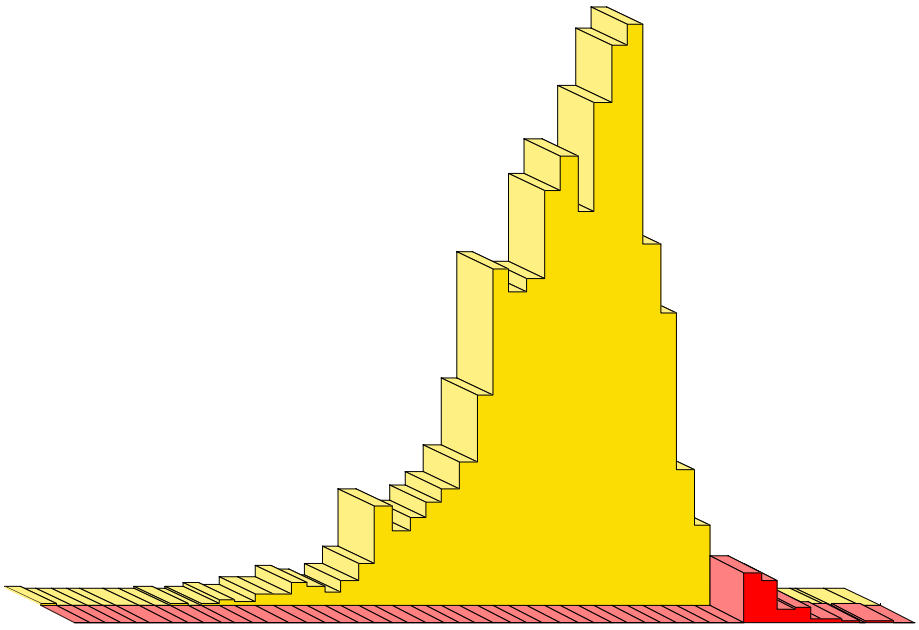
MCRA

a web-based program for Monte Carlo Risk Assessment

Release 3

User Manual

Waldo J. de Boer
Hilko van der Voet
Polly E. Boon
Gerda van Donkersgoed
Jacob D. van Klaveren



Biometris and RIKILT
Wageningen University and Research centre

March 2004

Biometris is the integration of the Centre for Biometry of Plant Research International and the Department of Mathematical and Statistical Methods of Wageningen University and is a part of Wageningen University & Research centre.

Post address
P.O. Box 100
6700 AC Wageningen
The Netherlands
Visiting address
Bornsesteeg 47, building no. 116
6708 PD Wageningen
Telephone: +31 (0)317 476925
Telefax: +31 (0)317 483554

RIKILT Institute of Food Safety is a part of Wageningen University & Research centre.

Post address
P.O. Box 230
6700 AE Wageningen
The Netherlands
Visiting address
Bornsesteeg 45, building no. 123
6708 PD Wageningen
Telephone: 0031 317 475400
Telefax: 0031 317 417717

Contents

Contents 3

1. Foreword 5

2. MCRA, an introduction 6

3. The MCRA website 7

4. MCRA 3.1 Data Selection 9

5. Preparing the data 13

5. 1. Overview *13*

5. 2. Specification of tables *13*
5.2.1. Compounds 14
5.2.2. Individual 14
5.2.3. Foodsurvey 14
5.2.4. Foodconsumption 15
5.2.5. Foodconversion 15
5.2.6. Concentration 16
5.2.7. Country 16
5.2.8. Histogramdata 17
5.2.9. Summarydata 18
5.2.10. Processing 18
5.2.11. Processingfactor 19
5.2.12. Products 20
5.2.13. Productcategory 20
5.2.14. Productgroup 20
5.2.15. Product_subgroup 21
5.2.16. VariabilityProd 21
5.2.17. VariabilityCompProd 21
5.2.18. VariabilityProcCompProd 22
5.2.19. Agriculturaluse 22

6. Performing a risk assessment 24

6. 1. Overview *24*

6. 2. MCRA program options: Settings *24*

6. 3. MCRA program options: Exposure model *25*
6.3.1. Acute or chronic risk 25
IESTI estimates 25
Chronic risk assessment 25

6. 4. MCRA Program options: Residues *26*
6.4.1. Full, summary or histogram data 26
6.4.2. Nondetects and percent crop treated 26
6.4.3. Empirical or parametric modelling 27
6.4.3.1 Non-parametric approach 27
6.4.3.2 Parametric approach 27
6.4.4. Processing 28
6.4.5. Unit variability 29
6.4.5.1 Bernoulli distribution 29

6.4.5.2 Estimated parameters for unit variability 30
6.4.5.3 Beta model 30
6.4.5.3 Lognormal distribution using estimated parameters 31

6. 5. MCRA Program options: Simulate *31*
6.5.1. Number of simulations 31
6.5.2. Bootstrap 32
6.5.3. Pseudo-random sampling 32
6.5.4. Memory use. 32

6. 6. MCRA Program options: Output *32*

6. 7. CHECK PROGRESS: Checking on the progress of job execution *33*

6. 8. VIEW OUTPUT: Viewing the output of an exposure assessment *33*

6. 9. HELP on CHARTs *33*
6.9.1. To display the Property Editor 34
6.9.2. Interacting with Chart 34
To Scale the Chart: 34
To Move the Chart: 34
To Graphics Zoom an Area of the Chart: 34
To Axis Zoom the Chart: 34
To Rotate the Chart (Bar/pie charts displaying 3D effect only): 34
To Reset to Automatic Scale and Position: 34

7. Example output **35**

7.1.1. Distribution of exposure 35
7.1.2. Percentiles 36
7.1.3. Exposure contributions 38

Appendix 1. Example of pooling commodities in parametric modelling **39**

8. References **40**

1. Foreword

This User Manual is intended to assist with the practical application of MCRA. MCRA is a computational tool for dietary exposure assessment of chemical substances based on measurements concerning the quality of agricultural products. Exposure assessment is an important step in the risk assessment of chemicals such as agricultural chemicals (pesticides, veterinary drugs), toxins (e.g. mycotoxins) and environmental contaminants (e.g. dioxins).

The MCRA system (Monte Carlo Risk Assessment) implements probabilistic modelling. Basically, it simulates daily consumptions by sampling a food consumption database, and combines these with a random sample from either a concentration database (empirical distribution) or a parametric distribution of compound concentrations. The result is a full *distribution* of intakes, rather than traditional deterministic methods which only provide a point estimate. Percentiles from the intake distribution can be used to assess risks by relating them to e.g. an acute reference dose (ARfD).

The current version of MCRA can calculate intake distributions for both short-term (acute) and long-term (chronic) exposures. Uncertainty of percentiles can be established by bootstrapping. It allows to include processing factors (e.g. the effect of cooking on the residue level) and variability factors (to correct for the fact that monitoring data are obtained from composite samples, whereas consumers may eat individual units). Analyses can be done for a total population or for a subpopulation (e.g. children, or consumption-days only). The effects of residue levels below analytical reporting limits can be assessed. Large portion consumption and the highest residue or median residue in case of bulking or blending in the composite sample are used in IESTI calculations.

The current MCRA system is Internet-based, and can be used by registered users at http://www2.rikilt.dlo.nl/mcra/mcra.html. It consists of a basic program to do the computations, written in the statistical package GenStat (2002), and of additional database selection possibilities implemented in HTML and Active Server Pages (ASP). MCRA runs with Component One Chart (1999) which offers the possibility to manipulate graphical output after it has been obtained. An earlier version of the GenStat MCRA program, as well as an implementation of the Monte Carlo method in @Risk (1996), have been described in van der Voet *et al*. (1999), and further elaboration was given in de Boer & van der Voet (2000, 2001) and van der Voet *et al*. (2001).

This documentation describes MCRA Release 3. It covers the current release 3.1 and all future updates starting with the same release number. So, for example, release 3.2 will to the second update of Release 3. For future updates additional documentation supplementing this manual will be placed on the website. Major updates of the program, encompassing new or improved facilities will be released with an increased Release number and a new manual.

MCRA is a result of an ongoing co-operation between RIKILT and Biometris since 1998. RIKILT co-ordinates the Dutch KAP programme (Quality of Agricultural Products) where results of monitoring programs for chemical residues in food are gathered in a national database. RIKILT also has a recipe database to link food codes from the Dutch food consumption table to primary agricultural products. Biometris contributes statistical models and programs for quantitative risk analysis.

For a complete description and justification of the statistical methods in MCRA the reader is referred to the MCRA Release 3 Reference Guide (van der Voet et al. 2004).

2. MCRA, an introduction

This chapter first describes what you can expect from the use of MCRA. Secondly, it gives an introduction to the data that you need to have yourself in order to profit from the use of MCRA

The MCRA system (Monte Carlo Risk Assessment) can be used for assessment of risks due to the intake of residues on food products. MCRA provides the following options:

- Acute probabilistic risk assessment: MCRA will calculate the exposure distribution (mg residue per kg body weight) from input data on consumption and residue concentrations in the food.
- Percentiles: The exposure distribution can be characterised by percentiles, *i.e.* residue concentration levels exceeded with only a small specified probability (for example the 99th percentile p99 is exceeded only in 1 % of the cases).
- Uncertainty due to small samples: Bootstrap sampling of consumers and residue concentrations to assess the uncertainty of the percentiles in the form of an approximate confidence interval.
- Calculation of point estimates (IESTI), and comparison with Monte Carlo results.
- Conversion of food consumption to the consumption of primary agricultural commodities, e.g. convert pizza consumption to consumption of wheat, tomato, cheese, etc.
- Parametric or empirical modelling of residue levels: MCRA can resample the residue concentration data directly (empirical model), or it can sample from a binomial-lognormal model fitted to the concentration data (parametric model). Note: consumption data are always resampled empirically from the consumer data set.
- Modelling of processing effects: Sometimes it is known that residue levels are reduced by food processing, e.g. cooking. MCRA can incorporate processing factors as fixed effects or by sampling from a processing factor distribution. The latter possibility requires the specification of a nominal and an upper value for the processing factor.
- Modelling of unit variability: Residue concentrations are often measured in large composite samples, thus hiding part of the variability that exists between individual units. MCRA has extensive possibilities to model unit variability.
- Modelling of nondetects levels: Residue concentrations are often only known above a certain limit, the Limit Of Reporting (LOR). In a worst-case analysis all non-detect measurements may be replaced by the LOR value.
- Subset selection: Extensive possibilities to select the data on age, weight or sex of consumers, day of consumption, foods, primary products, and year, country and sampling type of concentration data.
- Calculate exposure distribution for consumption-days only
- Chronic risk assessment: Chronic risk assessment requires consumption data for > 1 day per individual. MCRA calculates the between-individual distribution of chronic exposure, following the method of Nusser *et al.* (1996). The chronic exposure distribution is characterised by percentiles and bootstrap confidence intervals on these percentiles.

Which data do you need to have in order to run MCRA simulations?
The MCRA system has a central database which offers a selection of example databases. So it is possible to run example analyses without having data yourselves. However, MCRA is primarily designed to work with user databases, or with a mix of user data and centrally available data. For example, you can provide your own data on residue concentrations, and combine these with the consumption survey data in the central database to produce an risk assessment for this residue.

All data for MCRA are stored in Microsoft Access database tables according to fixed formats. So, if you want to use own data, you will need to prepare Ms Access database files off-line, and then upload them to your personal user area on the MCRA website. As a registered MCRA user you have complete control over the file management in your personal area.

3. The MCRA website

To use MCRA navigate your web browser to http://www2.rikilt.dlo.nl/mcra/mcra.html . The opening screen gives some general information, as well as links to the latest versions of the User Manual (this document) and the Reference Guide.

As a potential new user, you will first have to fill in a registration form. Here you have to specify your name, organisation, address and email address, and you can choose a user name (no spaces allowed) and a password for the use of MCRA system. Clicking the **OK** button will send the request to the MCRA webmaster at RIKILT, and you will get a response by email as soon as possible.

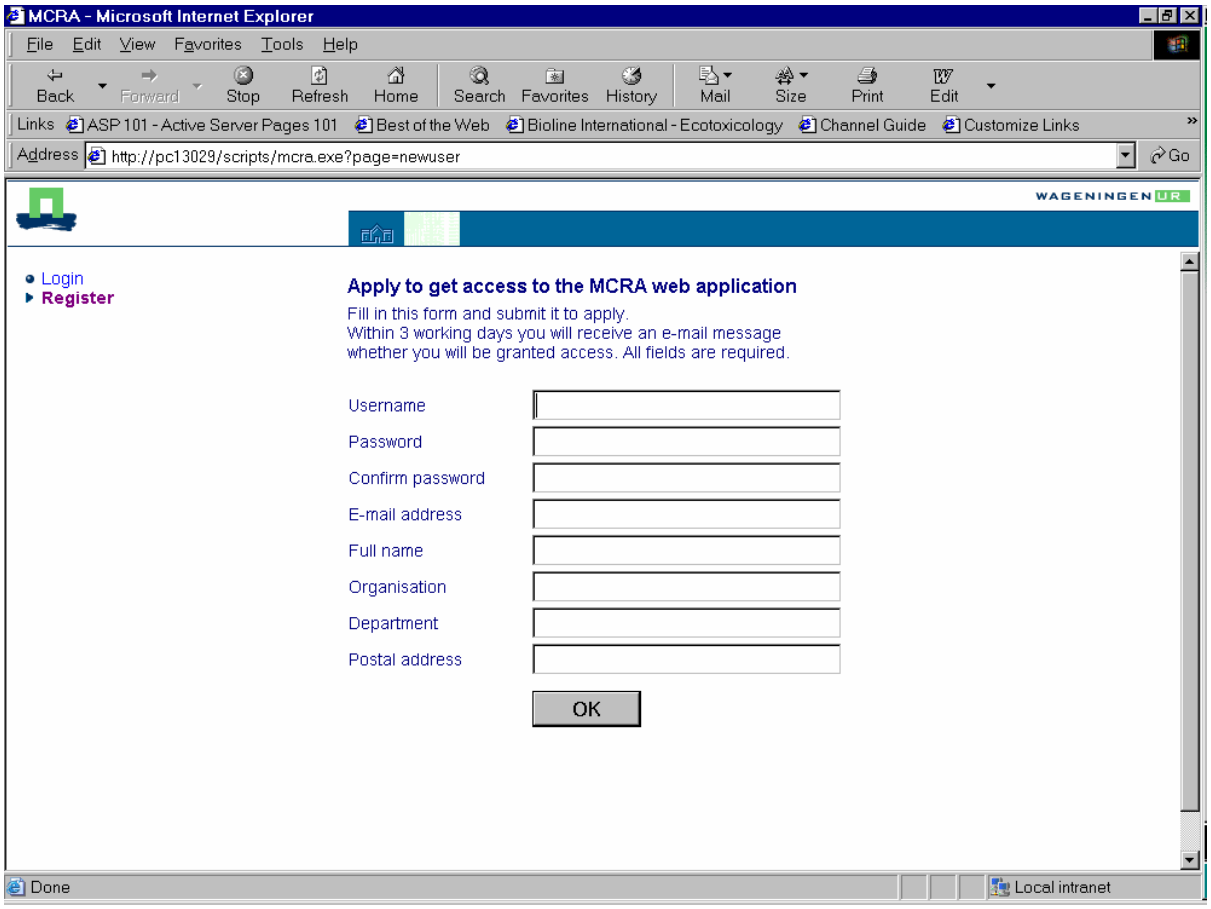


Figure 1: Registration form

Registered users can login with their username and password. The first screen after login is the **MCRA Main Menu**.

