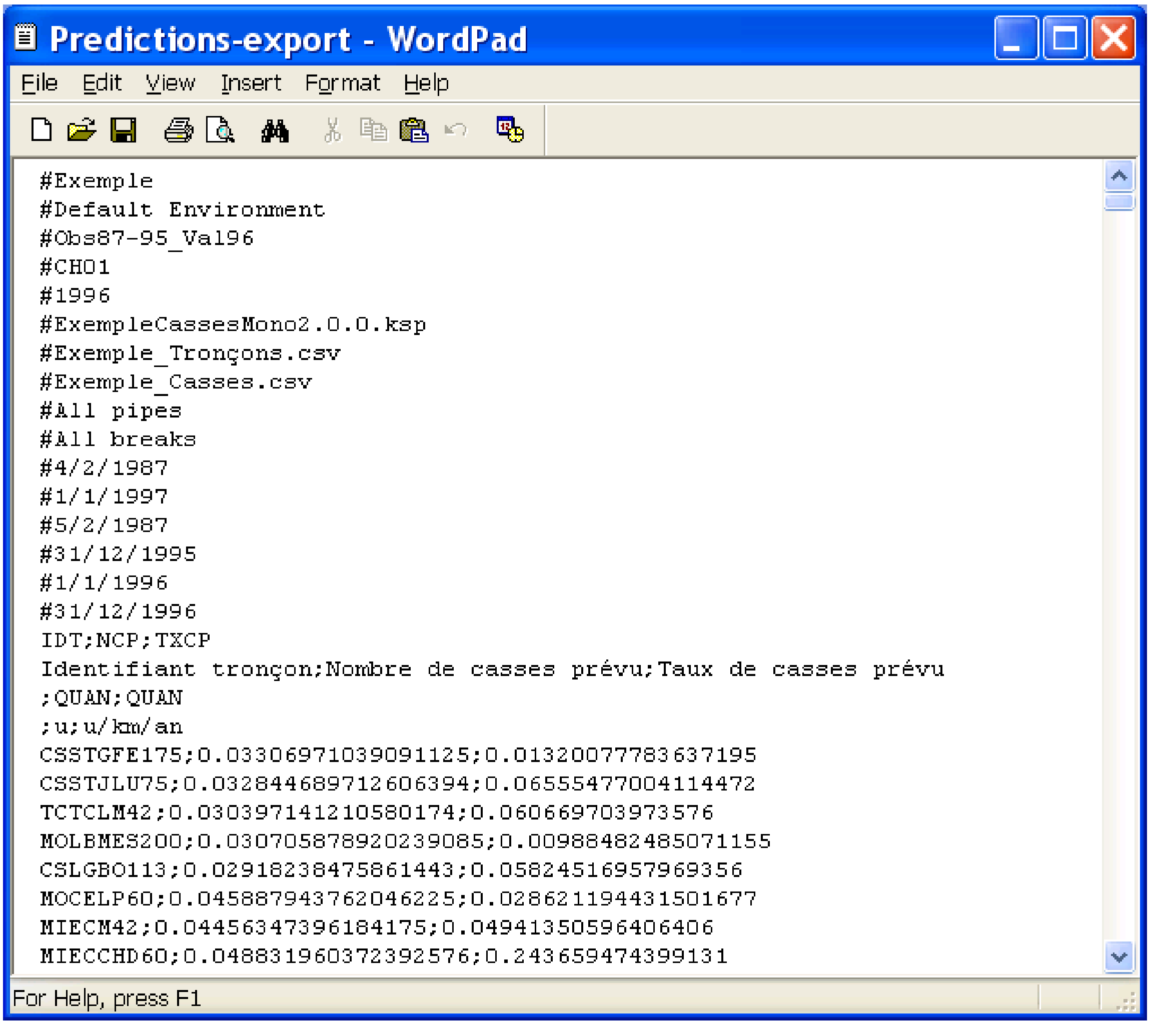
IDT is the pipe ID.

PBN is the number of breaks predicted over the prediction period.

PBR is the predicted break rate in breaks per kilometre per year.

Then there is one row per pipe with the different values being separated by a semi-colon.

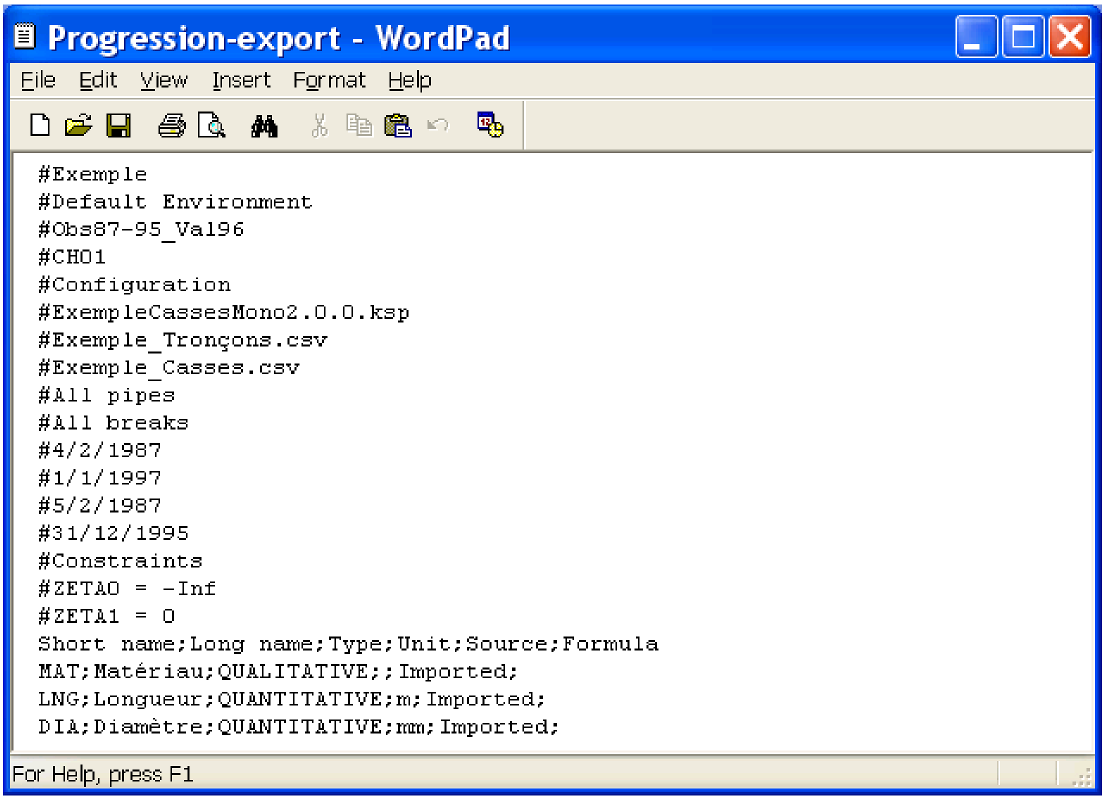
The exportation is carried out in the same way for a Validation, the prediction period being replaced with the validation period.



