MCRA

a GenStat program for

Monte Carlo Risk Analysis

Release 2002-02-27

*

Waldo J. de Boer
Hilko van der Voet

Biometris and RIKILT
Wageningen University and Research centre

2002

1. MCRA, an introduction

The program MCRA can be used for assessment of acute and chronic risks due to the intake of residues on food. The program is the result of an ongoing co-operation between RIKILT and Biometris since 1998. The program is written in the statistical package GenStat (2000), and therefore requires that GenStat is installed. An earlier version of the GenStat program as well as a simple version in @Risk (1996) have been described in Van der Voet *et al*. (1999). Most options in the program described here can also be found in Van der Voet *et al*. (2001) and De Boer & Van der Voet (2000, 2001).

1. 1.Overview

Before running a Monte Carlo risk assessment it is useful to give a short description of the program, options and data files to get you acquainted as quickly as possible.
A Monte Carlo risk assessment is performed with MCRA*ddmmyyyy*.gen with string *ddmmyyyy* representing the date of the current version of the program. MCRA, in short, is available as stand-alone version or as internet application. The stand-alone version requires that the specification form, library and all necessary datafiles are placed on a local disk drive. The internet version requires that the program MCRA, datafiles, library and supporting ActiveX-controls are installed on a server.

MCRA provides the following options:

- acute risk assessment
- chronic risk assessment
- parametric or non-parametric modelling of residue levels
- modelling of processing effects
- modelling of sample variability
- modelling of non-detects levels
- restrictions on age and/or days
- consumers only

To run a Monte Carlo risk assessment with a stand-alone version of MCRA the following files are needed:

Basic files	
MCRA*ddmmyyy*.gen	GenStat program
MCRAlib	procedure library
MCRA-input.xls	specification spreadsheet
individuals.lis	data on consumer characteristics
Standard data files	
####_con.lis	consumption data and processing codes
####_res.lis	residue concentration data
####_nos.lis	total number of samples
####_pro.lis	commodity code and labels
####_cmp.lis	residue label
Optional data files	
####_varf.xls	variability factors
####_proc.xls	processing factors
####_crtr.xls	data on percent crop treatment
####_histo.xls	histogram data
####_sum.xls	summary data
proccode.xls	processing codes and labels

Table 1: Requirements for a stand-alone version of MCRA. #### indicates a residue specific code.

For a risk assessment based on a model in its simplest form (see 4. 1), only the basic and standard data files (see Table 1) are required. To run a more extended version of the model, all optional data files are needed.
The following diagram illustrates the MCRA program in its working environment and shows what output is generated.