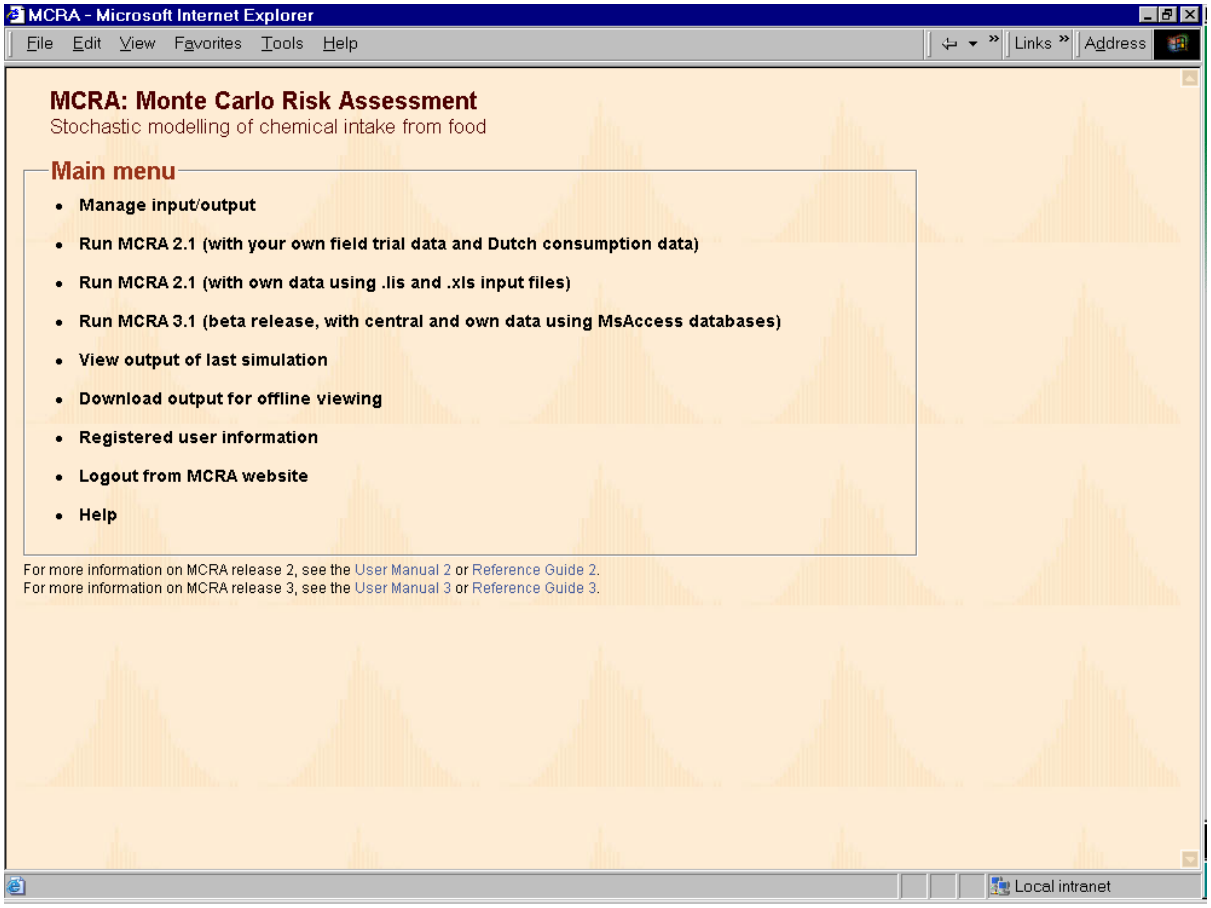


Figure 2: MCRA Main Menu

From here you can do the following:

1. If you want to use your own data, choose **Manage input/output**. This brings you to a screen where you can upload your data files. Each user has a personal data area, with two subdirectories named IN and OUT. The IN directory can be used to upload your own Access database files. See Chapter 5 for the required data formats. Files can be uploaded directly or in zipped form. You can also zip, rename or delete files. Acces database files and zip files can also be downloaded. Output from MCRA will be written to your personal OUT directory.
2. If you want to run MCRA, choose one of the **Run MCRA** options. It may be true that different versions are available, which can be older versions of the program or specialised versions for specific groups of users. This manual describes the general MCRA 3.1. system. For a description of the Run MCRA 3.1 menu see Chapter 4. For a description of the input options see Chapter 6.
3. If you want to look at the results of your last MCRA run, choose **View output of last simulation**. This brings you to a screen where both tabular and graphical results can be viewed. See Chapter 7 for examples.
4. It is also possible to **Download output for offline viewing**. Output files are downloaded in a zipped format. The download includes a file *viewoutput.htm* which gives you the same options to study your output as available on the website.
5. You can view which information about you is stored in the user database by clicking **Registered user information**.
6. You can **Logout from the MCRA website**. Your personal data files, latest output files and the latest input options remain stored for later use.

4. MCRA 3.1 Data Selection

The MCRA 3.1 system is essentially a cluster of activities build around a central menu. Each activity can be started from the central menu window and after finishing the activity the user returns to the central menu. New users are automatically brought to the Selection of databases screen. In all subsequent cases the MCRA 3.1 central menu shows the latest selection of database, consumption survey and compound.

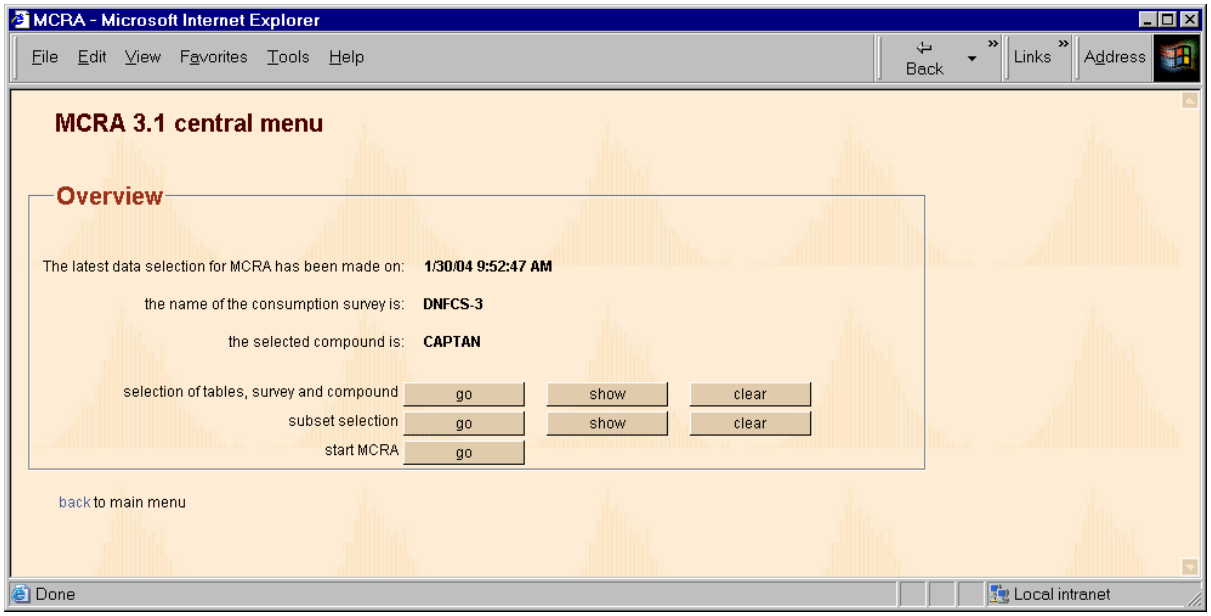


Figure 3: MCRA 3.1 central menu

The possible choices from the Central Menus are:

1. **Selection of databases, tables, compound and survey**.
   You can select tables from a central database or supply your own data (see Figure 4). You can select whole databases or select tables from different databases making combinations from centrally available data or user data.
   If there are consumption data from multiple surveys in the selected Food consumption survey table, you are required to choose one survey. If there are more compounds in the selected Compounds table, you are required to select one compound.
2. **Subset selection**.
   The data are stored in a relational database. The interface menu for subset selection utilises the relationships between tables, offering a flexible tool to select data according to your own wishes. The user first defines the population of consumers through the use of scroll-down menu's for age, weight and sex. Then, selection of day(s) of consumption, foods, primary agricultural products and year, country and sampling type of concentration data follows, again using scroll-down menu's to define your selections. The hierarchical structure of the productcode may be used to select a group of primary agricultural products or foods at once.
3. **Show current subset selection**.
   This shows all selected levels for those variables on which selections are active. For variables that are not shown all available data from the selected database tables is used.
4. **Start MCRA**.
   See Chapter 6.

After selecting tables a check is made on the number of primary agricultural products involved. If there are products for which there are consumption data, but no concentration data, the user is confronted with a choice: either continue with only the products for which concentration measurements are available, or let the program check the Agriculturaluse table. In the latter case products for which no information is found, or for which agricultural use is indicated to be forbidden (coded as 0), are deleted from the analysis, whereas the remaining products (agricultural use allowed, but no concentration measurements available) will be included in the analysis using a user-defined worst-case value. These choices are postponed until, after subset selection, the Start MCRA button in the Central Menu is pressed.

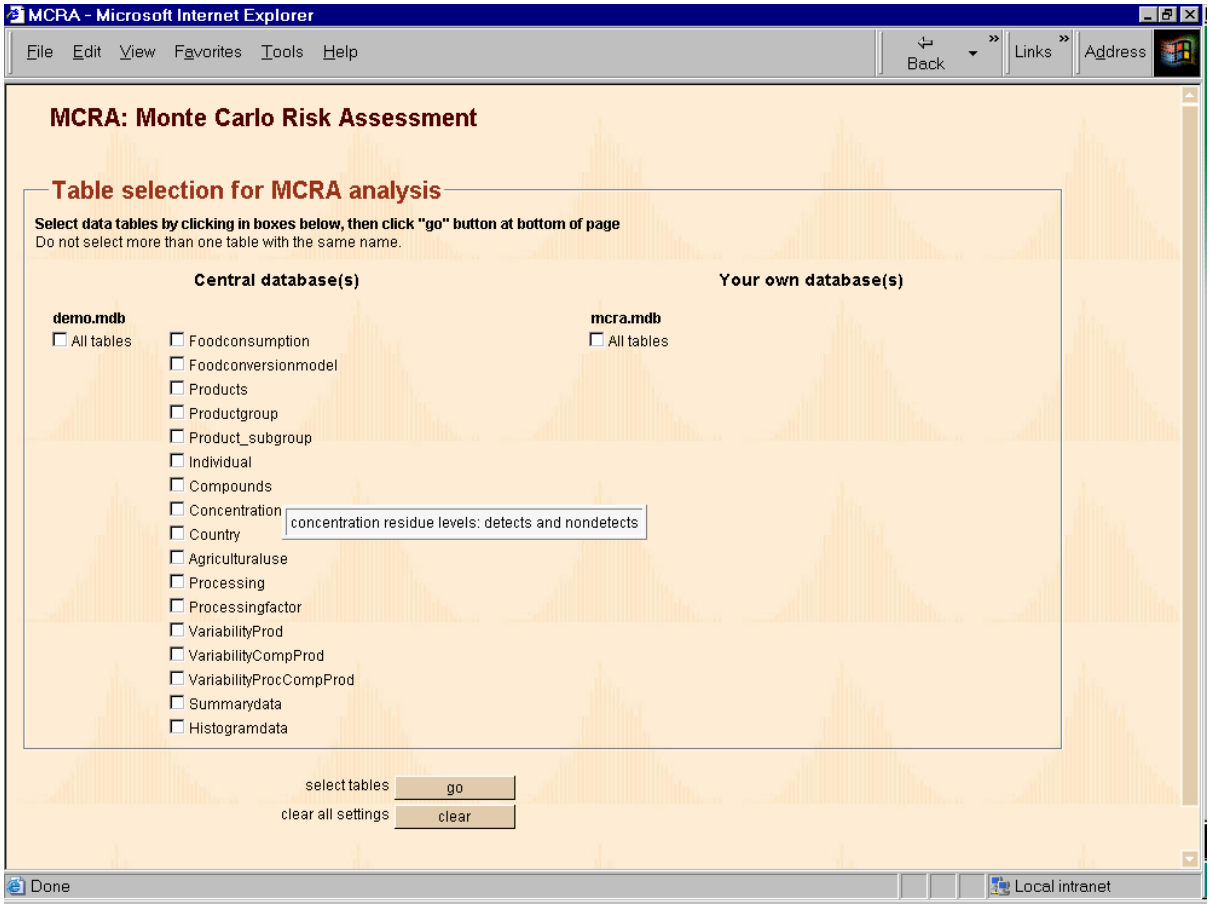


Figure 4: Table selection for MCRA analysis

The first subset selection screen (see Figure 5) allows you to select the consumer population based on characteristics of the individuals (age, weight, sex). For example to select children, check the box for variable age, and click the upper button. Then, a new screen appears where you can choose minimum and maximum ages (unit depending on which survey you use, see table Food consumption survey). After clicking the upper button, the system automatically adjusts the levels of the other variables. For example, for children 1-6 yers old, the maximum weight will be lower than the maximum of 150 kg for the whole database. To begin with the current selection is equal to all data in the selected database tables. If in a further selection step you want to include previously excluded levels, just click the radio button labeled “database” in the “select from:” column.

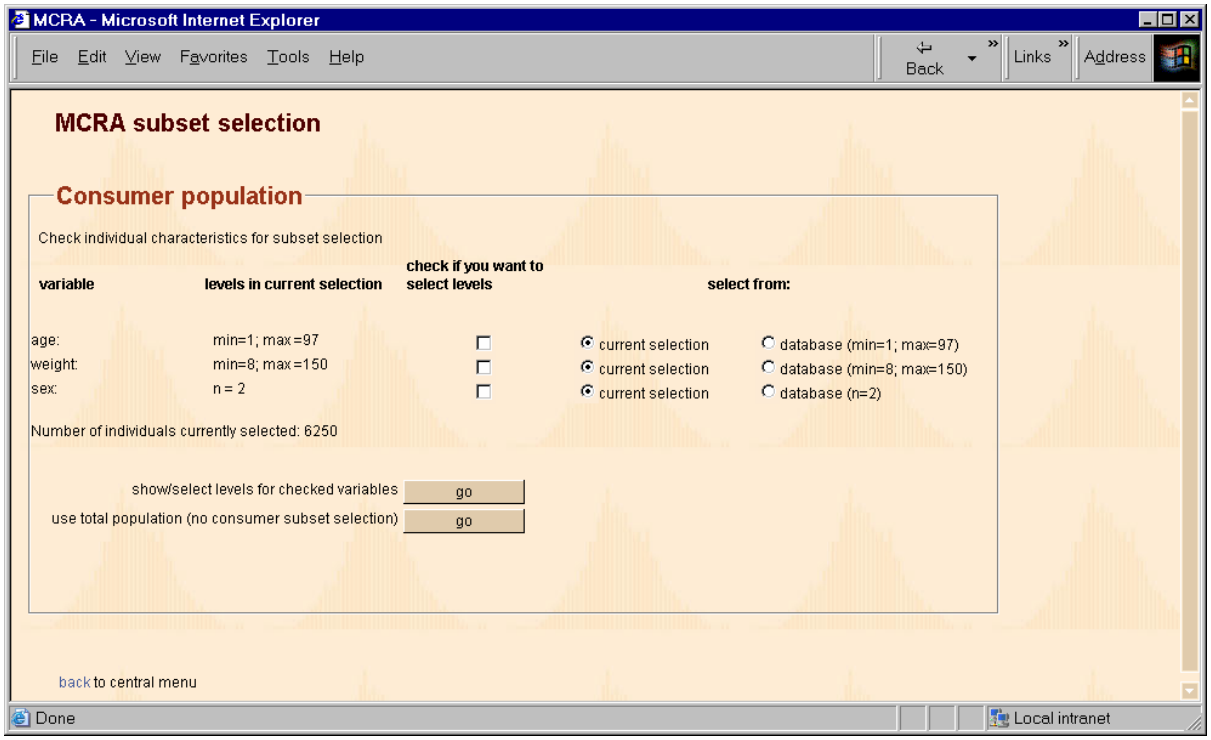


Figure 5: MCRA subset selection: Consumer population

The second subset selection screen allows you to select which consumption and concentration data should be used (see Figure 6). Variables on which a subset selection can be made are:

- food, food group: to restrict the analysis to specific foods or food groups;
- product, product subgroup and product group: to restict the analysis to specific primary agricultural products;
- (consumption) day: to restrict typical consumption survey data to specific days (e.g. only the first);
- year: to restrict the concentration data to specific years;
- samplingtype: to include only concentration data from a specific samplingtype;
- country: to include only concentration data from products originating in specific countries.