X.3 Exporting the results of a Progression

X.3.i Exporting a configuration

Information on the configuration of a Progression can be exported in .csv format.




```
#Exemple
#Default Environment
#Obs87-95_Val96
#CHO1
#Configuration
#ExempleCassesMono2.0.0.ksp
#Exemple_Tronçons.csv
#Exemple_Casses.csv
#All pipes
#All breaks
#4/2/1987
#1/1/1997
#5/2/1987
#31/12/1995
#Constraints
#ZETA0 = -Inf
#ZETA1 = 0
Short name;Long name;Type;Unit;Source;Formula
MAT;Matériau;QUALITATIVE;;Imported;
LNG;Longueur;QUANTITATIVE;m;Imported;
DIA;Diamètre;QUANTITATIVE;mm;Imported;
```

The first rows are as follows:

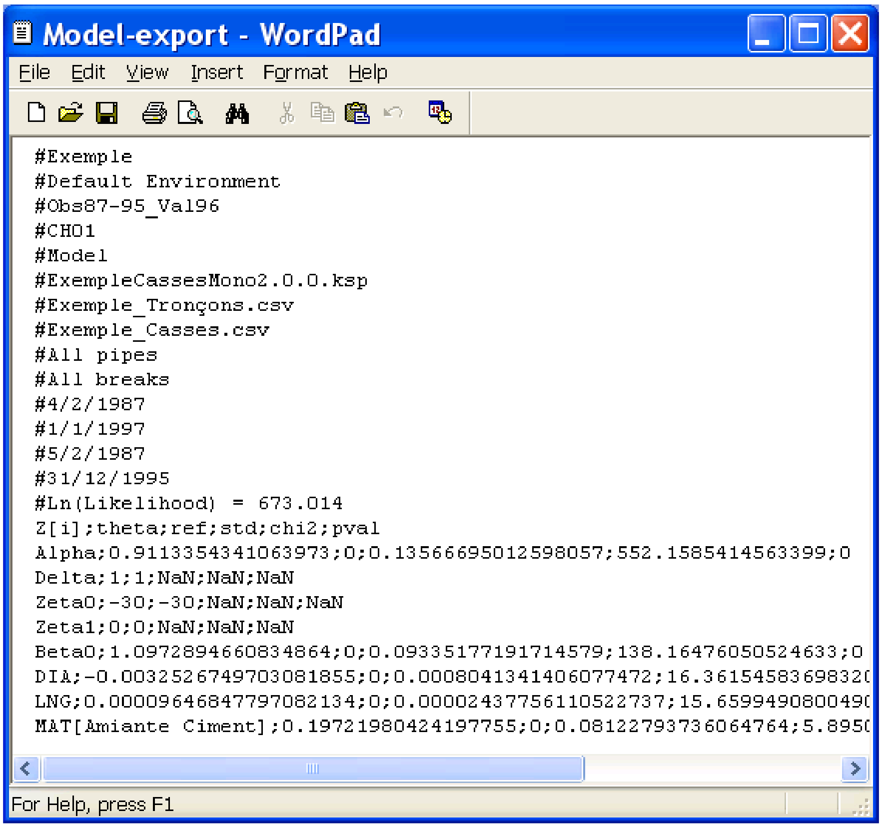
#Project name
#Environment name
#Sub-project name
#Progression name
#Casses filename.ksp
#Pipes filename.csv
#Breaks filename.csv
#Pipe set name
#Break set name
#Recording period start date
#Recording period end date
#Observation period start date
#Observation period end date
#Constraints
#A row starting with # for each constraint
Short name;Long name;Type;Unit;Source;Formula

Then there is one row per covariate with the different values being separated by a semi-colon.

Dates are in the format d/m/y.

X.3.ii Exporting a model

Information on the model parameters of a finalised Progression can be exported in .csv format.



The first rows are as follows:

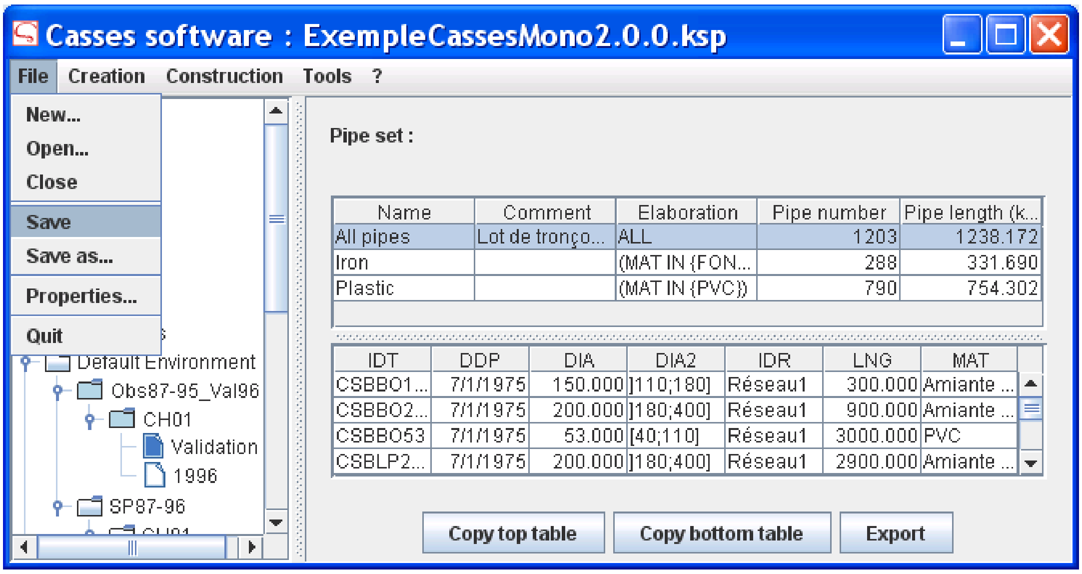
#Project name
#Environment name
#Sub-project name
#Progression name
#Casses filename.ksp
#Pipes filename.csv
#Breaks filename.csv
#Pipe set name
#Break set name
#Recording period start date
#Recording period end date
#Observation period start date
#Observation period end date
#Ln(Likelihood) = *value of the model*
Z[i];theta;ref;std;chi2;pval

Then there is one row per model parameter with the different values being separated by a semi-colon.

Dates are in the format d/m/y.