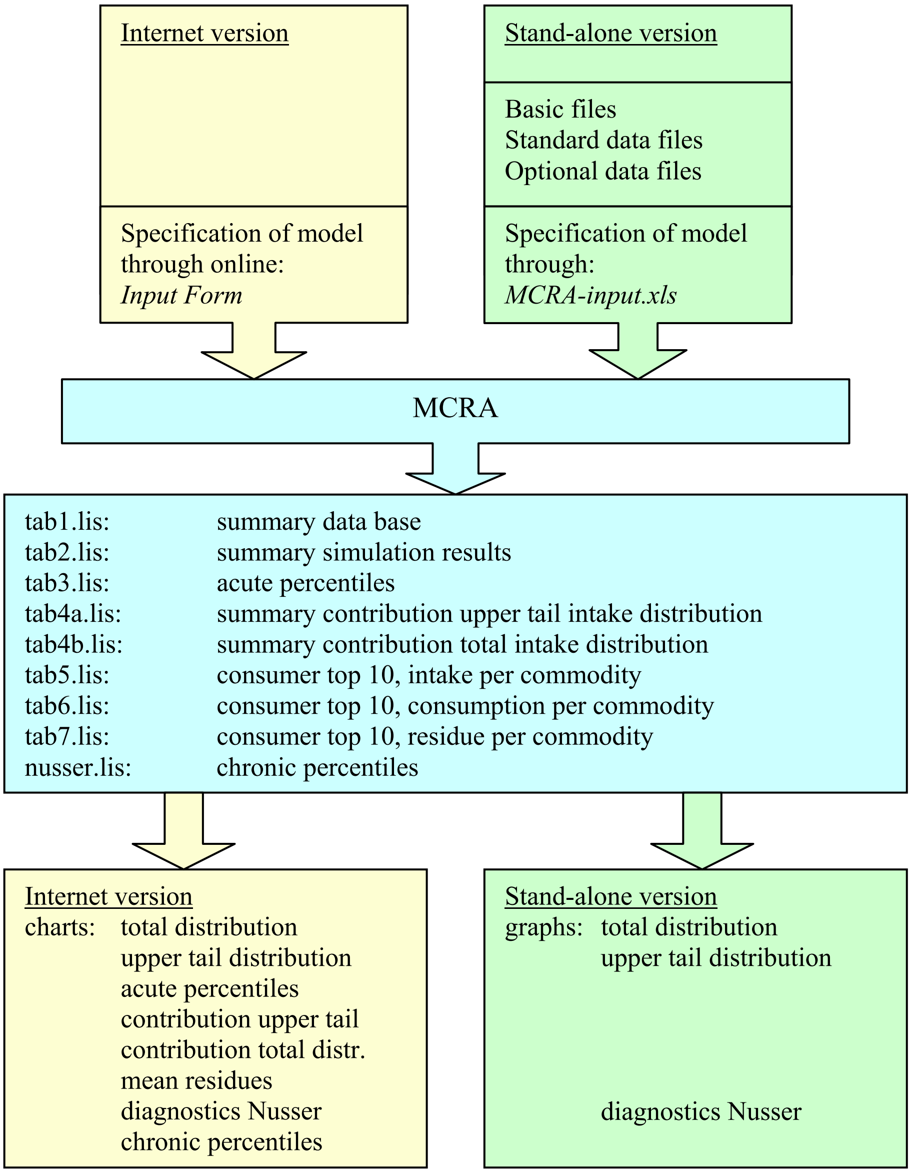

Internet version
Specification of model through online: *Input Form*
Stand-alone version
Basic files
Standard data files
Optional data files
Specification of model through: *MCRA-input.xls*
MCRA
tab1.lis: summary data base
tab2.lis: summary simulation results
tab3.lis: acute percentiles
tab4a.lis: summary contribution upper tail intake distribution
tab4b.lis: summary contribution total intake distribution
tab5.lis: consumer top 10, intake per commodity
tab6.lis: consumer top 10, consumption per commodity
tab7.lis: consumer top 10, residue per commodity
nusser.lis: chronic percentiles
Internet version
charts: total distribution
upper tail distribution
acute percentiles
contribution upper tail
contribution total distr.
mean residues
diagnostics Nusser
chronic percentiles
Stand-alone version
graphs: total distribution
upper tail distribution
diagnostics Nusser

Figure 1: Environment of the stand-alone version and the internet application of MCRA

Figure 1 shows that for the stand-alone version a specification spreadsheet MCRA-input.xls is needed to specify the model. For the internet application, the user is requested to fill out on-line an input form and submit this to the server. In both applications, basically, the same program MCRA is running. In the internet version, graphs are customised to communicate with an ActiveX-aware browser. The client-side ActiveX control comprises a ComponentOne Chart control[1]. Clients are allowed to modify the chart they view on the web page. The stand-alone version generates three graphs which can be printed or saved as .emf. Both versions generate tables with results on percentiles, summaries of the total and upper tail of the intake distribution, contributions of commodities and consumer top 10's.