For example if you only want an analysis on Greek grapes (see Figure 7) check the boxes for the variables product and country, and click the upper button. Then, a new screen appears where you can choose grape in the first scroll-down menu and Greece in the third (assuming these exist in the database). After clicking the upper button, the system automatically adjusts the levels of the other variables. For example, all foods that contain no grape are removed, as you can check by checking the food box, and using the show levels button again.

To begin with the current selection is equal to all data in the selected database tables. If in a further selection step you want to include previously excluded levels, just click the radio button labeled “database” in the “select from:” column.

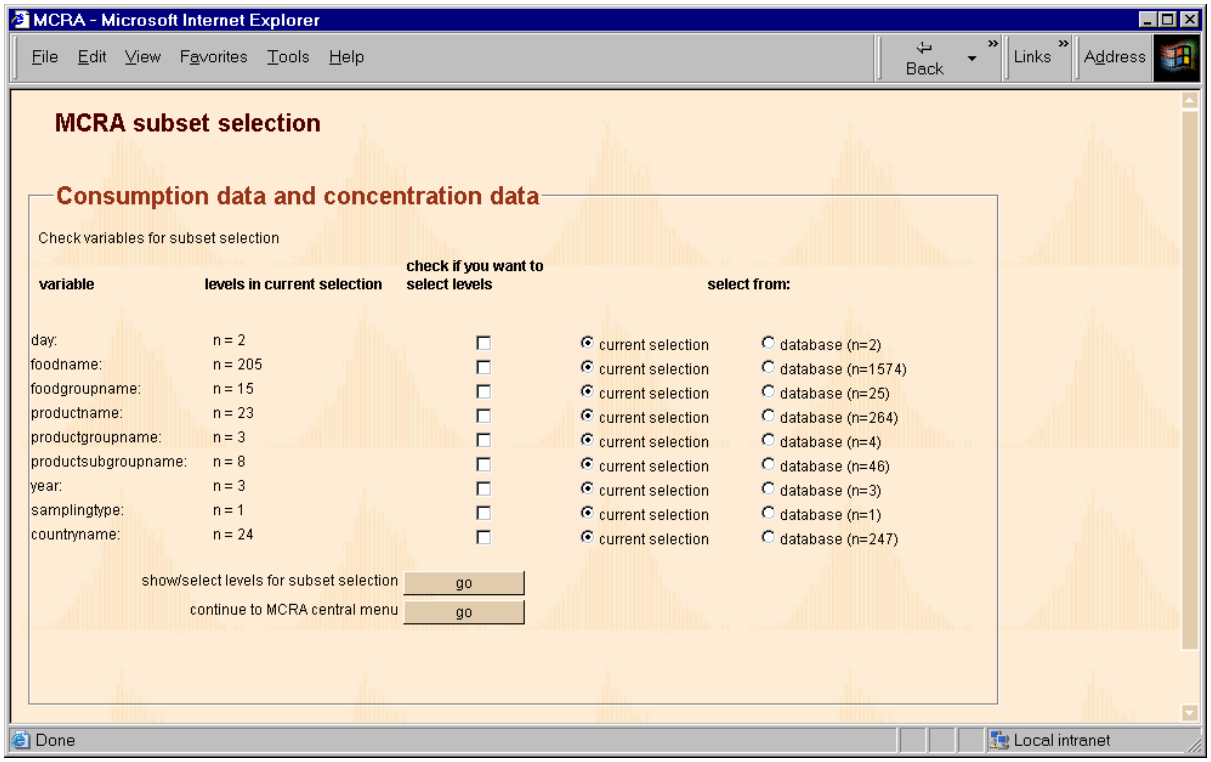


Figure 6: MCRA subset selection: Consumption data and concentrations

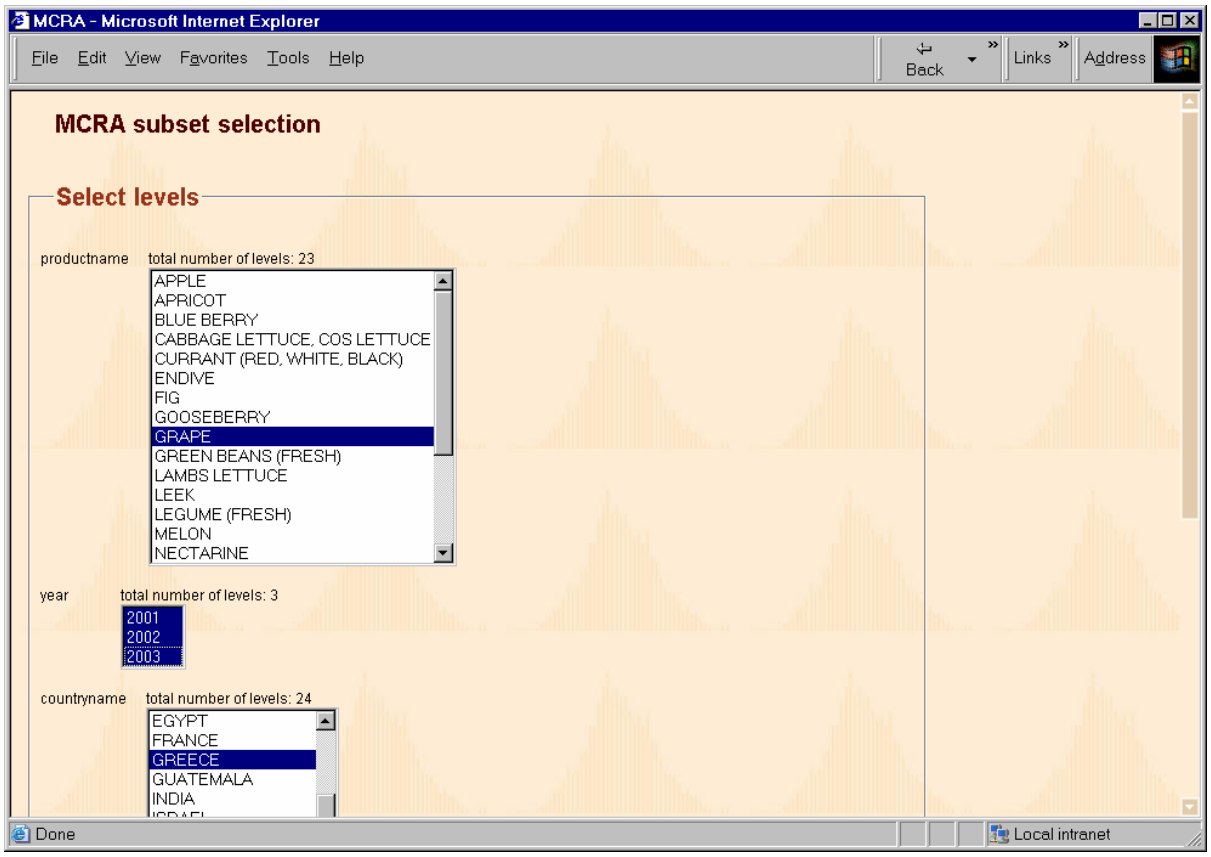


Figure 7: MCRA subset selection: Select levels

5. Preparing the data

5. 1.Overview

The internet version of MCRA requires that all information needed for an exposure assessment is stored in a database, residing in one or more Microsoft Access database files. The database organises data into a number of tables. See Table 1 for an overview. A table is a collection of data about a specific subject or topic. The data are organised into columns (called fields) and rows (called records). Select queries are used to retrieve data from one or more tables according to criteria that are specified by the user (see Chapter 4). The database structure offers flexibility in handling data and assessing risks.

Required tables	
Individual	consumer characteristics
Foodconsumption	consumption of foods
Foodconversionmodel	conversion factors from foods into primary agricultural products
Products	product codes, labels and product specific information
Compounds	compound codes, labels, agricultural and toxicological limits
Country	country codes, labels (concentration data)
one of:	
Concentration	compound concentration data (full data)
Summarydata	compound concentration data (as summary statistics)
Histogramdata	compound concentration data (table of frequency counts)
Additional tables (for querying or specific options)	
Processing	processing codes, labels
Processingfactor	processing factors
VariabilityProd	unit variability factors
VariabilityCompProd	unit variability factors, compound-specific
VariabilityProcCompProd	unit variability factors, processing- and compound-specific
Agriculturaluse	information on the agricultural use of compounds (e.g. use allowed, percent crop treated, maximum residue limits)
Productcategory	first level of product hierarchy, code and label
Productgroup	second level of product hierarchy, code and label
Product_subgroup	third level of product hierarchy, code and label

Table 1: Overview of tables.

To run a Monte Carlo Risk Assessment, tables called ‘required’ should always be supplied. The need to supply ‘additional’ tables depends on the model specification (see 6).

5. 2.Specification of tables

In the next paragraphs, the format of tables needed for a Monte Carlo Risk Assessment are described.
General remarks:

- Missing values are indicated with code 9999, unless stated otherwise. In general, an empty cell is also interpreted as a missing value. However, occasionally the use of empty cells leads to errors in retrieving data. Therefore, it is advised to use the code 9999 to indicate missing values
- Table and column (field) names should be exactly as indicated in the sections below.
- Product codes and compound codes are entered in multiple columns, corresponding to a hierarchical coding system.
- The order of columns should not be changed.
- The use of quotes (‘) in product, compound, country or surveynames is not allowed

In the next sections, the heading displays the table name, followed by block with field names and a short description of each field. A table should contain all fields except for tables Histogramdata and Summarydata. For each field, numbers are assumed unless stated otherwise (in parentheses). Each section ends with some notes and a short example of the content of a table.

5.2.1.Compounds

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
arfd	ARfD (acute reference dose), in µg/kg bw/day
adi	ADI (acceptable daily intake), in µg/kg bw/day
compoundname (text)	compound label (name of compound)

- Each compound is characterised by a code built hierarchically from 3 numbers.
- Missing values for arfd and adi: 9999.

Example:

```
11  5  1  0.02  60  9999  IPRODIONE
```

5.2.2.Individual

field name	description
individualid	consumer identification number
age	age (*e.g.* in years, months or days)
weight	body weight (*e.g.* in kg or g)
sex (text)	gender
foodsurvey (text)	name of survey

- In table Foodsurvey the unit for age and weight is defined.
- No missing values allowed.

Example:

```
 101  302  13  female  EU-BABY
4901   47  87    male  DNFCS-3
4904   43  67    male  DNFCS-3
5342   84  78  female  DNFCS-3
5674   98  71    male  DNFCS-3
```

5.2.3.Foodsurvey

field name	description
foodsurvey (text)	name of survey
year	year of survey
country (text)	country of survey
agein (text)	unit of age
weightin (text)	unit of weight

- Defines characteristics of the survey.
- No missing values allowed.

Example:

```
DNFCS-3   1997  NL  Y  kg
EU-BABY   1998  NL  D  kg
```

5.2.4.Foodconsumption

field name	description
individualid	consumer identification number
day	day (sequential number in food consumption survey)
foodcategory	food category number
foodgroup	food group number
foodsubgroup	food subgroup number
food	food number
foodq	food quality number
amountconsumed	consumed portion of food (g)
foodsurvey (text)	name of survey

- This table contains data on consumed foods. Days without consumptions are not recorded. The number of available days per consumer is inferred from this table and is assumed to be the same for each consumer in the survey.
- Each food is characterised by a food code built from 5 numbers (not hierarchically):
  1 – food category number, 1 digit: 1 = primary agricultural product, 9 = food as eaten
  2 – food group number, max. 2 digits
  3 – food subgroup number, max. 2 digits
  4 – food number, max. 4 digits
  5 – food quality number, max. 2 digits
- Residue concentration data are usually measured on primary agricultural products. This means that conversion from consumed food to primary agricultural products should be done before performing a MCRA analysis (see van Dooren et al. 1995). Conversion factors should be provided in table Foodconversion.
- No missing values allowed.

Example:

```
4901   1   9   20   0   401   0     49.910  DNFCS-3
4901   2   9   18   0   436   0      0.964  DNFCS-3
5673   2   9   19   0   443   0     62.000  DNFCS-3
5673   1   9   21   0   743   0    131.776  DNFCS-3
5675   1   9   20   0   133   0     92.000  DNFCS-3
```

5.2.5.Foodconversion

field name	description
foodcategory	food category number
foodgroup	food group number
foodsubgroup	food subgroup number
food	food number
foodq	food quality number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
percentage	percentage of primary agricultural product each food consists of
proccode	processing code

- For each food, the composition is given in percentages of primary agricultural products.
- Each product is characterised by a code built hierarchically from 5 numbers:
  1 – product category number, 1 digit: 1 = primary agricultural product, 9= food as eaten
  2 – product group number, max. 2 digits
  3 – product subgroup number, max. 2 digits
  4 – product number, max. 3 digits

5 – product quality number, max. 2 digits
European commodity codes are a concatenation of 2, 3 and 4.

- No missing values allowed

Example:

```
9  20  0  401  0 1  8  1  2 1  8.4  12
9  20  0  401  0 1  8  2  2 1 91.6   3
9  18  0  436  0 1  9  1  1 1 80.9   3
9  19  0  443  0 1  6  1  2 1 62.0   3
9  19  0  443  0 1  6  1  2 2  5.0   5
```

5.2.6.Concentration

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
year	sampling year
month	number of month
country (text)	country of sample
samplingtype (text)	type of sampling (monitoring)
numberofsamples	count of the number of times the specified concentration or limit of reporting (LOR) occurs
value	concentration (mg/kg) or LOR (see below)

- The limit of reporting (mg/kg) is specified in column value using a minus (-) sign to make the distinction between a measured residue levels, e.g. –0.02 (see example first row).
- Concentration values are stored in column value and the number of times each value occurs in column numberofsamples, e.g. 0.21 and 1, respectively.
- The number of nondetects for product 1 6 1 1 1 is 10, the total number of samples taken is equal to 11 (see example, first and second row).
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing, the lowest residue value found in the database. A warning is generated when 1) and 2) are not possible.
- No missing values allowed for the other columns.