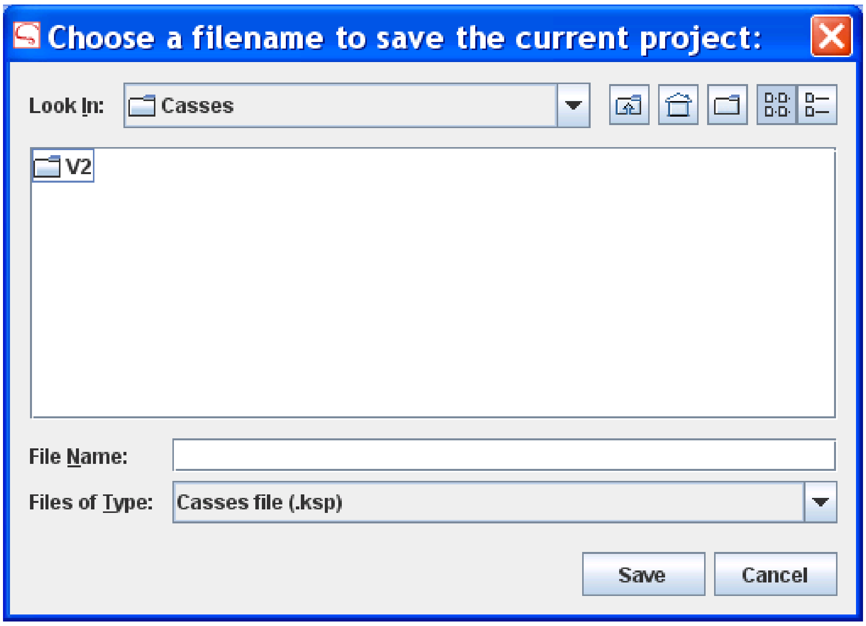
XI Closing a work session

XI.1 Save a project



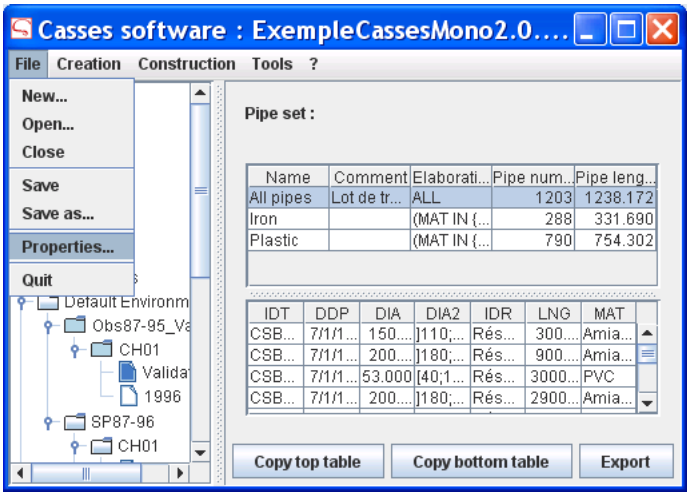
From the “File” menu the “Save” and “Save as...” functions can be accessed.

“Save as...” opens a dialog box which allows the filename and location to be chosen. The extension for a saved Project is .ksp.

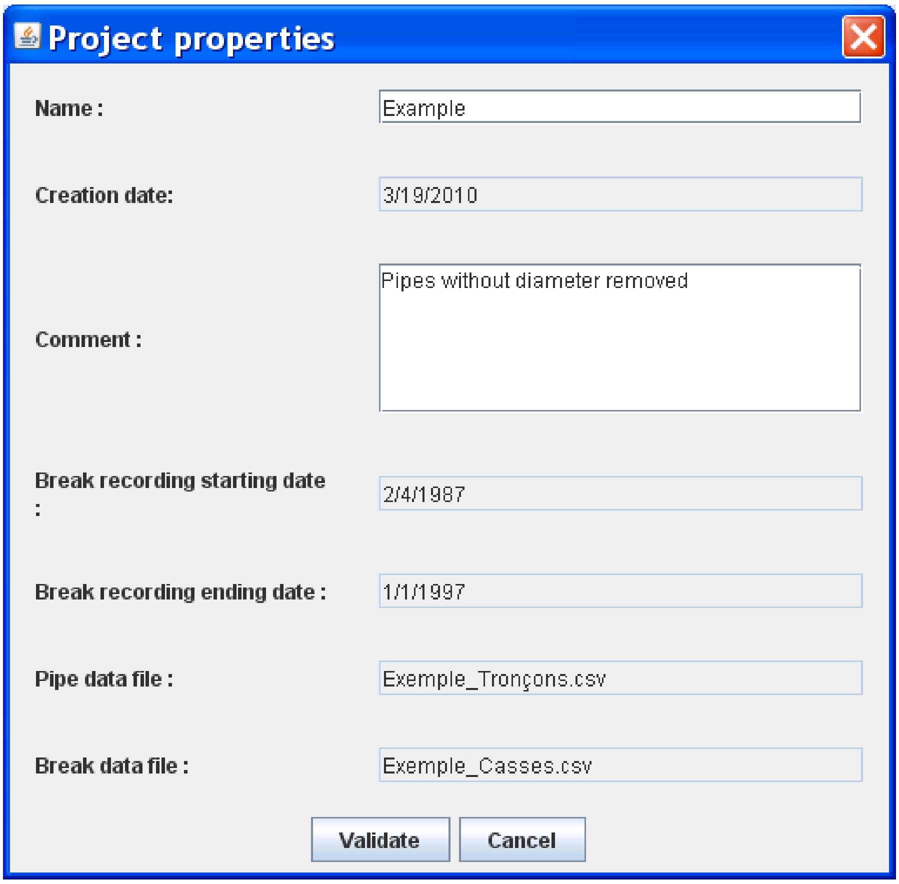


“Save” directly saves a Project that has already been saved without changing its name.

XI.2 Properties of a Project



The “Properties” of a Project are general information that is saved.

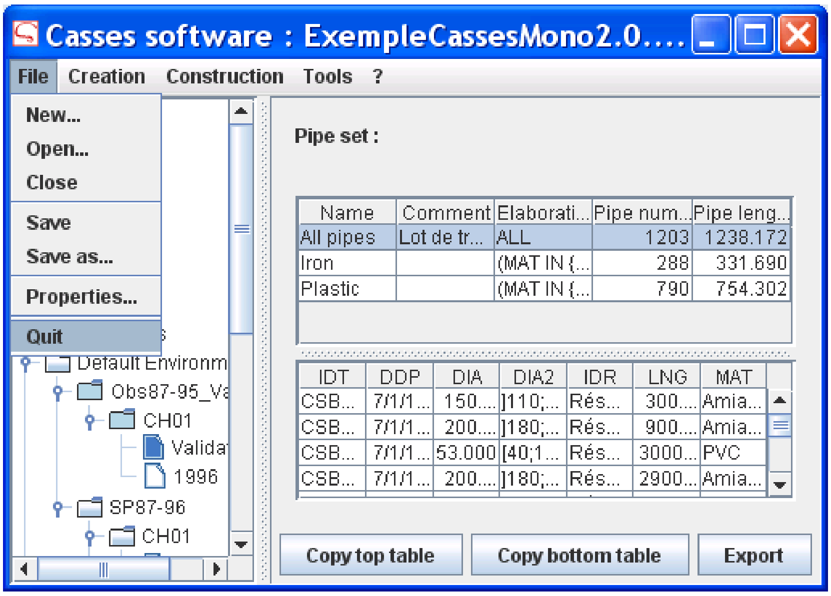


You can choose any Project name, the default being the first comment line in the Pipes file.

The “Comment” box is also freely defined, the default being the second comment line in the Pipes file.

The other properties are saved at the moment the data are imported. They are not modifiable.

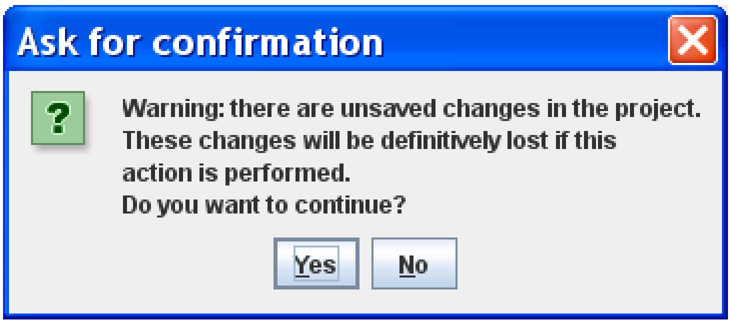
XI.3 Closing a Project



“Close”, closes the current Project but keeps the **Casses** application open.

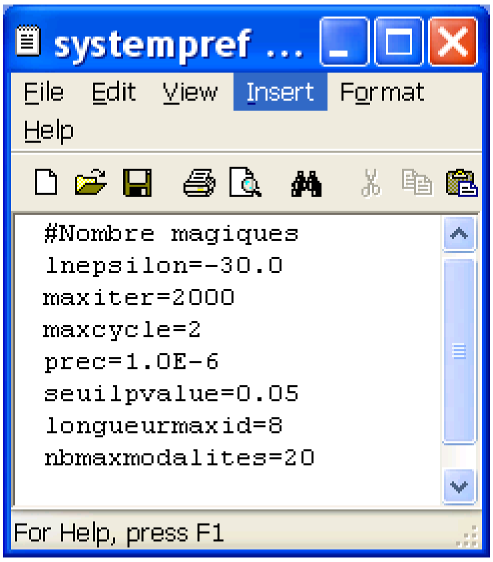
“Quit”, closes both the current Project and the application.