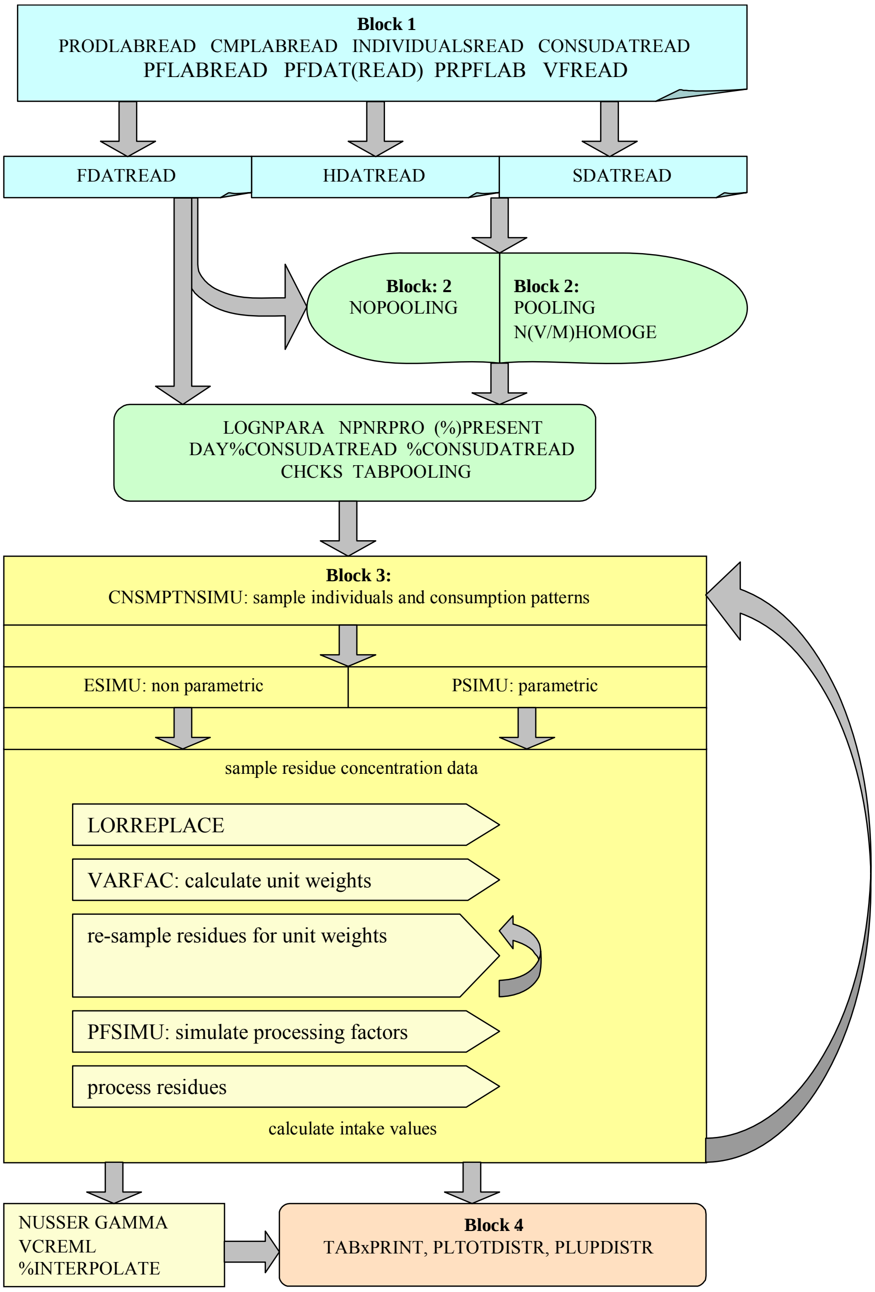
[1] ComponentOne WebChart 7.0, Charting tool for browser-independent Web server applications.

1. 2.Flow diagram of MCRA

After the model is set (specification spreadsheet or on-line input form), a Monte Carlo risk analysis is performed. The program MCRA is composed of a set of procedures which may be arranged into four main blocks. All procedures are shortly described in 5. 1. The main tasks of block 1 to 4 are:

1. reading of data (residue concentration data, consumption data, consumer characteristics, processing factors, variability factors, percent crop treatment)
2. pre-processing of datastructures (age and/or day restrictions, consumers only, processing or not, variability factors, estimation of parameters for a parametric model), determining number of loops, chunksize, etc.
3. simulation of exposure values (parametric, non-parametric)
4. generating output (intake distribution, contribution to upper tail, characteristics of consumers with the highest intake, etc…)

In Figure 2, a schematic outline of the blocks with the most relevant procedures is presented.