Example:

```
11  5  1  1  6  1  1  1  1999  1   NL   M  10  -0.02
11  5  1  1  6  1  1  1  1999  1   NL   M   1   0.21
11  5  1  1  7  1  2  1  1999  12  NL   M   1   2.78
11  5  1  1  8  1  2  1  2000  8   NL   M   1   1.20
11  5  1  1  8  1  2  1  2000  9   NL   M   1   0.63
```

5.2.7. Country

field name	description
country (text)	code for country
countryname (text)	name of the country, label

- No missing values allowed

Example:

```
NL  THE NETHERLANDS
BL  BELGIUM AND LUXEMBOURG
CA  CANADA
CN  CHINA
DK  DENMARK
```

5.2.8.Histogramdata

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
limitofreporting	limit of reporting (mg/kg)
numberofsamples	size of sample (detects and nondetects)
c%02	number of samples with a concentration between the value
c%05	extracted from the field name of the previous column (exception:
c%1	for the first column a value 0 is taken) and the value extracted from
c%2	the field name in the current column (mg/kg).
c%5	
c1	
c2	
c4	
cE10	

- MCRA release 3 has an option to input frequency counts (histogram data) for residue concentrations. These data are then used for estimating parameters of a parametric model (see 6.4.3).
- Field names for columns representing the number of frequency counts are constructed as follows:
  c = indicates class limit,
  % = represents the decimal point (if necessary),
  xx = is the value of the class limit.
  Thus: field name c%02 specifies class limit 0.02, field name c4 specifies class limit 4, field name cE10 specifies class limit $1*10^{10}$.
- The number of nondetects measurements is given as the difference between the numberofsamples and the sum of frequency counts, e.g. see example first record 377 – 1 = 376,.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing….. A warning is generated when 1) and 2) are not possible.
- For columns numberofsamples, c%02…cE10 no missing values are allowed. When no data are available for a product, delete the entire row. Classes without frequency counts are reported as 0.

Example:

```
                                    C%02 C%05 C%1  C%2  C%5  C1   C2   C4  CE10
11 5 1 1 8 1 1 1 0.02 377  0    0    0    0    0    1    0    0    0
11 5 1 1 8 2 1 1 0.01 274  1    0    0    0    0    1    0    0    0
11 5 1 1 7 1 1 2 0.02  17  0    0    1    0    0    1    2    0    0
11 5 1 1 6 1 3 1 0.03  51  0    0    0    0    7    1    1    0    0
11 5 1 1 6 2 2 2 0.01  68  0    0    0    0    2    3    3    1    0
```

5.2.9.Summarydata

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
limitofreporting	limit of reporting (mg/kg)
numberofsamples	size of sample (detects and nondetects)
numberofpositives	number of positive concentration values (detects)
the mean: mean *or* meanall	statistic for the mean
the median: med *or* medall	statistic for the median
max	statistic for the maximum
the variance: var *or* varall	statistic for the variance
the percentile: perc *or* percall	statistic for the percentile
percentile	specifies the percentage of the statistics **perc** and **percall**

- MCRA release 3 has an option to input summary data for residue concentrations. These data are then used for estimating parameters of a parametric model (see 6.4.3).
- Field names mean, meanall, med, medall, max, var, varall, perc, percall and percentile are optional and their order is free. Not all statistics need to be present in the table. See also last bullet.
- Statistics ending on '**all**' refer to statistics based on all samples including nondetects (concentrations below LORs), while statistics without suffix 'all' relate to statistics based on nonzero samples (nondetects) only.
- The use of equivalent statistics, like **mean** and **meanall**, for one product in the same row is not allowed.
- Be aware that statistics should be consistent e.g.: **med** is always smaller than **mean;** the calculated mean (nonzero samples only) that is derived from statistic **meanall** should be smaller than **max**; specifying **medall** implies that more than half the number of samples are detects (numberofpositives); specifying **percall** implies that the number of detects (numberofpositives) is greater than the percentage specified in column **percentile**.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing….. A warning is generated when 1) and 2) are not possible.
- Missing statistics are reported as 9999. A columns containing only missing values is not allowed and should be deleted.

Example:

```
                                        mean  meanall  med   max   var  perc percentile
11 5 1 1 7 1 1 1 0.02 377 1  .56    9999   9999 9999 9999 9999    9999
11 5 1 1 8 1 1 1 0.01 274 2  .59    9999   9999  .97 9999 9999    9999
11 5 1 1 8 1 2 1 0.02  17 4 9999     .23   9999 1.68  .56 9999    9999
11 5 1 1 9 1 1 1 0.02  51 9  .48    9999    .32 9999  .23  2.1      90
11 5 1 1 9 2 2 1 0.03  86 6 9999    9999    .62 9999 9999  2.5      95
```

5.2.10. Processing

field name	description
proccode	code of processing type
proctype (text)	description of processing type

disttype	indicator (1/2) whether simulated processing factors are restricted to the interval (0,1) using a logistic-normal distribution (1) or simulated processing factors are restricted to positive values using a log-normal distribution (2)
bulkingblending	indicator (0/1) for types of processing applied on large batches, e.g. juicing, sauce/puree (obligatory), 0 = no bulking/blending ; 1 = bulking/blending

- Information on bulking and blending is only relevant for modelling of processing effects in combination with unit variability and IESTI calculations, but should always be present in the table even when these effects are not explored.
- No missing values allowed.

Example:

```
 1  RAW                   1  0
 2  PEELING               1  0
 3  COOKING IN WATER      1  0
 4  BAKING OF BREAD       1  0
 5  CANNED/CONSERVED      1  0
 6  BREWING               1  0
 7  DRYING                2  0
 8  FRYING/BAKING IN FAT  1  0
 9  JUICING               1  1
10  MILLING               2  0
11  MARMALADE/JAM         1  1
12  OIL EXTRACTION        1  0
13  SAUCE/PUREE           1  1
14  CLEANING              1  0
15  WASHING/CLEANING      1  0
16  WINE MAKING           1  0
99  UNKNOWN               1  0
```

5.2.11.Processingfactor

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
proccode	code of processing type
procnom	nominal value (best estimate) of processing factor
procupp	upper value (estimate of 97.5th percentile or “worst case” estimate) of processing factor

- The value for procupp should be higher than the value for procnom.

Example:

```
11  5  1  1  6  1  1  1  9   0.48  1
11  5  1  1  6  1  1  1  15  0.40  0.64
11  5  1  1  7  1  2  1  7   1.04  2.63
11  5  1  1  8  1  2  1  15  0.40  0.64
11  5  1  1  8  1  2  1  9   0.35  0.97
```

5.2.12.Products

field name	description
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
productname (text)	product label
unitweight	nominal weight of a unit (gr)
edibleportion	edible portion (corrected large portion weight, gr)
largeportion	weight of a large portion (gr)

- Table Products contains both primary agricultural products as foods, both items can be distinguished by their productcategory number 1 or 9, respectively. Note that the fourth column has 3 digits for primary agricultural products whereas 4 digits are allowed for foods (see also example).
- When the nominal unit weight is unknown the value 0 is used.
- For foods, the value for unitweight, edibleportion and largeportion is not relevant, please fill all empty cells with value 0.
- Missing values for edibleportion and largeportion: 9999.

Example:

```
1   7  1    10  2  BEAN                              0     0  285
1   8  1    15  2  SPINACH                         111    90  562
1   8  4     1  1  POTATOES                        210  9999  420
9   9  0  1141  0  VEGETABLES MIX MEXICO FROZEN      0     0    0
9  18  0  1815  0  BAMI GORENG COOLFRESH MEAL        0     0    0
```

5.2.13. Productcategory

field name	description
productcategory	product category number
productcategoryname (text)	product category label

- No missing values allowed.

Example:

```
1  PRIM AGRICULTURAL COMMODITY
9  FOOD
```

5.2.14. Productgroup

field name	description
productcategory	product category number
productgroup	product group number
productgroupname (text)	product group label

- No missing values allowed.

Example:

```
1  6   GRAINS AND GRAIN PRODUCTS
1  7   PULSES, SEEDS, KERNELS AND NUTS
1  8   VEGETABLES, POTATOES, BEET AND TUBERS
1  9   FRUIT
```

```
9  18  MIXED DISHES
```

5.2.15. Product_subgroup

field name	description
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
productsubgroupname (text)	product subgroup label

- No missing values allowed.

Example:

```
1  7   1   PULSES
1  7   2   SEEDS
1  7   3   KERNELS
1  8   2   CABBAGES
1  8   4   POTATOES, ROOTS, TUBERS
```

5.2.16.VariabilityProd

field name	description
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (e.g. from special studies) for the lognormal and the beta distribution and the number of units in a composite sample.
- Estimates for unit variability are independent of the compound.
- All columns need to be present: use code 9999 when data are not available.
- When the parameter for unit variability is a coefficient of variation and the number of units equals 1, unit variability is ignored for this product.

Example:

```
1   7  1  10  2  5  9999   12
1   8  1  15  2  4  0.23   15
1   8  4   1  1  4  1.23  120
1   9  1   1  1  3  1.30   25
1   9  1   2  1  7  0.34   24
```

5.2.17.VariabilityCompProd

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number

productq	product quality number
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (e.g. from special studies) for the lognormal and the beta distribution that are dependent on the compound. Values for unit variability in table VariabilityProd are replaced by the new ones.

Example:

```
11  5  1  1   7  1  10  2 3  9999  12
11  5  1  1   8  1  15  2 2  0.10  15
12  4  1  1   7  1  10  2 3  0.20  12
12  4  1  1   8  1  15  2 3  0.15  15
13  4  1  1   9  1   2  1 7  0.34  20
```

5.2.18.VariabilityProcCompProd

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
proccode	processing type code
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (e.g. from special studies) for the lognormal and the beta distribution that are dependent on the combination of processing type and compound. Values for unit variability in table VariabilityProd and VariabilityCompProd are replaced by the new ones. This can be used for example to reset the variability factor to 1 for grape juice and raisins (dried grapes).

Example:

```
11  5  1  1  1  5  1  1    7    1  0.10    12
11  5  1  1  1  5  1  1    7    1  0.10    15
12  4  1  1  1  5  1  1    7    1  0.10    12
12  4  1  1  1  5  1  1    7    1  0.10    15
13  4  1  1  1  5  1  1   15    2  0.10    20
```

5.2.19.Agriculturaluse

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number

country (text)	code for country
year	year
useallowed	indicator (0/1) whether use of the compound for the commodity is allowed (1) or not (0)
perccroptreated	maximum percentage of the commodity that is treated with the compound
maximumresiduelimit	worst case value

- For combinations of compound and products that are not listed in table Agriculturaluse MCRA will assume that use is not allowed.

Example:

```
11  5  1  1  6  1  1  1  NL  1999   0   80  0.21
11  5  1  1  6  1  1  1  NL  2000   0   80  0.10
11  5  1  1  7  1  1  1  NL  1999   0   75  0.05
11  5  1  1  7  2  1  1  NL  2000   0   75  0.05
11  5  1  1  8  2  2  2  NL  1999   0   67  0.07
```

6. Performing a risk assessment

6. 1. Overview

From the Central Menu, MCRA is started by clicking the **go** button. In the next sections the input form for specifying MCRA program options is described. When the MCRA input form is entered for the first time, it looks like Figure 8 (only partly shown). The input form is divided into sections in which model options are to be specified. In section General, a short description of the MCRA program is displayed. After clicking any button in the form this section disappears and does not show up again until a new session is started. MCRA program options are specified by clicking the buttons. After clicking a button, the chosen option is displayed together with a **reset options** button. By clicking the **reset options** button, the input choice is erased and the program option buttons are shown again. Note, that only model specifications are displayed that are relevant to a specific choice e.g. empirical modelling is only a program option after clicking the **full** button in the Residues section.

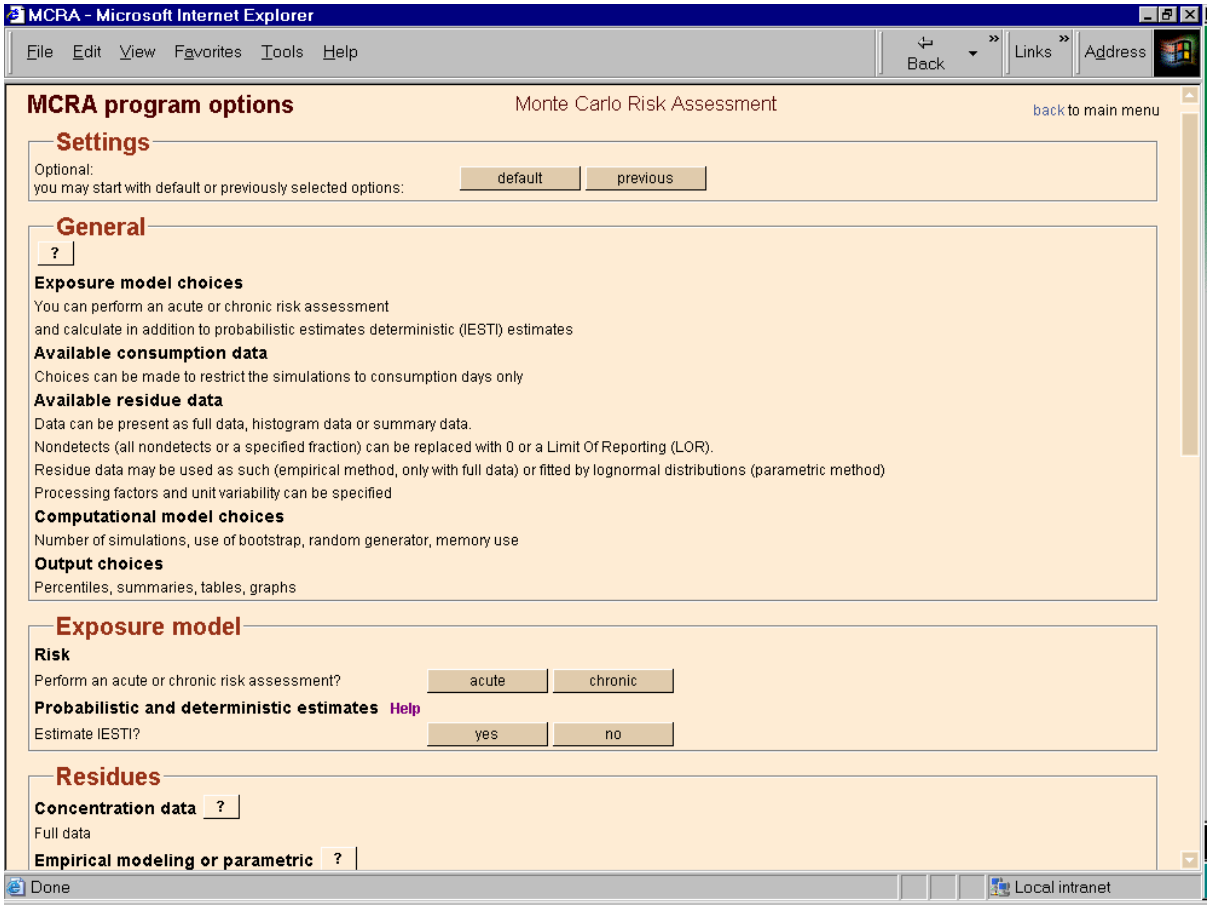


Figure 8: MCRA input form (only partly shown).

After introducing the input form with MCRA program options, a paragraph with information on graphical output is given.

6. 2. MCRA program options: Settings

Section Settings (see Figure 8) contains two buttons: the **default** button is used to specify in one click acute risk based on an empirical approach using full data. IESTI, percent crop treated, processing and variability options are not implemented. By clicking the **previous** button, all program settings from the last performed MCRA analysis are recalled and displayed in the input form. Note, the **previous** button is only displayed when a MCRA has been performed in the past.