In both cases if the Project has not been saved, a dialog box invites confirmation of the closure.



XII Configuration of system preferences

In the installation directory of **Casses**, the file “systempref.ini” contains certain parameters used in the software.



For normal use of **Casses** these parameters do not need modifying and it is not recommended for a user to do so without advice from the assistance service..

The parameters are as follows:

- lnepsilon, threshold for calculation precision
- maxiter, maximum number of LEYP iterations
- maxcycle, maximum number of LEYP cycles
- prec, relative precision of parameters
- seuilpvalue, threshold for p-val for the advice module
- longueurmaxid, maximum number of characters for short names
- nbmaxmodalites, maximum number of modalities for qualitative covariates

To be taken into account, modification of the parameters must be made before running **Casses**.

XIII Annexes

XIII.1 Glossary

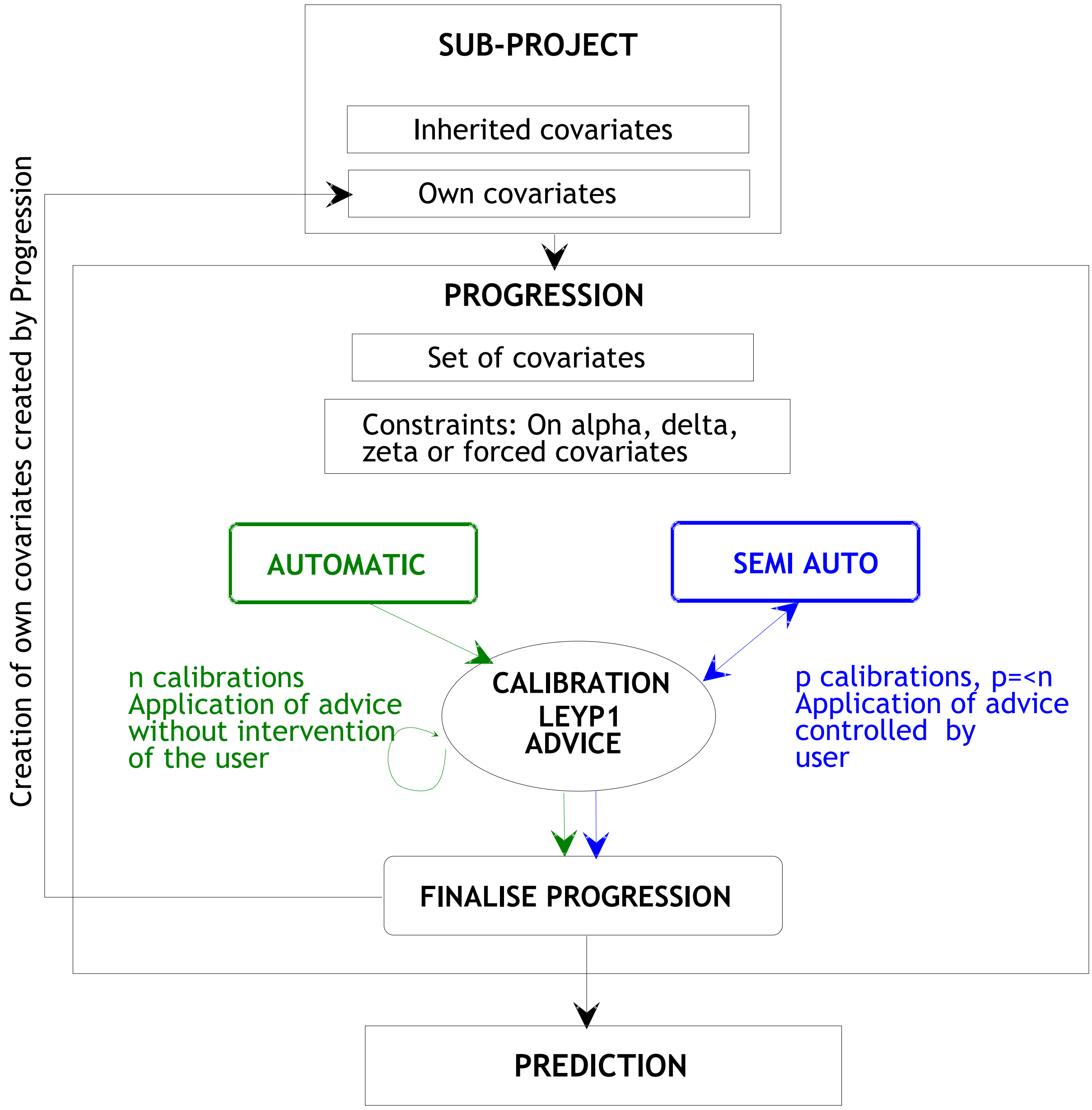
Actual number of breaks	By convention, the number of actual breaks concerns the real breaks occurring during the prediction window when it is included in the recording window (as distinct from the number of breaks observed which only concerns breaks during the recording period).
Break (or failure)	A rupture or leak on a pipe that leads to a repair.
Break set	A collection of breaks selected from those in the Project. It only contains breaks with valid data.
Calibration	A part of a Sub-project corresponding to an execution of the LEYP kernel applied to a set of covariates with view to calibrating a model (LEYP1 calculation). A calibration corresponds to a single execution for a fixed set of covariates.
Calibration period	In Validation mode it is the observation period used for calibrating the model.
Collection of Pipes	A group of pipes imported for use in a Project. Whether they belong to one or several networks, all the pipes are described with the same set of characteristics.
Covariate	Data attached to a pipe that is the function of one or more pipe characteristics (this function can be the ID). A covariate has a unique value for a single pipe. It can potentially (but not compulsory) be used in a model.
Covariate modality	Any particular value of a covariate.
Covariates set	A collection of covariates selected from those eligible associated with a particular pipe set.
Eligible covariate	The covariates that do not have a unique value for the pipe set concerned. Only eligible covariates can be part of a set of covariates linked to a calibration.
Environment	A part of a Project regrouping the Sub-projects created from a Pipe set and a Break set.
Environment recording period	The period delimited by the earliest recording start date and the earliest recording end date of the networks for which at least one pipe is included in the Environment.

Finalising a Progression	An action carried out on a Progression enabling predictions to be made from the last calibration as well as the creation (if necessary) of new covariates associated with the calibration. It is not possible to make further calibrations on a finalised Progression.
Forced covariate	An eligible covariate from a Sub-project that the user wished to include in the model even if it is not significant.
Likelihood	The maximum value for the likelihood function of a model. For each model, this value is calculated by the LEYP kernel; the result shown is the logarithm. A model has a closer fit the greater the value of the log(likelihood).
Mean observed break rate	The sum of the number of pipe breaks during their window of observation divided by the sum of the product of the pipe lengths and their observation window duration. $$\overline{\delta} = \frac{\sum_{i=1}^{n} Nf_i}{\sum_{i=1}^{n} L_i \times Df_i}$$
Modality indicators of a qualitative covariate	A quantitative covariate deriving from a qualitative one that takes the value 1 for pipes with the modality considered or else the value 0. In the statistical model, a qualitative covariate with n modalities is represented by n-1 indicators associated with n-1 modalities with the non-represented modality being referred to as the "reference modality".
Network recording period	The period delimited by the start and end dates for which breaks associated with pipes in the network have been observed.
Observation period	A past period during which break observed on the studied networks are taken into account in the calculations.
Pertinent covariate modality	A covariate modality is pertinent for a collection of pipes if at least one of the pipes has this modality as a value. Only pertinent covariate modalities can be included in calibrations.
Pipe	A collection of adjacent pipeline segments for which all the characteristics (other than length) have the same value.