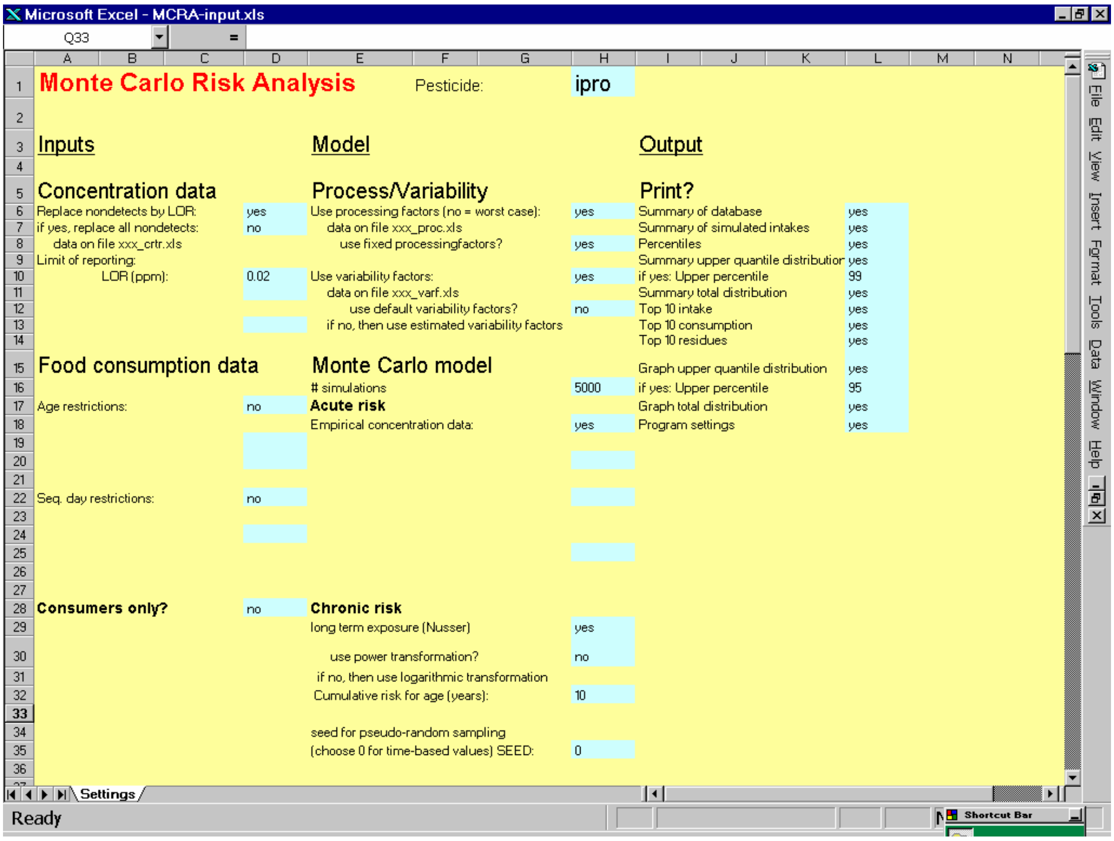


Block 1
PRODLABREAD CMPLABREAD INDIVIDUALSREAD CONSUDATREAD
PFLABREAD PFDAT(READ) PRPFLAB VFREAD
FDATREAD
HDATREAD
SDATREAD
Block: 2
NOPOOLING
Block 2:
POOLING
N(V/M)HOMOGE
LOGNPARA NPNRPRO (%)PRESENT
DAY%CONSUDATREAD %CONSUDATREAD
CHCKS TABPOOLING
Block 3:
CNSMPTNSIMU: sample individuals and consumption patterns
ESIMU: non parametric
PSIMU: parametric
sample residue concentration data
LORREPLACE
VARFAC: calculate unit weights
re-sample residues for unit weights
PFSIMU: simulate processing factors
process residues
calculate intake values
NUSSER GAMMA
VCREML
%INTERPOLATE
Block 4
TABxPRINT, PLTOTDISTR, PLUPDISTR