6. 3.MCRA program options: Exposure model

6.3.1.Acute or chronic risk

The MCRA program is used to perform an acute or chronic risk assessment. In Figure 9, acute risk is specified (click the **acute** button). When acute risk is specified, options concerning deterministic estimates (IESTI) become visible.

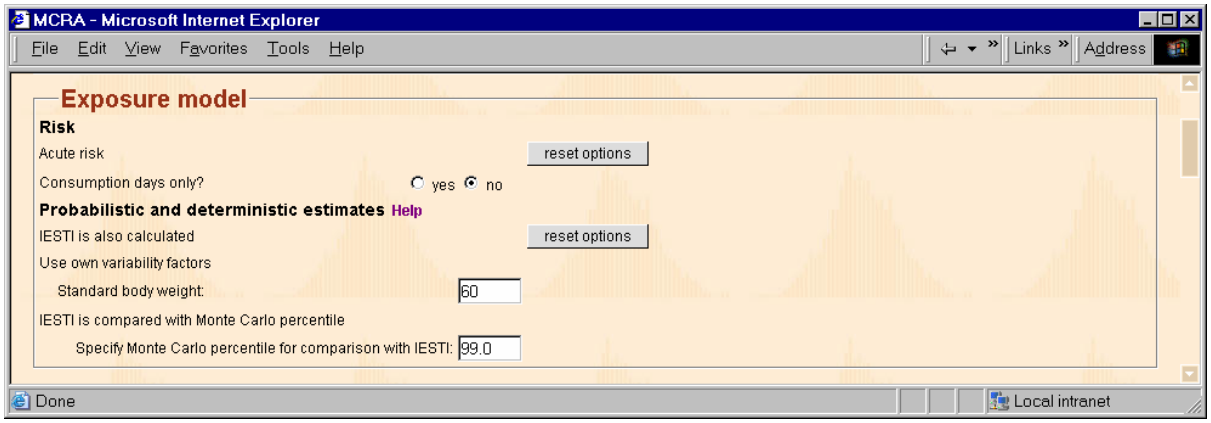


Figure 9:'Exposure model', acute risk, IESTI

IESTI estimates

The IESTI (International Estimated Short-Term Intake) is a prediction of the short-term intake of a compound on the basis of the assumptions of high daily food consumption per person and highest residues and, in case of blending and bulking, the median residue from supervised trials. The IESTI is expressed in mg/kg body weight and estimated per product.

In Figure 9, IESTI estimates are requested (click the **reset** and then the **yes** button). Own unit variability factors and a standard body weight of 60 kg is specified. The IESTI is compared with estimates of a specified percentile (per commodity) of the Monte Carlo simulation. In the output (not shown) two kinds of estimates of the Monte Carlo percentile are given: one for 'All days' and one for 'Consumption days only'. Be aware that specification of option 'Consumption days only = *yes*' may alter the interpretation (and estimate) of the percentile for 'All days'. In latter case, the estimate refers to a smaller subset containing consumption days only. However, note that still not every commodity is eaten on every consumption day. The interpretation and estimate of the percentile for 'Consumption days only' is not affected by setting option 'Consumption days only' to *yes*.

The IESTI calculations correspond to the definition of FAO/WHO (2002) that may be considered as the deterministic counterpart of the probabilistic approach used in Monte Carlo Risk Assessment.

Chronic risk assessment

In dietary risk assessment, chronic exposure is defined as the long-run average of daily intakes of a dietary component by an individual. Chronic exposure may be estimated for dietary components that are consumed on a nearly daily basis. Before estimating the chronic percentiles of the distribution, non-normal exposure data are transformed to approximate normality, following an approach proposed by Nusser et al. (1996). The cumulative risk is based on the estimated percentiles that are calculated on a daily basis and are simply summed over the number of specified years. In the MCRA program, before chronic exposure is estimated, the average intake is calculated as the consumption multiplied by the average value of the residues (nondetects and detects) and, if specified, applying processing and/or replacing zeros with the LOR (based on percent crop treated).

In Figure 10, chronic risk is specified (click the **chronic** button) and the options concerning 'long term exposure' become visible. Two transformations may be specified, a power transformation (*yes*) or a logarithmic transformation (*no*). Usually, a power transformation is satisfactory. In addition, the cumulative risk for a certain age (in years) is specified. A chronic risk assessment is only performed when the total fraction of zero intakes (due to nondetects or days without consumption) is below a

user-specified low percentage (e.g. 1%). Then, for computational purposes zero intakes are replaced by 0.0001 times the mean intake. When the proportion of zero intakes is higher, a warning message is printed and the simulation aborts.
Residue data for a chronic risk assessment may be present as full, summary or histogram data. For full data, a choice can be made between a parametric or a non-parametric (empirical) approach. For summary or histogram data a parametric approach is obligatory. Note, that option consumption-days only (see Figure 9) is not relevant for chronic risk.

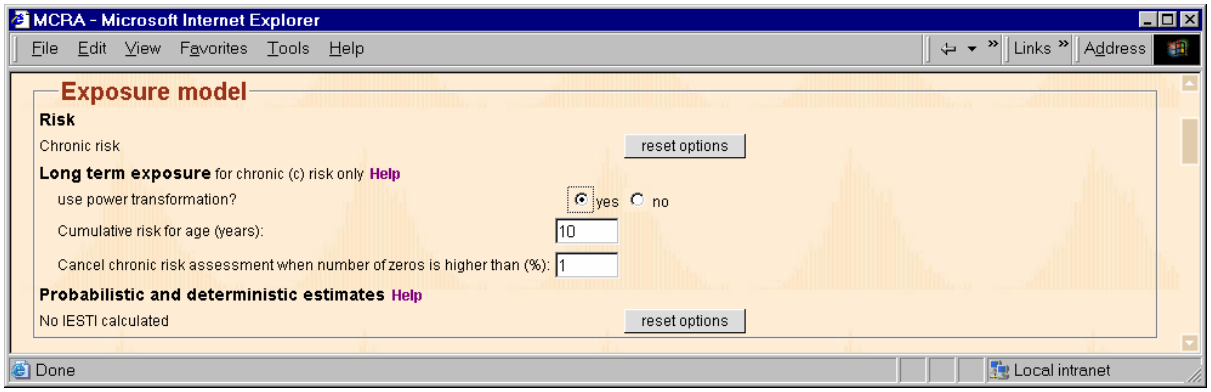


Figure 10: 'Exposure model', chronic risk, model options

6. 4. MCRA Program options: Residues

6.4.1. Full, summary or histogram data

Residue concentration data (see Figure 8) can be present as full data (a list of residue concentrations is available), summary data (only some summary statistics, for example means, percentiles or maxima are available) or histogram data (only numbers of observations classified in intervals are available). In Figure 11, full data are specified. See the Reference Manual for more information.

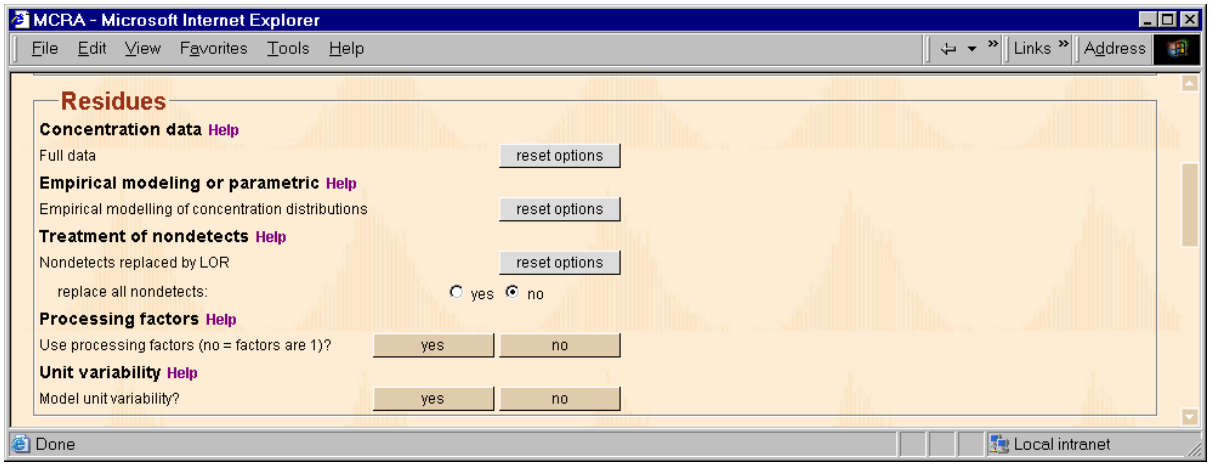


Figure 11: 'Residues', full data, empirical modelling, replace nondetects

6.4.2. Nondetects and percent crop treated

In many cases of residue risk assessment (e.g. pesticides) most monitoring measurements are nondetects, i.e. no quantitative measurement is reported. Only values higher than the Limit Of Reporting (LOR) are reported. When a compound can enter the food chain only via crop treatment, and when the percentage of crop treated is (approximately) known, then this knowledge may be used to infer that some of the monitoring measurements should be real zeroes, contributing nothing to the

intake, whereas other nondetects in the monitoring data could have any value below the limit of reporting.
Nondetects (all nondetects or a specified fraction) can be replaced with 0 or the LOR (see Figure 11). If percent crop treated data are available, then replacement by LOR can be restricted to an appropriate fraction of the nondetects by specifying 'replace all nondetects = *no*'.

6.4.3. Empirical or parametric modelling

In the probabilistic model, a distribution of residue data is used to sample from. A choice can be made between a parametric or a non-parametric (empirical) approach. Residue data may be used as such (empirical modelling, only with full data) or fitted by lognormal distributions (parametric modelling, based on either full data, summary data or histogram data).
Parametric modelling becomes important in data-scarce situations. The lognormal distribution with parameters μ and σ has been selected as being both theoretically sensible and practically useful (Shimizu & Crow 1988, Van der Voet et al. 1999). The non-parametric approach requires more data to obtain a satisfying representation of the full distribution.

6.4.3.1 Non-parametric approach

In the non-parametric approach, set 'empirical modelling' to *yes (*see Figure 11), residue values are sampled at random from the available data and combined with food consumption data to generate the intake distribution of exposure values.

6.4.3.2 Parametric approach

In the parametric approach, (in case of full data, set 'empirical modelling' to *no*), residue concentrations per food commodity are sampled from parametric distributions based on full, histogram or summary data. Parameters μ and σ of the lognormal distribution are estimated using the log-transformed non-zero residue concentrations (full data) or condensed data (summary or histogram data). Estimation of the variance and/or mean may fail because sometimes compounds on specific food commodities are sparse or even missing. A related question is the reliability of estimates based on a few degrees of freedom. To overcome these problems, basically, concentration data on other commodities are used to give sufficient data to base estimates upon. Commodities are classified into groups of similar products and missing or unreliable parameters are estimated using all concentration data in a group. This process of using concentration data on similar products to base estimates for μ and σ upon is called pooling.
When parametric modelling is specified, pooling of means and variances is optional. In Figure 12, parametric modelling and 'pooling of means and variances' is specified. Specifying 'no pooling' is only an alternative when enough data are available to estimate μ's and σ's for all commodities. In case of missing parameters, a warning message will be printed. MCRA should then be rerun with 'pooling of means /variances' set to *yes*.
Specifying pooling means that products are automatically assigned to productgroups and pooled. The identification of productsgroups is based on the first 3 numbers of the commodity code (productcategory, productgroup, productsubgroup). Commodities with the same numbers are placed into a productgroup.

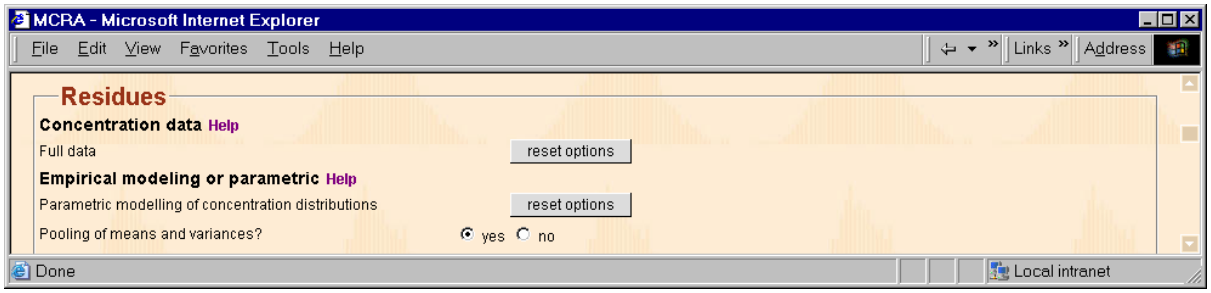


Figure 12: 'Residues', parametric modelling, pooling of means/variances

Pooling is performed in a two step procedure following the next scheme:

1. **Test homogeneity of variances within productgroups**
   if variances are homogeneous,
   pool variances and
   test homogeneity of means within productgroups
   if means are homogeneous,
   pool means.
2. **Test homogeneity of variances of commodities with df < 10 against overall-variance**
   if variances are homogeneous,
   replace variances with overall-variance

Results of step 1 and 2 are (sub)groups with:

a) pooled variances and pooled means,
b) pooled variances and the original (unpooled, heterogene) means,
c) the original (unpooled, heterogene) variances and original means.

An example of pooling is given in Appendix 1.

6.4.4. Processing

Concentrations in the consumed food may be different from the monitoring residue due to processing such as peeling, washing and cooking. Usually, processing lowers the residue level in the consumed food compared to the concentration in the raw product. The effect of processing is modelled by multiplying the monitoring residue by a factor f_k which will typically be between 0 and 1. Occasionally, the processing factor may also be > 1, e.g. for drying. Often, processing factors are not exactly known or information is of limited quality. These uncertainties may be entered into the model by specifying two values: $f_{k,nom}$, the nominal value, typically some sort of mean; and, $f_{k,upp}$, an upper 95% confidence limit. Distribution based processing factors require both values whereas for fixed factors only $f_{k,upp}$ need to be specified. No processing implies that $f_k = 1$.

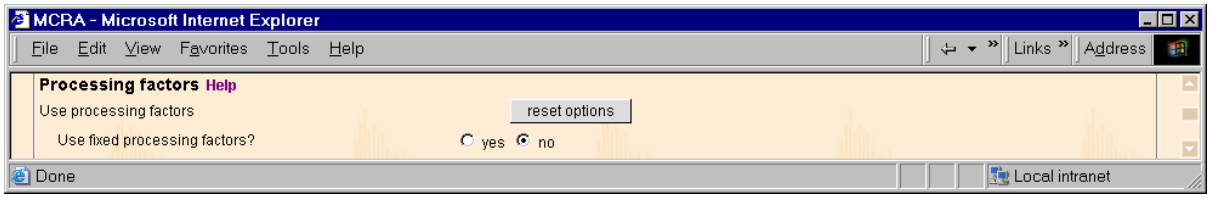


Figure 13: 'Residues', processing factors

The use of processing factors f_k is indicated by setting 'Use processing factors' to *yes*. Processing factors are read from table Processingfactor ($f_{k,nom}$ = proc_nom, $f_{k,upp}$ = proc_upp) and processing codes and labels from table Processing. Note that specifying no processing is a worst case scenario ($f_k = f_{k,upp} = 1$), e.g. no processing.