Pipe break rate	The number of breaks during a window of time divided by the pipe length and by the window duration. It is expressed in breaks per kilometre per year. $\delta = \frac{Nf}{L \times Df}$
Pipe characteristic	Data attached to a pipe describing its physical nature, its environment or its function.
Pipe network (or network)	A collection of pipes used for distribution in a geographic zone and for which homogenous information is available. In this application, service pipes and fittings are not included.
Pipe observation window	A period delimited by two dates during which the pipe is in service and the breaks on it are observed and recorded.
Pipe set	A collection of pipes selected from those in the Project. It only contains pipes with valid data.
Prediction	A part of a Sub-project regrouping the operations made and results obtained with the aid of the software for calculating break predictions from the data in the Sub-project for a defined set of covariates and a fixed prediction period.
Prediction period	A period during which break predictions are made. The start date of the prediction period is after the end date of the observation period.
Prediction window	The period delimited by two dates for which the pipe is in service and a prediction of breaks is calculated.
Progression	A succession of calibrations for which each new calibration (with the exception of the first) is determined by modifying the set of covariates in respect of the results of the previous calibration.
Project	A project regroups all the operations and results obtained using the software from a collection of source data from one or more networks.
Qualitative covariate	A covariate with a limited number of numeric or alphanumeric values (modalities). *Note: The terms "quantitative" and qualitative" can be used in the same way for break data or pipe characteristics.*
Quantitative covariate	A covariate with a measured value, expressed as a value with a unit. It can be used directly in a numerical calculation.

Recording window	A period delimited by two dates during which the pipe is in service and the breaks on it are recorded.
Stratification	The process of sub-dividing groups of pipes for which a qualitative covariate (imported or created) has the same value. The stratification of a collection of breaks is also possible.
Sub-project	A part of a Project regrouping the operations made and results obtained with the aid of the software from a Pipe set and a Break set over a fixed observation period. A Sub-project can belong to only one Environment.
Validation	The comparison of break predictions with actual breaks that have occurred during a defined period. In Validation mode the recording period is divided into two successive periods, the calibration period and the validation period.
Validation mode	The context of the calculations allowing the calculation of validity indicators. Validation mode is only applicable on Sub-projects where the observation period ends before the recording period for breaks in the collection of networks.
Validation period	In Validation mode, this is the prediction period. It immediately follows the calibration period and is included in the period for which breaks have been recorded for all networks in the Project.
Validity indicator	Values calculated from a Prediction made in Validation mode and giving a measure of the predictive performance of the model.

XIII.2 Schematisation of calculation process



XIII.3 Definitions and rules relative to dates

XIII.3.i Denominations and calculations of dates and ages

Abbreviation	Definition	Calculation
BRSD	Break record start date for network	1st January if imported in years.
BRED	Break record end date for network	31st December if imported in years.
DDP	Installation date of pipe	1st January if imported in years.
DHS	Removal date of pipe	31st December if imported in years.
DDC	Break date	1st July if imported in years.
t	Age of the pipe at the time of a break	t = DDC - DDP
BWSD	Break recording window start date for pipe	BWSD = max (BRSD; DDP)
BWED	Break recording window end date for pipe	BWED = min (BRED; DHS)
ESD	Environment break record period start date	EST = min (BRSDi) for the network participants
EED	Environment break record period end date	EED = min (BREDi) for the network participants
OPSD	Observation period start date	Fixed by the user
OPED	Observation period end date	Fixed by the user
OWSD	Observation window start date for pipe	OWSD = max (OPSD; BWSD)
a	Age of pipe at observation start date	a = OWSD - DDP
OWED	Observation window end date for pipe	OWED = min (OPED; BWED)
b	Age of pipe at observation end date	b = OWED - DDP
PPSD	Prediction period start date	Fixed by the user
PPED	Prediction period end date	Fixed by the user
PWSD	Prediction window start date for pipe	PWSD = max (PPSD; DDP)
c	Age of pipe at prediction start date	c = PWSD - DDP
PWED	Prediction window end date for pipe	PWED = min (PPED; DHS)
d	Age of pipe at prediction end date	d = PWSD - DDP
VPSD	Validation period start date (equivalent to PPSD for Validation)	Fixed by the user
VPED	Validation period end date (equivalent to PPED for Validation)	Fixed by the user

To describe a passage of time, the term **period** is used when a collection of pipes is concerned and the term **window** is used when it concerns a single pipe.

For pips still in service, DHS is not defined therefore it can be ignored in the rules it appears (for example, min (BRED; DHS) = BRED).

The “window” of a pipe relative to the “period” only has a sense if the pipe respects certain constraints:

- Pipes with DHS ≤ BRSD or DDP ≥ BRED don’t have a break recording window and a refused in the importation.
- Pipes with BWED ≤ OPSD or BWSD ≥ OPED don’t have a break observation window and are excluded in the calibrations (LEYP 1).
- Pipes with DHS ≤ PPSD or DDP ≥ PPED don’t have a break prediction window. Pipes with DHS ≤ PPSD or DDP ≥ PPSD are excluded from the prediction calculations (LEYP 2). *Note: Pipes with PPSD < DDP < PPED can have a prediction window but are still excluded from the calculations.*

Dates are at the day precision. When imported dates are expressed in years, they are converted according to the following rules:

- DDP, 1st January of the year
- DHS, 31st December of the year
- DDC, 1st July of the year

Ages are expressed in decimal years calculated by dividing the number of days by 365.25.

XIII.3.ii Principal rules of dates:

Rules of the existence of periods:

- DDP < DHS
- BRSD < BRED
- OPSD < OPED
- PPSD < PPED

Rules of recorded breaks:

- BWSD ≤ DDC ≤ BWED
- DDP < DDC (t > 0)

Rule of succession of observation and prediction periods:

- OPED<PPSD

Rules of consistency of observation and recording periods

- OPSD≥ESD
- OPED≤EED

Rules of the existence of windows (see above):

- BRSD < DHS and DDP < BRED
- OPSD < BWED and BWSD < OPED
- PPSD < DHS and DDP < PPED expanded to DDP < PPSD

Rules of consistency of recording and validation periods:

- ESD < VPSD
- VPED ≤ EED

XIII.3.iii Duration of windows

The duration of a window in days is equal to the difference of the dates + 1:

Dfx = (EDFx – SDFx) +1

LEYP model for recurrent failures of water mains

Cemagref, Research Unit "Networks, Purification and Water Quality"
50 avenue de Verdun, 33612 Cestas Cedex (France)

May 2010

Abstract

This document aims at presenting the theoretical bases of the Linear Extension of the Yule Process (LEYP) designed to model recurrent failures of water mains. The LEYP model is implemented in the computation code of the "Casses" software.

1 Introduction

The Linear Extension of the Yule Process (LEYP) allows to give a parametric representation of the process of recurrent failures a pressure main is likely to undergo. The LEYP model enables to compute the possible number of failures that may affect a main with known characteristics within any time interval, even in the future. The water mains that make up a water supply network can consequently be ranked according to their failure risk, hence allowing to build annual pipe renewal programmes, and to compare in the medium and long term asset management strategies.