Figure 2: Schematic outline of MCRA

2. MCRA stand-alone version, documentation and specification of input

In this section, a description is given of the stand-alone version of the program MCRA.
Before running a Monte Carlo risk analysis with MCRA, *inputs*, the *model* and requested *output* are specified in a specification spreadsheet MCRA-input.xls (Figure 3).



Microsoft Excel - MCRA-input.xls
Q33
=
File Edit View Insert Format Tools Data Window Help
Settings
Ready
Shortcut Bar

	A–C	D	E–G	H	I–K	L
1	Monte Carlo Risk Analysis		Pesticide:	ipro		
2						
3	Inputs		Model		Output	
4						
5	Concentration data		Process/Variability		Print?	
6	Replace nondetects by LOR:	yes	Use processing factors (no = worst case):	yes	Summary of database	yes
7	if yes, replace all nondetects:	no	data on file xxx_proc.xls		Summary of simulated intakes	yes
8	data on file xxx_crtr.xls		use fixed processingfactors?	yes	Percentiles	yes
9	Limit of reporting:				Summary upper quantile distributior	yes
10	LOR (ppm):	0.02	Use variability factors:	yes	if yes: Upper percentile	99
11			data on file xxx_varf.xls		Summary total distribution	yes
12			use default variability factors?	no	Top 10 intake	yes
13			if no, then use estimated variability factors		Top 10 consumption	yes
14					Top 10 residues	yes
15	Food consumption data		Monte Carlo model		Graph upper quantile distribution	yes
16			# simulations	5000	if yes: Upper percentile	95
17	Age restrictions:	no	**Acute risk**		Graph total distribution	yes
18			Empirical concentration data:	yes	Program settings	yes
19						
20						
21						
22	Seq. day restrictions:	no				
23						
24						
25						
26						
27						
28	**Consumers only?**	no	**Chronic risk**			
29			long term exposure (Nusser)	yes		
30			use power transformation?	no		
31			if no, then use logarithmic transformation			
32			Cumulative risk for age (years):	10		
33						
34			seed for pseudo-random sampling			
35			(choose 0 for time-based values) SEED:	0		
36						

Figure 3: MCRA stand-alone version: input form MCRA-input.xls

2. 1.Inputs

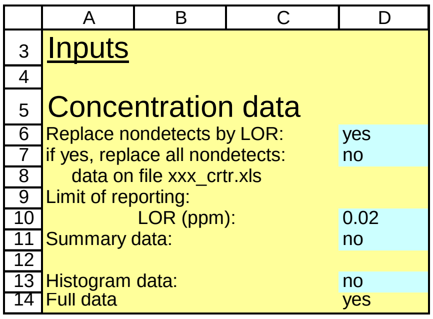
First of all, the name of the residue is specified using a three or four letter code 'xxxx' which refers to residue specific files (see 5.3.2, 5.3.3). For example, Iprodione may be specified using 'ipro' which refers to e.g. residue concentration data file ipro_res.lis. All inputs concerning concentration and food consumption data are specified in the **Inputs** part of the form by setting the cell content of the relevant option to *yes* or *no*.

2.1.1.Percent crop treated

In Figure 3, non-detects are replaced (*yes*) by the LOR (0.02 ppm) (see 4.2.4). and replacement is based on the percent crop treated data (replace all non-detects = *no*) on file ipro_crtr.xls (see 5.3.3).

2.1.2.Data

In the example, the simulation is based on empirical concentration data (*yes*) (see 5.3.2), so options summary or histogram data are not relevant (options are suppressed). When option empirical concentration data is set to *no,* a parametric version of the model is running based on either full data, summary data or histogram data. Usually, a parametric version based on full data is specified by setting options summary data and histogram data to *no* (Figure 4).