The program multiplies residue values with fixed processing factors (in which case the conservative value $f_k = f_{k,upp}$ is used), or with random values sampled from a normal distribution with parameters μ and σ. The mean and standard deviation are based on transformed values of $f_{k,upp}$ and $f_{k,nom}$. The type of transformation for each processing type is specified in the last column of table Processing. Choose disttype = 1 for a logistic-normal distribution or disttype = 2 for a log-normal distribution. In Figure 13, distribution based processing factors are specified (set 'Use fixed processing factors' to *no*).

To process simultaneously some commodities using fixed factors and others distribution based, set 'Use fixed processing factors' to *no*. Now, fixed factors f_k are obtained by providing only $f_{k,upp}$ whereas random factors f_k are sampled when both $f_{k,upp}$ and $f_{k,nom}$ are given.

It is not necessary to fill out a complete list of processing factors for all commodities. Missing values of $f_{k,nom}$ and $f_{k,upp}$ are, by default, replaced by the value 1.

6.4.5. Unit variability

Monitoring measurements are typically made on homogenised composite samples. Each sample is composed of nu_k units with nominal unit weight wu_k each. The weight of a composite sample is often larger than a daily consumer portion. This implies that the mean level of the monitoring residue may be a fair estimate of the mean level of the raw commodity, but the variability of the monitoring measurements is certainly not appropriate to estimate the variance. Therefore, acute risks may be higher than would follow from a direct use of the composite sample data. This problem has been addressed by modelling unit variability. Set 'Model unit variability' to *yes* to display all program options concerning unit variability (see Figure 14).

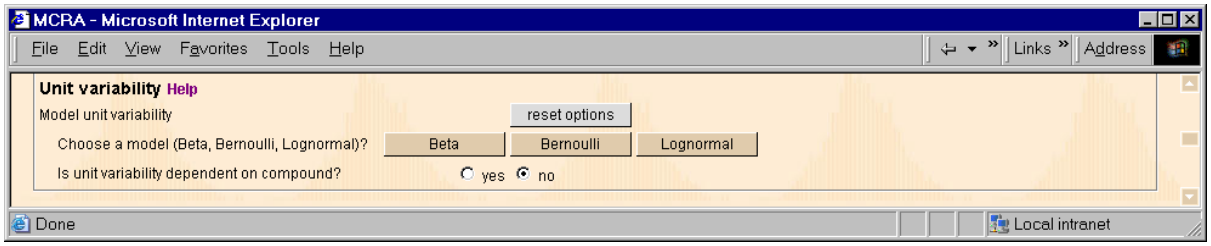


Figure 14: 'Residues', sample unit variability

In MCRA the following three models for unit variability can be selected:

1. **Beta model**, requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates);
2. **Bernoulli model**, requires only knowledge of the number of units in a composite sample (results are always conservative);
3. **Lognormal model**, requires only knowledge of the variability between units (realistic or conservative estimates).

6.4.5.1 Bernoulli distribution

In practice, measurements on individual units to obtain a measure for unit variability are not very common. Therefore, the number of units nu_k in the composite sample is used to define the parameter for unit variability (see van der Voet et al. 2001). When the number of units nu_k in the composite sample is missing, the nominal unit weight wu_k is used to calculate the parameter for unit variability. The following variability factors v are recommended: for large crops (wu_k > 250g) value v = 5; for medium crops (wu_k 25- 250g) v = 7; and for small crops ($wu_k \leq$ 25g) v = 1 (FAO/WHO, 1997). For commodities which are processed in large batches, e.g. juicing, marmalade/jam, sauce/puree, v = 1. The latter information is specified in field bulkingblending of table Processing (see 5.2.10). The number of units within a consumption is calculated and for each unit a Bernoulli distribution is used to sample the monitoring residue itself with probability (v-1)/v or a multiple v of it with probability 1/v (see Figure 15).

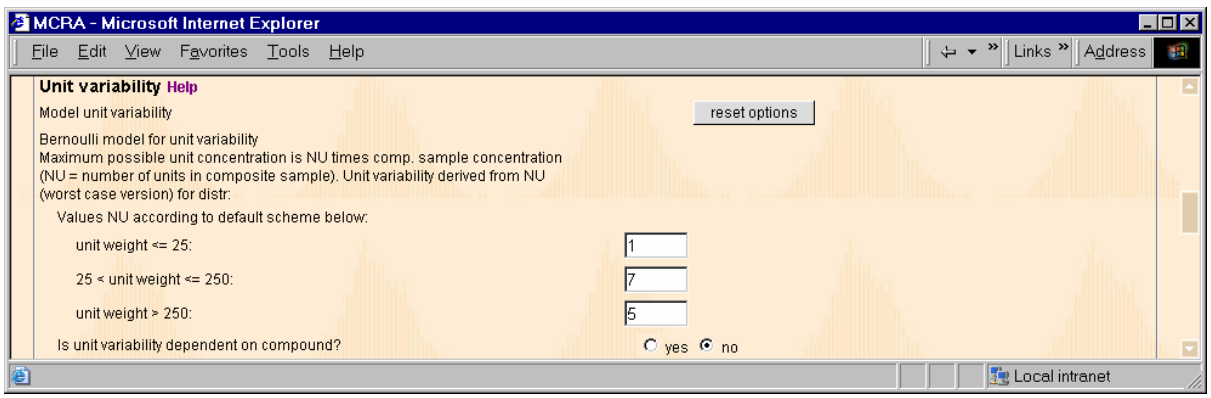


Figure 15: 'Residues', Bernoulli distribution

6.4.5.2 Estimated parameters for unit variability

When parameters for unit variability based on empirical studies are available, these are used to simulate concentrations for a unit, assuming a parametric form for the unit-to-unit variability within a batch e.g. the beta or lognormal distribution

Table 2 describes the four options when a parametric form for unit variability is specified. Residues are simulated for a new unit in the batch using a lognormal distribution or for a unit belonging to the composite sample leading to the use of the beta distribution.

	Simulate for new unit in batch (log-normal distribution)	Simulate for unit belonging to composite sample (beta distribution)
Estimates of unit variability are realistic (R)	• no censoring at cm_k • no upper limit to the unit concentration	• no censoring at cm_k • unit values never higher than $nu_k \cdot cm_k$
Estimates of unit variability are conservative (C)	• unit values will be left-censored at cm_k • no upper limit to the unit concentration	• unit values will be left-censored at cm_k • unit values never higher than $nu_k \cdot cm_k$

Table 2: Choices for estimated variability factors. cm_k = value of the composite sample concentration, nu_k = number of units in composite sample.

6.4.5.3 Beta model

When the beta distribution is chosen (see Figure 16 the parameter for unit variability is specified as a variability factor *v* or a coefficient of variation *cv* of the unit values in the composite sample. Variability factors *v* (97.5th percentile divided by mean), coefficient of variation *cv* (standard deviation divided by mean) and number of units *nu* in the composite sample are retrieved from table VariabilityProd when, as in Figure 14, unit variability is independent of the compound and processing type. If the variability factor is dependent on compound and processing type, data are expected in tables VariabilityCompProd or Variability ProcCompProd, respectively. If the parameter for variability is missing, zero variability is assumed, and the unit concentrations are equal to the sampled composite sample concentrations.

A choice is to be made whether the supplied values for variability are realistic or conservative estimates. In the latter case, unit values are left-censored at the value of the mean (composite sample concentration). If there are no user-defined values for the number of units in the composite sample these are taken using a default scheme of nominal unit weights. This scheme follows in principle the definition of FAO/WHO (1997), as illustrated in 0, but can be modified by the user.

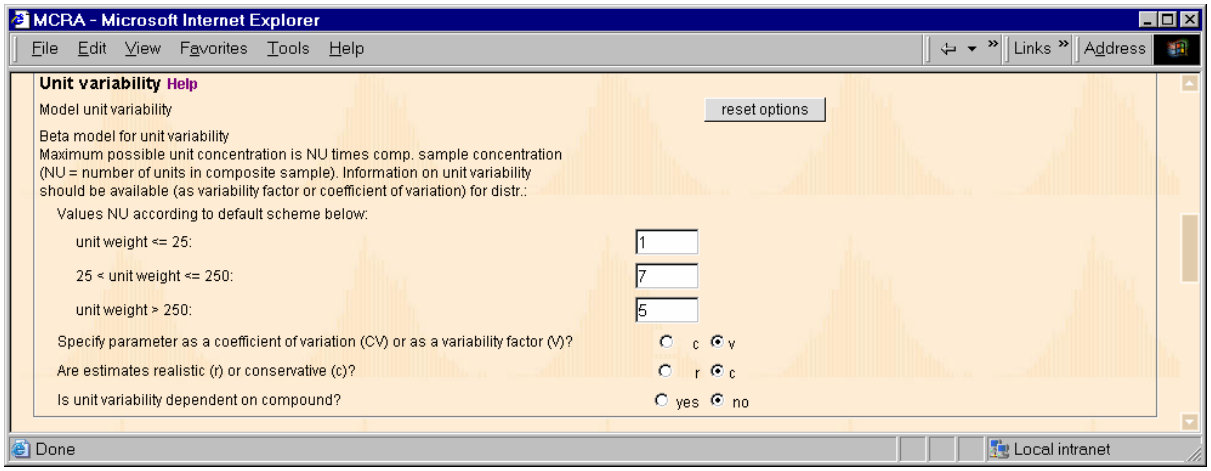


Figure 16:'Residues', Beta distribution

6.4.5.3 Lognormal distribution using estimated parameters

In Figure 17, a parametric form for the unit-to-unit variability is specified. Residue values are simulated for new units in the batch leading to the lognormal distribution. The parameter for unit variability is specified as a coefficient of variation *cv* or a variability factor *v*.

The conversion of a variability factor into parameters of the lognormal distribution requires an exact definition of what is meant. Here, the variability factor is defined as the 97.5th percentile of the concentration in the individual measurements divided by the corresponding mean concentration seen in the composite sample. Finally a choice is made whether a realistic or a conservative approach is modelled. In the conservative approach, unit residue values of the composite sample are left-censored at the value of the monitoring residue. When a realistic approach is defined, the unit value may be lower than the value of the monitoring residue.

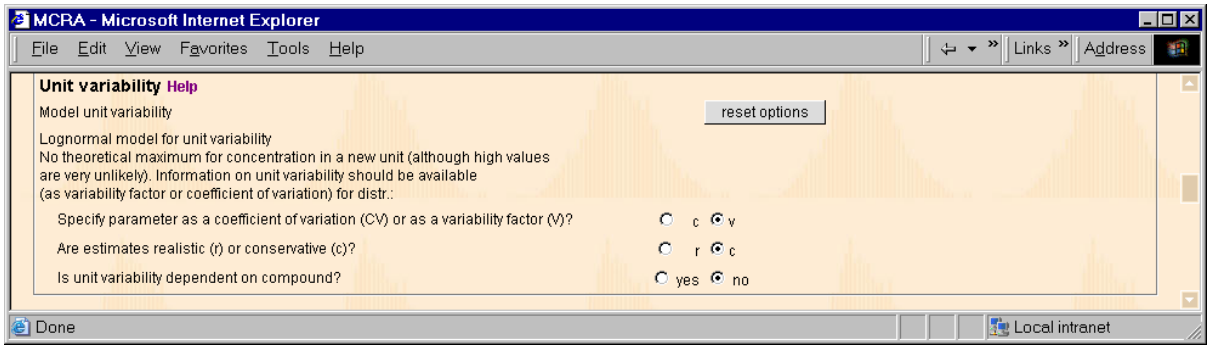


Figure 17:'Residues', lognormal distribution

6. 5. MCRA Program options: Simulate

6.5.1. Number of simulations

Consumers are randomly sampled from the consumer database. Each time a consumer is sampled, it contributes to the probability distribution of intakes. Each individual contribution is called a simulation.

The number of simulations or total number of sampled consumers specified in Figure 18 is 500000. There is no upper limit. Keep in mind that the number of commodities, certainly when processing and variability factors are involved, is a major factor affecting run-time.

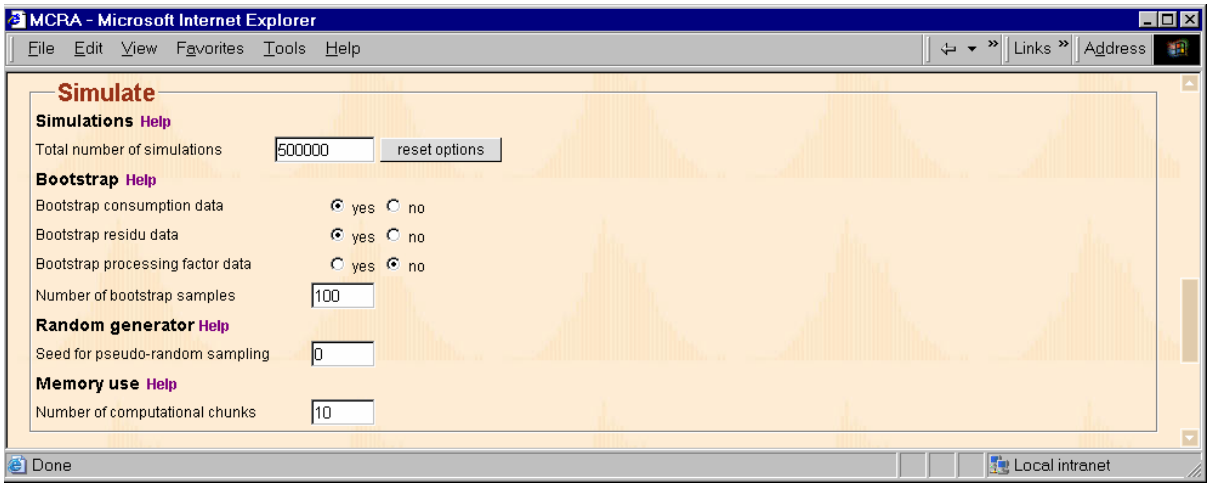


Figure 18: 'Simulate', computational model choices

6.5.2. Bootstrap

The uncertainty of output statistics (e.g. mean or percentiles of the exposure distribution) is assessed by bootstrapping. The bootstrap can be applied on the level of consumers and/or residues (in future versions also processing factor distributions). From each level, with replacement, a bootstrap sample is drawn and the corresponding intake distribution is calculated. The sample provides a mean, maximum and specified percentiles. This process is repeated according to specified number of bootstrap samples. All replicates contain information that can be used to make inferences from the data, e.g. to establish the uncertainty of mean, maximum and percentiles.
In Figure 18, 100 bootstrap sample are specified. Suppose, the number of simulations is 10000, then each bootstrap sample consists of 100 values. The number of values within a bootstrap sample restricts which percentiles are displayed. Here, the highest possible percentile for which uncertainty information can be calculated is the 99.0^{th}, for a bootstrap-sample containing 1000 values it is the 99.9^{th} percentile.
Whereas 10000 simulations (100 bootstrap samples of 100 simulations) may be appropriate for testing purposes, a larger number (e.g. 500000 = 1000 bootstrap samples of 500 simulations is recommended for accurate results.

6.5.3. Pseudo-random sampling

The Monte Carlo simulation uses a pseudo-random number generator that is initialised by setting the seed. To get time-based values, set seed to zero and the generated sequence of random numbers is based on a default value which is printed in the program output. Using this value in a second run will result in identical simulation results provided that the model or number of iterations did not change.