2 The counting process $N(t)$

Repeated failures may affect a single main at random times T_j, $j \in \mathbb{N}^*$. Their cumulated number define the so-called "counting" random function, also known as the "counting process" denoted $N(t)$, namely a right-continuous left-bounded "step" function, defined for any $t \geqslant 0$ and incremented by one unit at each failure occurrence:

$$N(0) = 0$$
$$N(T_j-) = j - 1, N(T_j) = j$$

Equivalently, the differential $\mathrm{d}N(t) = N(t + \mathrm{d}t) - N(t)$ takes the value 0 everywhere except at failure times where it takes the value 1:

$$\forall t \in \mathbb{R}_+ - \{T_j, j \in \mathbb{N}^*\} : \quad \mathrm{d}N(t) = 0$$
$$\forall t \in \{T_j, j \in \mathbb{N}^*\} : \quad \mathrm{d}N(t) = 1$$

Fig. 1 illustrates the construct of the counting process $N(t)$ and of its differential $\mathrm{d}N(t)$.

1

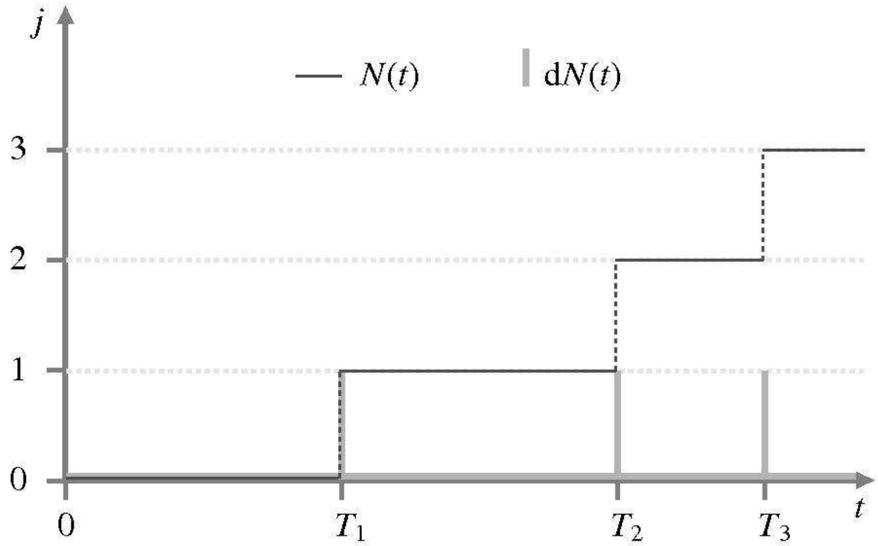


Figure 1: Counting process

3 The LEYP intensity function

The average failure rate within the time interval $[t, t+h]$ of length $h > 0$ can be defined as $\mathrm{E}\left(N(t+h) - N(t)\right)/h$, *i.e.* the average failure number within that interval divided by the interval length. When divided by the pipe length, this rate is commonly used in the field of infrastructure engineering ("failure rate"). The limiting value of the average failure rate when $h \to 0+$, $\mathrm{E}\mathrm{d}N(t)$, is a pivotal object of the counting process theory, called "process intensity".

In the literature dedicated to the reliability of pressure water mains, all reported studies seem to agree that the failure rate:

- tends to increase with the number of failures already undergone;
- tends to increase with the age of the mains;
- can significantly vary according to characteristics of the mains (such as material, size, joint type, internal or external coating, cathodic protection, *etc.*) and of their environment (soil corrosivity, traffic intensity, service pressure, *etc.*)

These technical evidences have lead to define the LEYP intensity function with parameter $\boldsymbol{\theta}^{\mathrm{T}} = (\alpha, \delta, \boldsymbol{\beta}^{\mathrm{T}})$, related to a main the characteristics of which compose the vector $\boldsymbol{Z}$ (vector of covariates), as the conditional expectancy:

$$\mathrm{E}_{\theta}\left(\mathrm{d}N(t) \mid N(t-), \boldsymbol{Z}\right) = \left(1 + \alpha N(t-)\right) \delta t^{\delta-1} \exp\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}\right) \mathrm{d}t$$

The LEYP intensity is then built as the product of three factors:

- the so-called Yule factor, $(1 + \alpha N(t-))$, linear function of the number of failures undergone until just before t; the scalar parameter $\alpha > 0$ measures thus the tendency of the failures to accumulate on the the same mains;
- time power factor, $\delta t^{\delta-1}$ with $\delta \geqslant 1$, which models ageing;
- the so-called Cox factor, $\exp(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta})$, making the LEYP model belong to the class of Proportional Hazard Models, parameterized by the regression coefficient vector $\boldsymbol{\beta}$.

4 Counting process distribution

The negative binomial distribution of the counting process constitutes a pivotal property of the LEYP model:

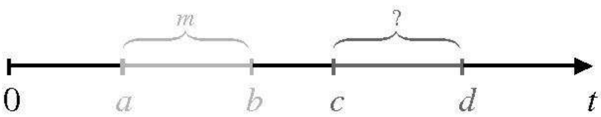
$$\mathrm{P}_{\theta}\{N(t) = m \mid \boldsymbol{Z}\} = \frac{\Gamma(\alpha^{-1} + m)}{\Gamma(\alpha^{-1})\, m!} \exp\{-\Lambda(t; \boldsymbol{Z})\} (1 - \exp\{-\alpha\Lambda(t; \boldsymbol{Z})\})^{m}$$

$$\text{with :} \quad \Lambda(t; \boldsymbol{Z}) = \int_0^t \delta u^{\delta-1} \exp(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}) \mathrm{d}u = t^{\delta} \exp(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta})$$

This explicit distribution property makes the counting process expectancy easy and quick to compute:

$$\mathrm{E}_{\theta} N(t) = \frac{\exp\{\alpha\Lambda(t; \boldsymbol{Z})\} - 1}{\alpha}$$

This result can without difficulty be extended to the computation, interesting from a practical point of view, of the number of failures likely to occur in a prediction interval $[c, d]$, given the number of failures within the observation interval $[a, b]$:



The conditional distribution of the process $N(d) - N(c)$ given $N(b) - N(a) = m$ is negative binomial:

$$[N(d) - N(c) \mid N(b) - N(a) = m, \boldsymbol{Z}] \sim \mathcal{NB}\left(\alpha^{-1} + m, \frac{\mu(b; \boldsymbol{Z}) - \mu(a; \boldsymbol{Z}) + 1}{\mu(d; \boldsymbol{Z}) - \mu(c; \boldsymbol{Z}) + \mu(b; \boldsymbol{Z}) - \mu(a; \boldsymbol{Z}) + 1}\right)$$

$$\text{avec :} \quad \mu(t; \boldsymbol{Z}) = \exp\{\alpha\Lambda(t; \boldsymbol{Z})\}$$

5 Parameter estimation procedure

The Casses software expects two input datasets:

- one related to the mains,
- the other related to failures.

The mains are n in number and indexed by $i = 1, \ldots, n$. The installation date of each main must be documented as well as the abandoned date (left empty if the main is still in service), and are used in connection with the beginning and stopping observation dates of the network to calculate the ages a_i and b_i between which the main was observed. The main description is required to include the length, and optionnally other important characteristics potentially explanatory of the failure rate, such as the material, the diameter, and also depending on their availability the soil corrosivity, the service presure, the location under roadway versus sidewalk, the traffic intensity *etc.* The characteristics kept as failure risk factors make up the covariate vector $\mathbf{Z}_i$.