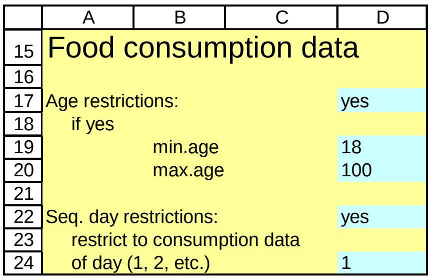


	A	B	C	D
3	Inputs			
4				
5	Concentration data			
6	Replace nondetects by LOR:			yes
7	if yes, replace all nondetects:			no
8	data on file xxx_crtr.xls			
9	Limit of reporting:			
10		LOR (ppm):		0.02
11	Summary data:			no
12				
13	Histogram data:			no
14	Full data			yes

Figure 4: Parametric simulation based on full data

2.1.3.Restrictions

In Figure 3, age and days are unrestricted, e.g. all consumers and all days are taken to sample from. Note that option Consumers only is set to *no*, meaning that all consumers are taken to sample from irrespective of their actual consumption. When options age or day restrictions are specified, the cell content is set to *yes* and the minimum and maximum age (years) and day-number is filled in (Figure 5). When option Consumers only is set to *yes,* in the upper left corner of the spreadsheet the message '**Consumers Only**' is shown.



	A	B	C	D
15	Food consumption data			
16				
17	Age restrictions:			yes
18	if yes			
19		min.age		18
20		max.age		100
21				
22	Seq. day restrictions:			yes
23	restrict to consumption data			
24	of day (1, 2, etc.)			1

Figure 5: Restrictions for age and day

2. 2.Model

Under **Model**, the exposure model is specified.

2.2.1.Processing factors

The use of processing factors (see 4.2.2) is indicated by setting the cell contents to *yes*. Note that *no* is a worst case scenario ($f_k = f_{k,upp} = 1$). If commodities are processed, processing factors are fixed ($f_k = f_{k,upp}$) or random e.g. sampled from a normal distribution with parameters μ and σ for mean and standard deviation based on transformed values of $f_{k,upp}$ and $f_{k,nom}$. The transformation should be specified (logarithm or logit). Processing factors are read from the file xxxx_proc.xls and codes and labels from proccode.xls (see 5.3.3). In Figure 3, fixed processing factors are specified (*yes*). To process simultaneously some commodities using fixed factors and others using factors based on a distribution, set option use fixed processing factors to *no*, e.g. processing is based on a distribution. 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 in xxxx_proc.xls a complete list of processing factors on all commodities. Missing $f_{k,nom}$ and $f_{k,upp}$ are, by default, replaced by the value 1. When processing factors are based on a distribution, sampled values are not necessarily < 1. Depending on the values of $f_{k,upp}$ and $f_{k,nom}$, occasionally values above 1 occur. To force $f_k < 1$ a logit transformation is specified (see also 5. 3 and Figure 12).

2.2.2. Variability factors

In Figure 3, option variability factors (see 4.2.3) is set to *yes*. Estimated factors (see 4.2.3.1) are found in file xxxx_varf.xls (see 5.3.3). When sample variability is based on default factors (see 4.2.3.2) variability factors are according to Table 2 in 4.2.3.

2.2.3. Number of simulations

The number of simulations (iterations) specified in Figure 3 is set to 5000. This is the number of consumers that is randomly drawn from the consumer data base. There are no upper limitations to this number. Note: a Monte Carlo risk assessment incorporating both processing and variability factors will increase simulation time considerably.

2.2.4. Acute risk

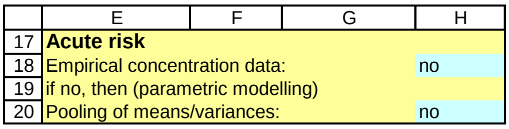
Residue concentration data in the various food commodities are independent and therefore can be modelled by univariate distributions. Two approaches are implemented.