6.5.4. Memory use

In general, the capacity of the internal memory restricts the size of the simulation. To overcome this problem, a Monte Carlo Risk Assessment can be performed in computational chunks. The results of each chunk are stored for later use. The chunk size is equal to the total number of simulations divided by the number of chunks (= simulations/chunks). Chunk size and number of chunks affect the processing time. When the number of simulations within a chunk is too high, the performance of the computer is seriously degraded and swapping will occur. Advised is to rerun the simulation with a higher number of chunks. The estimated CPU-time in the log-file may be an indication to determine the optimal number of chunks. Best strategy is to keep the number of chunks as low as possible conditional upon chunk size. Factors influencing the chunk size are: the total number of commodities and/or the total number of combinations of commodities and processing types. The number of chunks should always be less than the number of bootstraps.

6. 6. MCRA Program options: Output

In Figure 19, percentiles are specified, each percentile separated by a space. The next item is needed for summarising the contribution of commodities to the right tail of the exposure distribution and to display a graph of the upper tail. The percentile may be specified, but specifying an exposure value instead overrules the percentage. After all program option buttons are set, the form is submitted and a MCRA analysis is started. Note, that the input form can only be submitted when no program option buttons are shown anymore (except for the **default** and **previous** button). As long as program option buttons are not clicked, the form cannot be submitted.
The corresponding exposure values are written to file UpperDistrIntake.lis.
Default output is: a summary of the data and information on percentiles. When the bootstrap is specified, also information on the uncertainty of percentiles is available.
For acute risk, in addition, results of the simulation, summaries of the contribution of commodities to the upper right tail and total exposure distribution, graphs of the upper right tail and total exposure distribution, information on the average residue on commodities and consumer-characteristics of the 10 consumers with the highest intake is available. For chronic risk, a diagnostic graph is plotted.
Output is viewed using the link on the homepage.

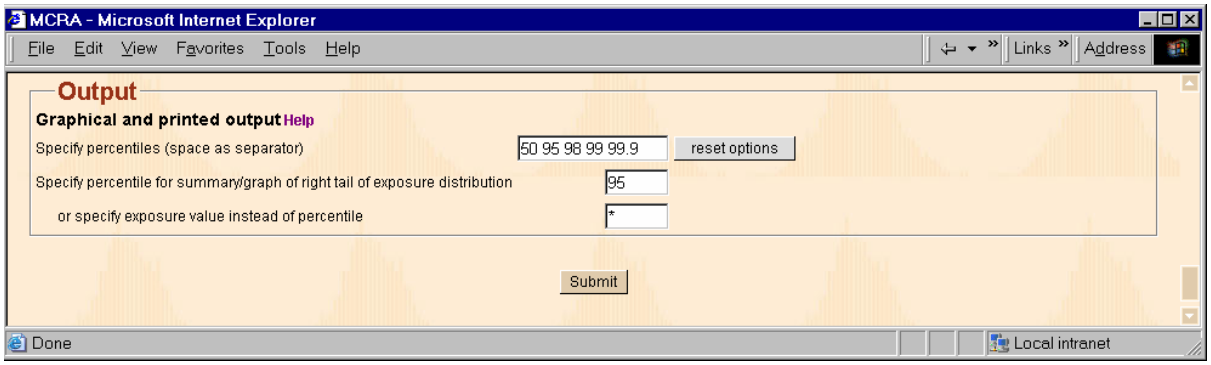


Figure 19: 'Output', percentiles and summary/graph of right tail of exposure distribution

6. 7. CHECK PROGRESS: Checking on the progress of job execution

After submitting the form, all model specifications are passed to the server and the analysis is started. The user is requested to wait until all calculations are performed and output is written to the user directory. Performing an exposure assessment may take considerable time if the data files are large and/or if the number of simulations is very high. The progress of program execution can be followed by clicking de link SHOW PROCESSING TIME in the wait window.

6. 8. VIEW OUTPUT: Viewing the output of an exposure assessment

If you want to look at the results of your last MCRA run, choose **View output of last simulation**. This brings you to a screen where both tabular and graphical results can be viewed. To view ComponentOne Charts, an ActiveX-aware browser is needed. Users can still manipulate the chart they view by right-clicking the chart, which brings up the Property Editor. Some more information is available under the link InfoCharts on the View Output screen. ComponentOne Chart graphs can be printed, or copied to the Windows clipboard, for later inclusion in documents.
Tabular output is available in separate ASCII output text files in the output directory. Alternatively, text can be copied and pasted from the VIEW OUTPUT window into another document. In order to obtain a proper lay-out the function Paste Special from the Edit menu should be used, selecting "Unformatted text".

6. 9.HELP on CHARTs

ComponentOne Chart is comprised of a 2D Control (ActiveX) for use in Windows applications. The control is stored in a so-called Cabinet-file, Olectra.CAB. A licence pack file Olectra.LPK is needed to register the control. To be able to view a ComponentOne Chart, the cabinet file and license pack have to be downloaded. Depending on the security level of your Internet Explorer, you may get the chart. You can change the security settings by doing the following:
Click:
1. Tools
2. Internet Options
3. Security
4. Local Intranet
5. Custom Level: set ActiveX controls and plug-ins: 4 x Enable
6. OK
7. OK

To be sure that changes to the charts (after rerunning the program) are displayed by the browser, you may need to do the following.
Click:
1. Tools
2. Internet Options

3. General
4. Settings
5. Check: every visit to the page
6. OK
7. OK

6.9.1.To display the Property Editor

Click the right mouse button over any part of the chart and select properties of the pop-up menu. Select the tab that corresponds to the element of the chart that you want to edit. Click **OK** or **Cancel** to close the Property Editor.

6.9.2.Interacting with Chart

You can interact with the chart as it is running to examine data more closely or visually isolate a part of the chart. The interactions described here affect the chart displayed inside the ChartArea; other chart elements like the header are not affected. ComponentOne Chart provides users with 2 different mechanisms for zooming the chart: Graphics zoom and Axis zoom. Performing a Graphics zoom enlarges the selected area of a chart, while not necessarily showing the axes. Performing an Axis zoom changes the minimum and maximum data values to those selected, and redraws only that data with axes. Scaling, moving, or graphics zooming the chart sets the PlotArea margin properties, so the chart will not automatically control margins anymore when other chart properties change.

To Scale the Chart:

1. Press CTRL, and hold down both mouse buttons (or middle button on 3-button mouse).
2. Move the mouse down to increase chart size, or move the mouse up to decrease chart size.

To Move the Chart:

1. Press SHIFT, and hold down both mouse buttons (or the middle button on 3-button mouse).
2. Move the mouse to change the positioning of the chart inside the ChartArea.

To Graphics Zoom an Area of the Chart:

1. Press CTRL, and hold down left mouse button.
2. Drag mouse to select zoom area and release the mouse button.

To Axis Zoom the Chart:

1. Press SHIFT, and hold down left mouse button.
2. Drag the mouse to select the zoom area and release the mouse button.

To Rotate the Chart (Bar/pie charts displaying 3D effect only):

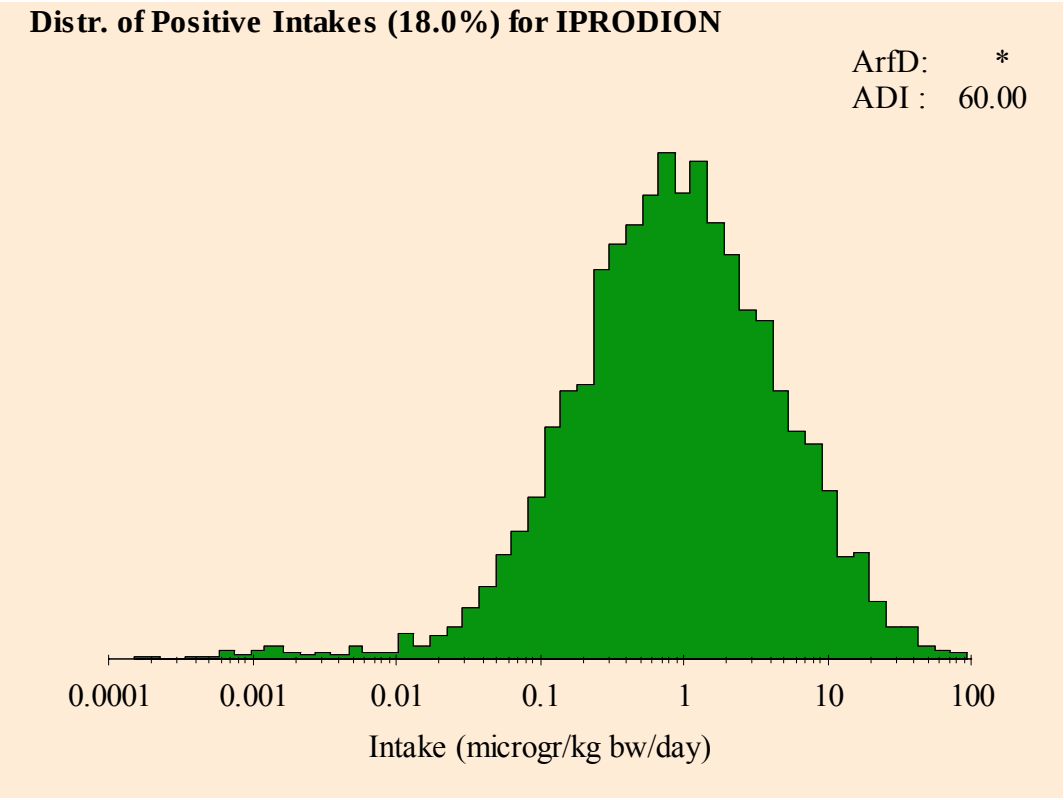
1. Hold down both mouse buttons (or middle button on 3-button mouse).
2. Move mouse up or down to change the 3D inclination.
3. On bar charts, you can also move mouse left or right to change the 3D rotation angle.

To Reset to Automatic Scale and Position:

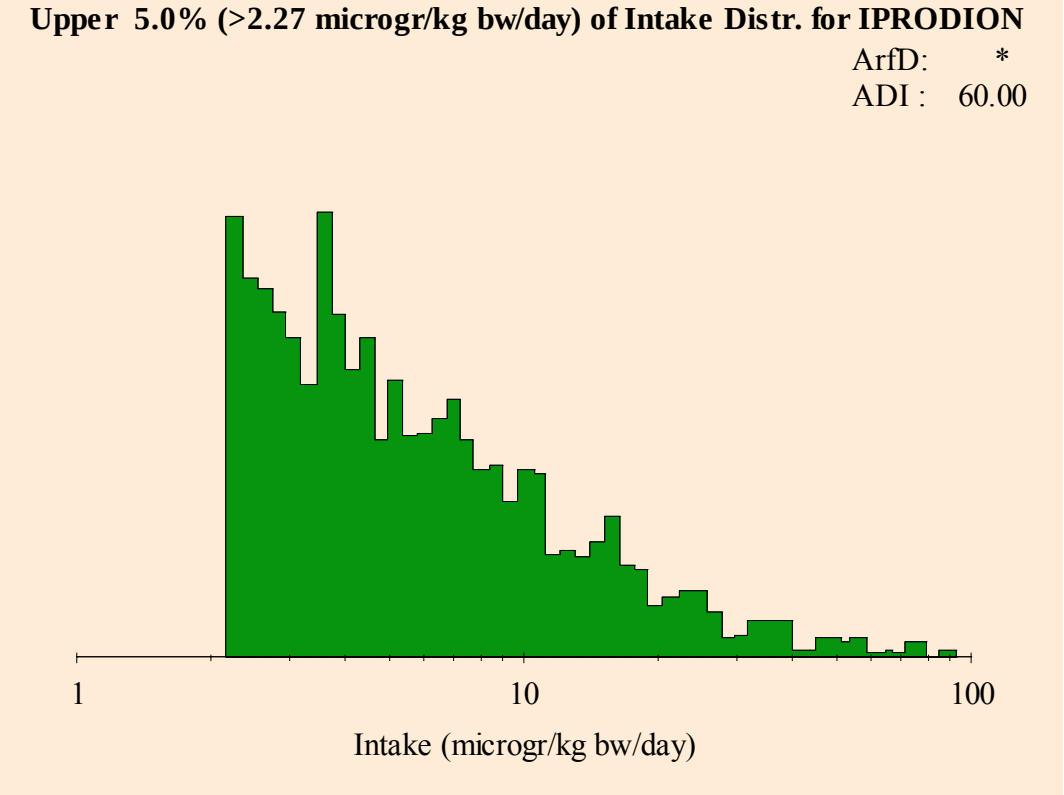
Press the “r” key to remove all scaling, moving, and zooming effects; chart regains control of PlotArea margins.

7. Example output

7.1.1. Distribution of exposure



Total Distribution



Upper Tail Distribution

7.1.2. Percentiles

Residue: IPRODION

****
Percentiles, maximum and average intake
****

Variability of intakes:

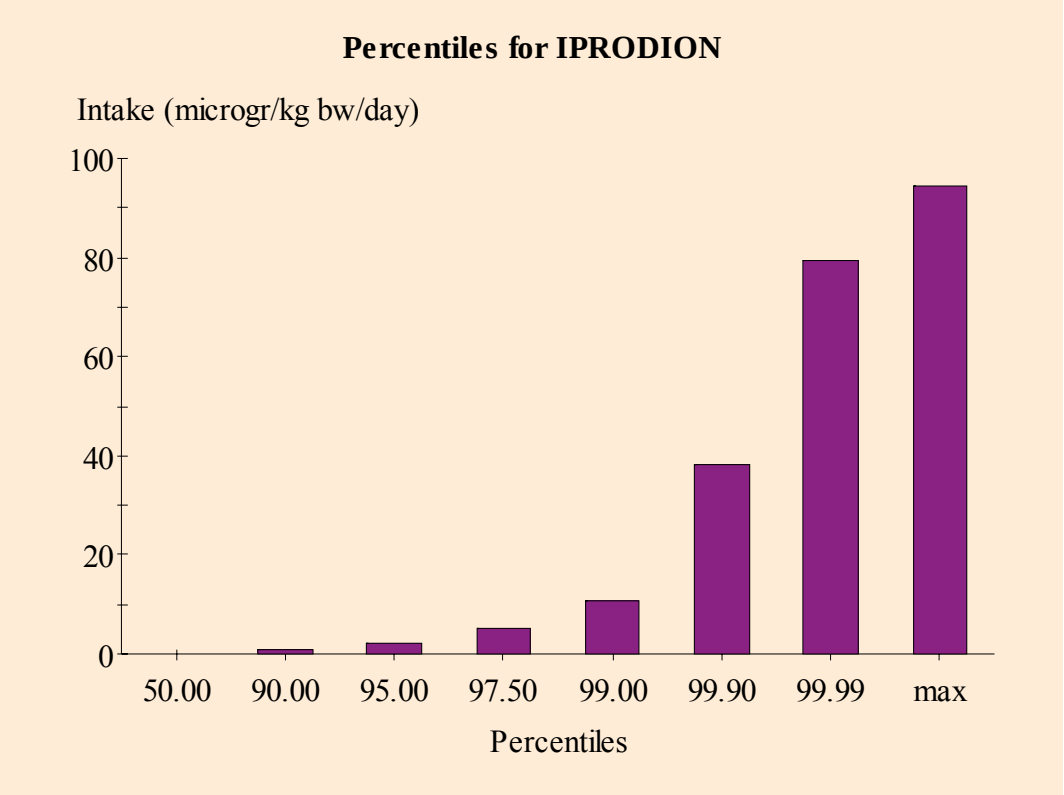
Percentile	IPRODION (microgr/kg bw/day)
50.00	0.00000
90.00	0.70588
95.00	2.27114
97.50	5.00653
99.00	10.87610
99.90	38.11885
99.99	79.19644