The failure dataset lists for every main i which was observed to fail at least once, the event times $t_{ij} \in [a_i, b_i]$, $j = 1, \ldots, m_{ij}$.

The information available to estimate the model parameters is so formalised as the following set O:

$$O = \left\{ (\mathbf{Z}_i, a_i, b_i, \{t_{ij}, j = 1, \ldots, m_{ij}\}), i = 1, \ldots, n \right\}$$

The natural logarithm of the likelihood function of the parameter vector $\boldsymbol{\theta}$ given the observation set O is then written as follows:

$$\begin{aligned} \ln L(\boldsymbol{\theta}; O) = \sum_{i=1}^{n} \Big\{ & m_i \ln\alpha + \ln\Gamma(\alpha^{-1} + m_i) - \ln\Gamma(\alpha^{-1}) \\ & - (\alpha^{-1} + m_i) \ln\left(\mu(b_i; \mathbf{Z}_i) - \mu(a_i; \mathbf{Z}_i) + 1\right) \\ & + \sum_{j=1}^{m_i} \left\{ \ln\lambda(t_{ij}; \mathbf{Z}_i) + \alpha\Lambda(t_{ij}; \mathbf{Z}_i) \right\} \Big\} \\ \text{with : } \quad & \lambda(t; \mathbf{Z}) = \delta t^{\delta - 1} \exp\left(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\beta}\right) \end{aligned}$$

The LEYP model parameters are estimated by the vector $\hat{\boldsymbol{\theta}}$ that maximises $\ln L(\boldsymbol{\theta}; O)$.

6 Consideration of main abandonment and selective survival bias

Ensuring the practical relevance of the LEYP model when applied to a set of mains that comprises a notable proportion of very old pipes requires considering main abandonments (*i.e.* most often replacements). The study of such datasets frequently shows indeed that the oldest pipes oddly undergo very few failures. This seems to be due to selection: mains installed a long time ago, but having undergone repeated failures, are likely to have been abandoned for that reason before observation starts. Observation is consequently subjet to the so-called "selective survival bias".

This leads to introduce the random service time T and the function $\varsigma(t) \in [0, 1]$ which gives the probability that the main be repared following a failure undergone at age t, whereas the probability that the main be abandoned following this failure is $1 - \varsigma(t)$.

5

Assuming $\varsigma(t) = 1$ for any $t > b$ (*i.e.* beyond the observation window), the conditional distribution of the number of failures within the prediction window remains negative binomiale:

$$[N(d) - N(c) \mid N(b) - N(a) = m, T \geqslant b]$$
$$\sim \mathcal{NB}\left(\alpha^{-1} + m, \frac{\mu(b) - \int_0^a \varsigma(u)\mathrm{d}\mu(u)}{\mu(d) - \mu(c) + \mu(b) - \int_0^a \varsigma(u)\mathrm{d}\mu(u)}\right)$$

The Casses software uses the model:

$$\varsigma(t) = \exp(-\exp(\zeta_0 + \zeta_1 t)), \zeta_0 \in \mathbb{R}, \zeta_1 \in \mathbb{R}_+$$

where the probability that the main be kept in service following a failure decreases with the age of the main ($\zeta_1 \geqslant 0$).

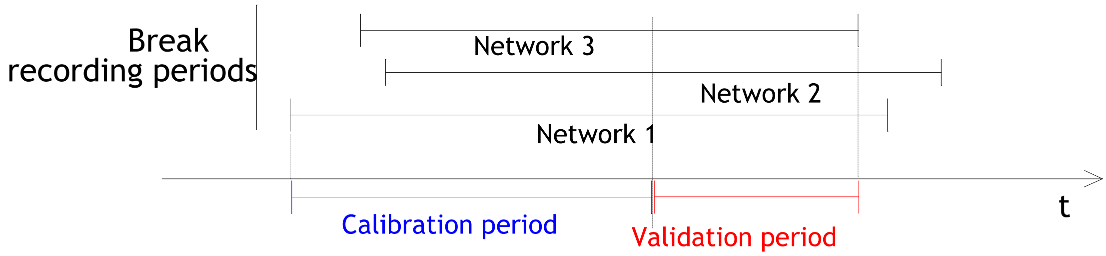
XIII.5 Consistency of the validation

XIII.5.i Principle of the validation

The basic principle of the validation is to compare the break predictions with the actual breaks for a period when breaks were observed.

To perform the validation, two distinct periods are defined from the break recording period – a calibration period and a subsequent validation period.

In the case of multiple networks, the validation period is a period during which all the networks were subject to break observations.



The validation applies to a prediction for which the period of observation coincides with the calibration period and the prediction period coincides with the validation period.

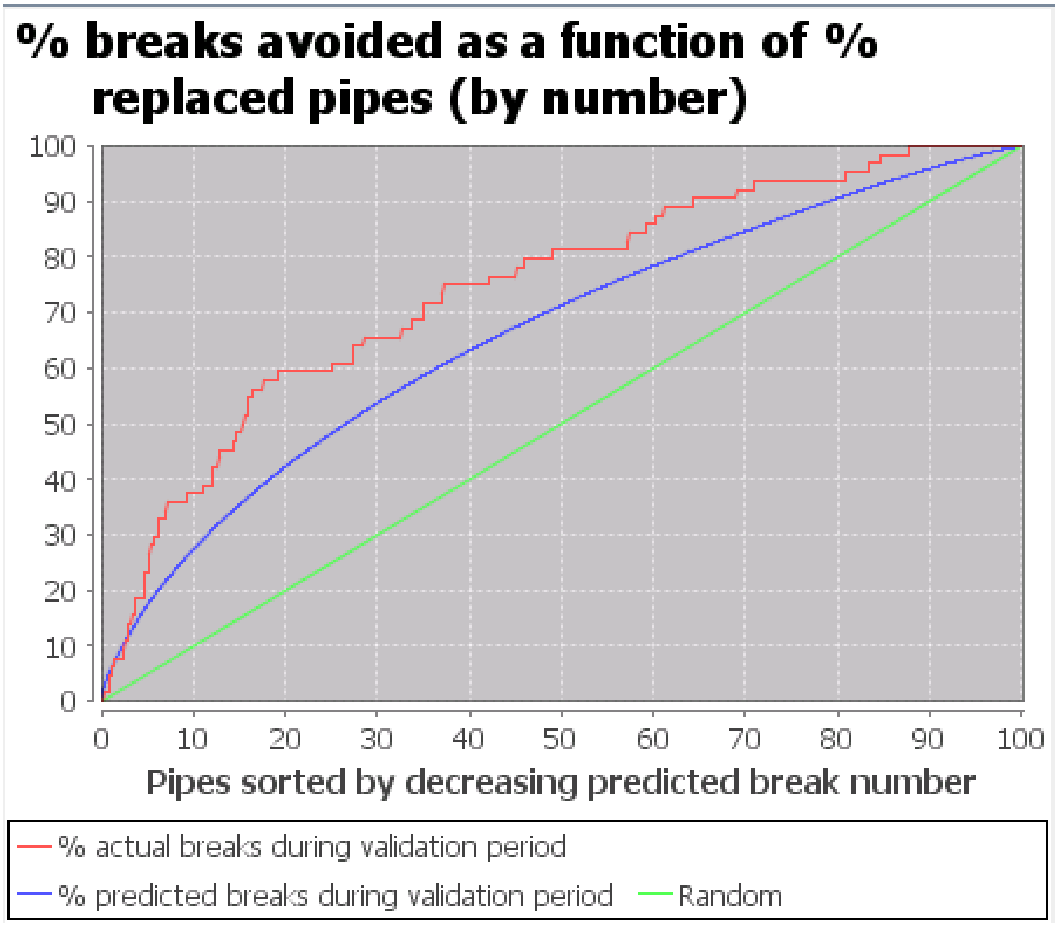
The calibration and validation periods are defined by the recording period of the Environment which is delimited by:

- ESD: Environment break record start date. It is the earliest recording start date of the networks with at least one pipe present in the Environment (min(BRSDi)
- EED: Environment break record end date. It is the minimum recording end date of the networks with at least one pipe present in the Environment (min(BREDi)

In the case of a mono-network Project, for all environments ESD coincides with BRSD and EED coincides with BRED.