Uncertainty of percentiles (Confidence limits)

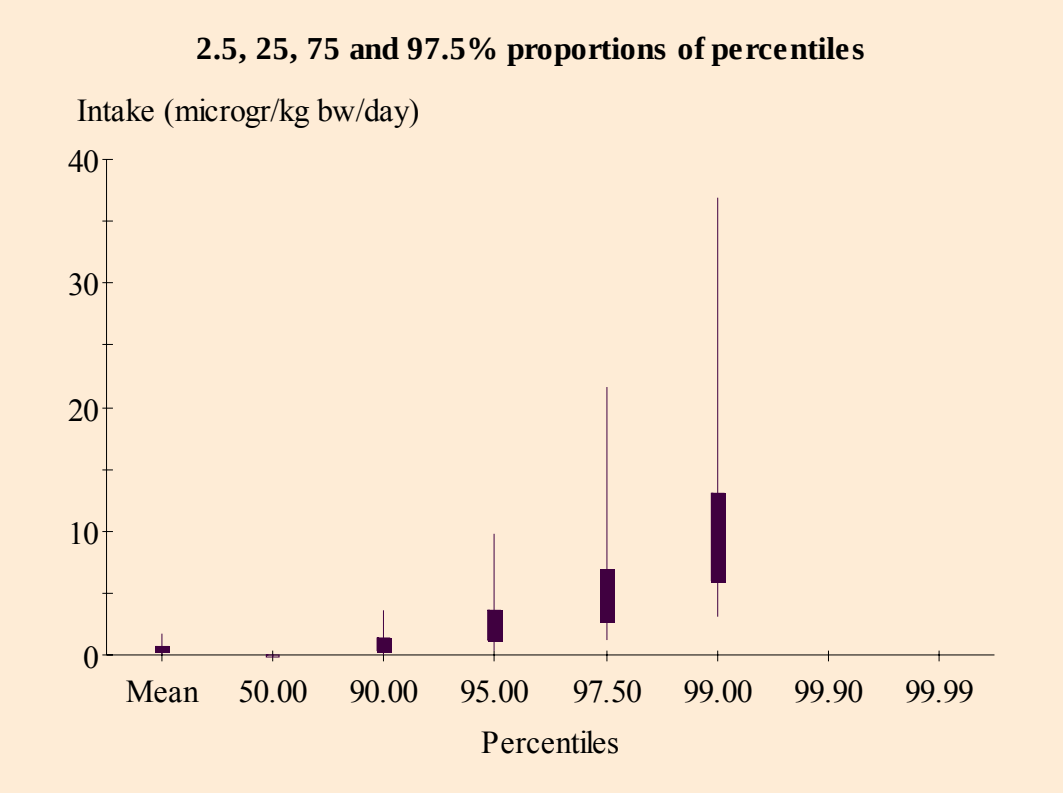
	2.5%	25%	75%	97.5%
50.00	0.0000	0.0000	0.0000	0.0000
90.00	0.1723	0.4105	1.3320	3.5846
95.00	0.4739	1.1596	3.6496	9.8238
97.50	1.3536	2.7433	6.9442	21.5466
99.00	3.2220	5.9429	13.0070	36.8480
99.90	*	*	*	*
99.99	*	*	*	*

maximum				
94.609	5.5256	12.5487	34.9944	80.6897

mean				
0.502	0.1565	0.2672	0.6465	1.6779

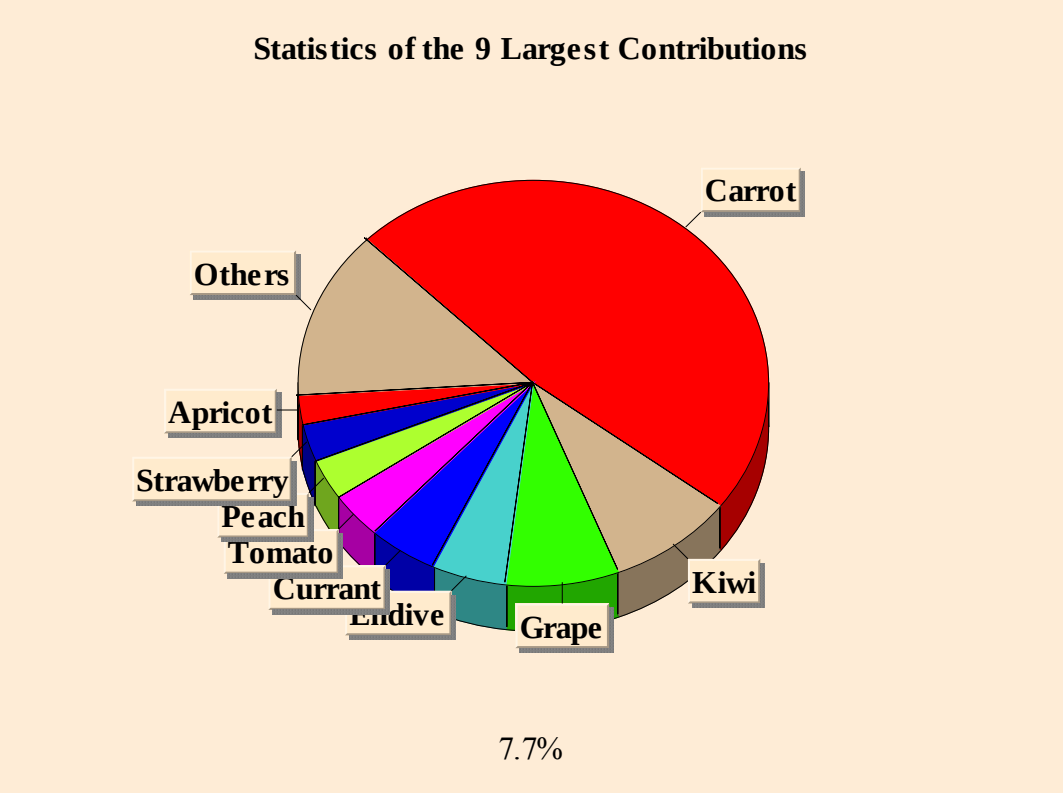


Acute Percentiles

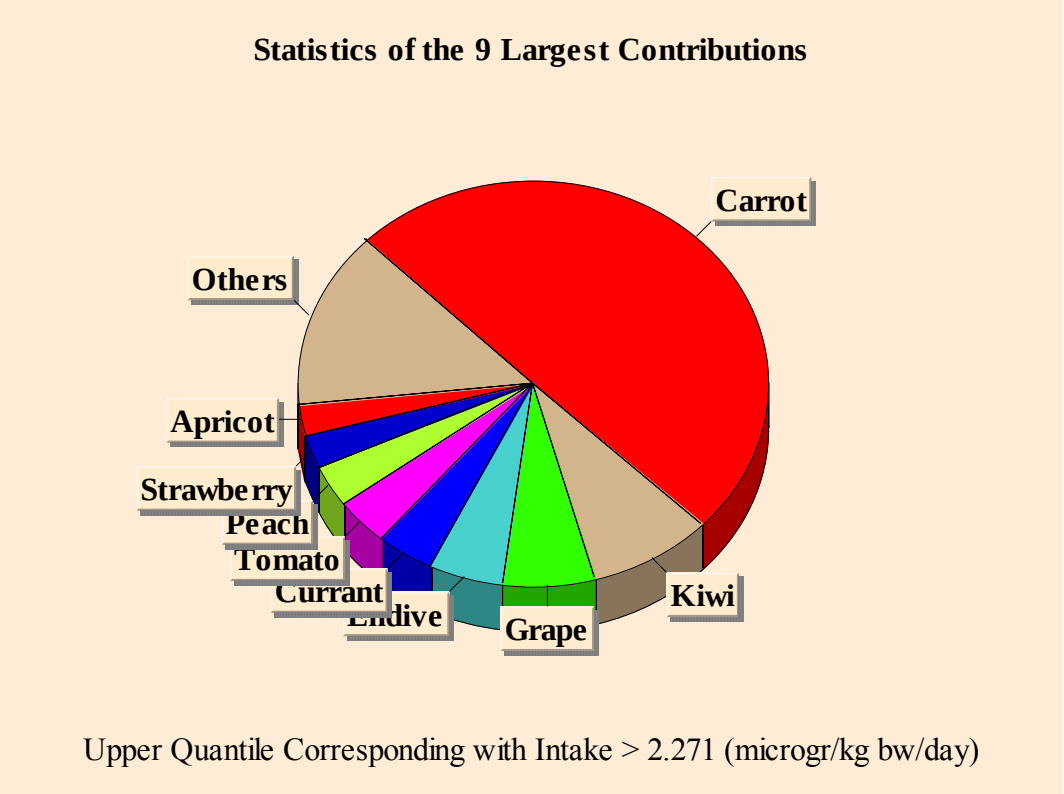


Bootstrap Intervals

7.1.3. Exposure contributions



Contribution Total Distribution



Contribution Upper Tail Distribution

Appendix 1. Example of pooling commodities in parametric modelling

Example: Automatic pooling of means and/or variances

The variances and means of products are pooled within a product group. Product groups are formed, based on columns productcategory, productgroup and productsubgroup in combination with factor Allowed (see example, factor Levels (= commodity code) and Allowed). Each commodity is characterised by a commodity code built hierarchically from 5 numbers. The first 3 numbers define a product group used in pooling. In the example, only the first 3 numbers (= 5 digits) are printed. Factor Allowed indicates whether a chemical substances is allowed on a product or not.

Productgroup 10101 contains products Bean and Sperzieboon. Productgroup 10201 is split into two subgroups (according to factor Allowed): one group with Chicory, Endive, Cabbage Lettuce and Curly Lettuce and a group with Roodlof and Spinach. On the last two products the use of a chemical substance is not allowed.

In the example, the original mean and sigma are displayed together with the number of observations. The last three columns show the parameters μ and σ of the lognormal distribution together with the degrees of freedom after pooling. For Bean and Sperzieboon, a pooled μ (= -1.67) and σ (= 1.31*)* are used. For Chicory, Endive, Cabbage Lettuce and Curly Lettuce only sigma is pooled (σ = 1.47), the original means are maintained. For Roodlof (1 observation) the overall sigma (= 1.36) is used to estimated the variance.

Result of pooling:

			Original			Automatically pooling		
Labels	Levels	Allowed	Sigma	Mean	Nobs	Sigma	mean	Nobs
BEAN	10101	1	1.60	-1.17	8	1.31	-1.67	12
SPERZIEBOON	10101	1	0.75	-2.33	6	1.31	-1.67	12
CHICORY	10201	1	1.38	-2.69	4	1.47	-2.69	382
ROODLOF	10201	0	*	-2.3	1	1.36	-2.3	729
ENDIVE	10201	1	1.52	-0.91	92	1.47	-0.91	382
CABBAGE LETTUCE	10201	1	1.46	-1.44	286	1.47	-1.44	382
CURLY LETTUCE	10201	1	1.08	-2.14	4	1.47	-2.14	382
SPINACH	10201	0	1.18	-0.57	10	1.36	-0.57	729
BRUSSELS SPROUT	10301	0	1.14	-2.7	2	1.36	-2.7	729
CHINESE CABBAGE	10301	1	1.62	-2.32	21	1.62	-2.32	20
OXHEART/CONICAL	10301	0	*	-2.3	1	1.36	-2.3	729
ONION (SMALL)	10301	1	0.08	-1.66	2	0.08	-1.66	1
FENNEL	10301	1	0.16	-2.38	3	0.16	-2.38	2
POTATO	10401	0	0.62	0.19	2	0.59	0.19	50
WINTER CARROT	10401	0	0.62	-2.55	14	0.59	-2.55	50
CARROT	10401	0	0.54	-2.71	36	0.59	-2.71	50
RADISH	10401	1	1.52	-2.91	6	1.36	-2.91	729
CELERIAC	10401	0	1.31	-2.07	2	0.59	-2.07	50
GRAPE	10501	0	1.14	-1.06	25	1.14	-1.06	24
STRAWBERRY	10501	1	1.14	-1.57	169	1.14	-1.57	168
RASPBERRY	10501	1	1.73	-1.04	9	1.36	-1.04	729
BLACKBERRY	10501	1	1.15	-0.89	17	1.15	-0.89	16
BLUE BERRY	10501	1	1.83	-1.24	3	1.36	-1.24	729
CURRANT	10501	1	1.87	-0.62	30	1.87	-0.62	29
OTHER FRUIT, NUT	10601	0	*	-1.51	1	1.36	-1.51	729

End of example

8. References

- @Risk (1996). Advanced risk analysis for spreadsheets, Windows version. Pallisade Corporation, Newfield, NY, USA.
- Apex Software Corporation (1999). ComponentOne WebChart 7.0.10, Charting tool for browser-independent Web server applications
- Crossley, S.J. (2000). Joint FAO/WHO Geneva consultation – acute dietary intake methodology. *Food Additives and Contaminants*, 17: 557-562.
- de Boer, W.J. & van der Voet, H. (2000). Dietary risk assessment concerning acute exposure to residues and contaminants using summary data. Note WDB-2000-01, Centre for Biometry Wageningen, Wageningen.
- de Boer, W.J. & van der Voet, H. (2001). Dietary risk assessment concerning long-run average daily intakes. Note WDB-2001-01, Biometris, Wageningen.
- FAO/WHO (1997). Food consumption and exposure assessment of chemicals. Report of an FAO/WHO Consultation, Geneva, Switzerland. 10-14 February 1997.
- FAO (2002). Submission and evaluation of pesticide residues data for the estimation of maximum residue levels in food and feed, FAO, Rome.
- GenStat (2002). GenStat for Windows. Release 6.1. Sixth Edition. VSN International Ltd., Oxford.
- Hamey, P.Y. (2000). A practical application of probabilistic modelling in assessment of dietary exposure of fruit consumers to pesticide residues. *Food Additives and Contaminants*, 17: 601-610.
- IUPAC (1995). Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry* 67: 1699-1723.
- Nusser SM, Carriquiry AL, Dodd KW & WA Fuller, 1996. A semi-parametric transformation approach to estimating usual daily intake distributions. JASA 91(436): 1440-1449.
- Shimizu, K., and Crow, E.L. (eds). (1988). Lognormal distributions: theory and applications. Marcel Dekker, INC. New York.
- van der Voet, H., de Boer, W.J. & Keizer, L.C.P. (1999). Statistical instruments for dietary risk assessment concerning acute exposure to residues and contaminants. Report August 1999, Centre for Biometry Wageningen, Wageningen.
- van der Voet, H., de Boer, W.J. & Boon, P. (2001). Modelling exposure to pesticides. Note HVT-2001-03, Centre for Biometry Wageningen, Wageningen.
- van der Voet, H., de Boer, W.J., Boon, P.E., van Donkersgoed, G. & van Klaveren, J.D. (2004). MCRA, a GenStat program for Monte Carlo Risk Assessment, Release 3, Reference Guide. Report, Biometris and RIKILT, Wageningen.
- van Dooren, M. M. H. , Boeijen, I., van Klaveren, J. D. and van Donkersgoed G. (1995). Conversie van consumeerbare voedingsmiddelen naar primaire agrarische produkten. RIKILT-report. Wageningen, RIKILT-DLO (Report No: 95.17).

links:
Number of simulations
Pseudo-random sampling
Memory use
MCRA Program options: Output