XIII.5.ii Ranking according to the number of predicted breaks

After having sorted the pipes by descending **number of predicted breaks per year**, the proportion of the number of actual breaks during the validation period can be expressed as a function of the **number of pipes**.



X-axis: % of number of pipes
Y-axis, red curve: % actual breaks during the validation period
Y-axis, blue curve: % predicted breaks during the validation period

A random ranking of pipes corresponds closely to that described by the function y = x (green curve).

Two indicators are defined:

- An: Area under the red curve.
- C5n: Percentage of actual breaks during the validation period on 5% of the number of pipes sorted by descending number of predicted breaks.

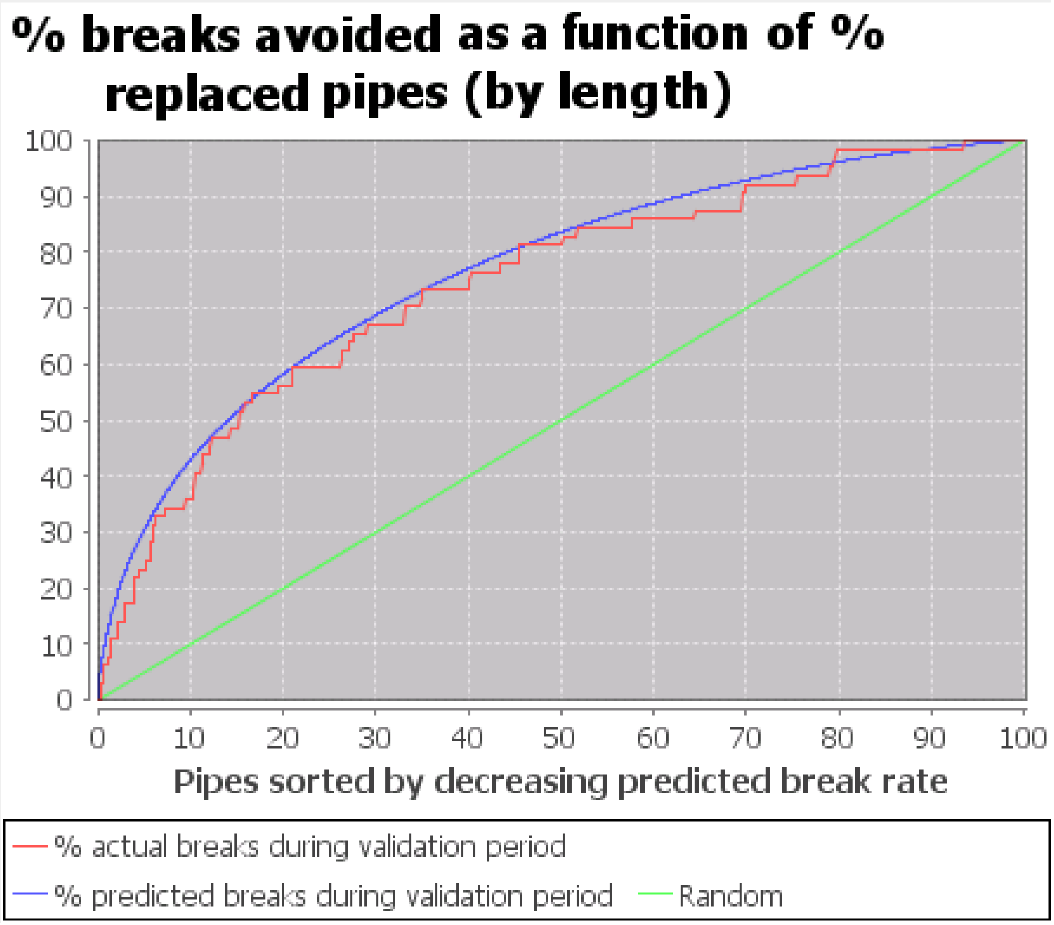
More generally, Cxn is the percentage of actual breaks during the validation period on x% of the number of pipes sorted by descending number of predicted breaks. Typical values are: 0.1 / 0.5 / 1 / 5.

For a random ranking, An is close to 0.5 and C5n is 5%. The prediction is therefore more satisfying when An and C5n are greater. In all cases, An and C5n are less than 1 (100%).

If a significant proportion of long pipes make up the pipes most at risk then this might lead to an optimistic vision of the model quality. 5 % of the number of pipes could, for example, represent 15% of the network length. For this reason an alternative ranking method is proposed in complement.

XIII.5.iii Ranking according to predicted break rate

After having sorted the pipes by descending **predicted break rate**, the proportion of the number of actual breaks during the validation period can be expressed as a function of the **cumulative length of pipes**.



X-axis: % of length of pipes
Y-axis, red curve: % actual breaks during the validation period
Y-axis, blue curve: % predicted breaks during the validation period

A random ranking of pipes corresponds closely to that described by the function y = x (green curve).

Two new indicators are defined:

- AI: Area under the red curve.
- C5I: Percentage of actual breaks during the validation period on 5% of the total length of pipes sorted by descending predicted break rate.

More generally, CxI is the percentage of actual breaks during the validation period on x% of the total length of pipes sorted by descending predicted break rate. Typical values are: 0.1 / 0.5 / 1 / 5.

XIII.5.iv Number de breaks predicted

To measure the validity of s prediction in terms of the number of breaks predicted, the following indicators are defined:

- PBN: Total number of predicted breaks for all pipes.
- ABN: Total number of actual breaks for all pipes.
- Rn: Ratio between the number of predicted and actual breaks for all pipes during the validation period.
- Rxn: Ratio between the number of predicted and actual breaks for x% of the pipes sorted by descending number of breaks predicted during the validation period.
- Rxl: Ratio between the number of predicted and actual breaks for x% of the cumulated length of pipes sorted by descending predicted break rate during the validation period.

The prediction is therefore more satisfying when Rn is close to 1. The same goes for Rxn and Rxl.

XIII.6 Computing information

1. Software installed:

- CASSES (software)
- ITHEA Actikey (driver for the protection key and management software for the keys)

2. Types of files created:

- Casses files with KSP extension
- Tabular data export file with CSV (Comma-separated value) extension
- Information files, Lanceur.log, Rimp.txt, Rcal.txt
- fichier des préférences utilisateur : casses.ini

3. Relevant directories:

- For CASSES: C:\Program Files\Cemagref\CassesMono or C:\Program Files\Cemagref\CassesMulti according to the version bought
- For the protection key: C:\Program Files\ithea and C:\WINDOWS\system32
- For the information files: \\home\user$\Casses, created at installation
- For Sun JVM (Java Virtual Machine): C:\Program Files\Java\

4. Minimum configuration requirements:

- OS: Windows XP
- Memory: 512 Mb
- Hard disk space: 96 Mb
- Sun JVM (Java Virtual Machine): 89 Mb