2.2.4.1 Non parametric approach

In the non-parametric approach, option empirical concentration data = *yes (*see Figure 3), residue values are sampled at random from the available data and combined with food consumption data to generate the intake distribution of exposure values. Note: when option empirical concentration data is specified, all cell information concerning parametric modelling is ignored.

2.2.4.2Parametric approach

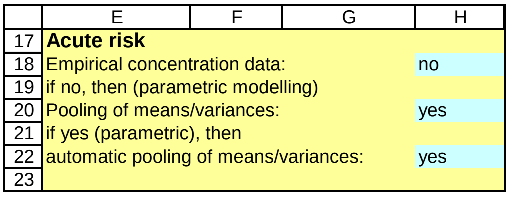
In the parametric approach, option empirical concentration data = *no*, residue concentrations per food commodity are sampled from parametric distributions based on full data, histogram data or summary data (see 2.1.2). When parametric modelling is specified, pooling of means and variances is optional (see 4.5.2, Figure 14). In Figure 6, parametric modelling and no pooling of means and variances within productgroups is specified. Note that specifying option no pooling is only a choice when μ's and σ's of residue concentration data are present for all commodities. If some parameters are missing, a warning message is printed by the program. MCRA should be rerun with option pooling set to *yes*.



	E	F	G	H
17	**Acute risk**			
18	Empirical concentration data:			no
19	if no, then (parametric modelling)			
20	Pooling of means/variances:			no

Figure 6: Parametric modelling: no pooling of means/variances

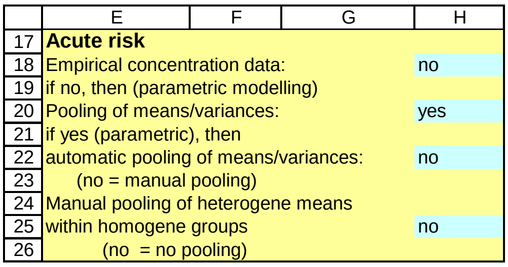
In case of missing μ's and/or σ's the simulation is preceded by a pooling step to obtain all necessary parameters. In Figure 7, option automatic pooling is specified (*yes*) and the Monte Carlo risk analysis cycles to the end without any interrupts.



	E	F	G	H
17	**Acute risk**			
18	Empirical concentration data:			no
19	if no, then (parametric modelling)			
20	Pooling of means/variances:			yes
21	if yes (parametric), then			
22	automatic pooling of means/variances:			yes
23				

Figure 7: Parametric modelling: automatic pooling of means/variances

If option manual pooling is set (Figure 8), the user is guided through the pooling process through the use of pop-up menu's (not for web application).



	E	F	G	H
17	**Acute risk**			
18	Empirical concentration data:			no
19	if no, then (parametric modelling)			
20	Pooling of means/variances:			yes
21	if yes (parametric), then			
22	automatic pooling of means/variances:			no
23	(no = manual pooling)			
24	Manual pooling of heterogene means			
25	within homogene groups			no
26	(no = no pooling)			

Figure 8: Parametric modelling: manual pooling of means/variances and no manual pooling of heterogene means within homogene groups

2.2.5.Chronic risk

In Figure 3, option long term exposure according to Nusser (see 4. 4) is set to *yes*. Two transformations may be specified, a power or a log transformation. Usually, a power transformation is satisfactory. In addition, the cumulative risk for a certain age (in years) is specified. Chronic and acute risk assessment may be performed simultaneously, that is in the same run. Both results are reported. However, a chronic risk assessment is only possible when the total number of non-detects is below 1%. When the number of non-detects is higher, a warning message is printed. Note that a chronic risk assessment is time-consuming. When options Consumers only and replace all non-detects are both specified, a chronic risk assessment is always completed succesfully. Specifying option replace all non-detects only, is in most cases not sufficient to run succesfully a chronic risk assessment.

2. 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 are not changed.

2. 4.Output

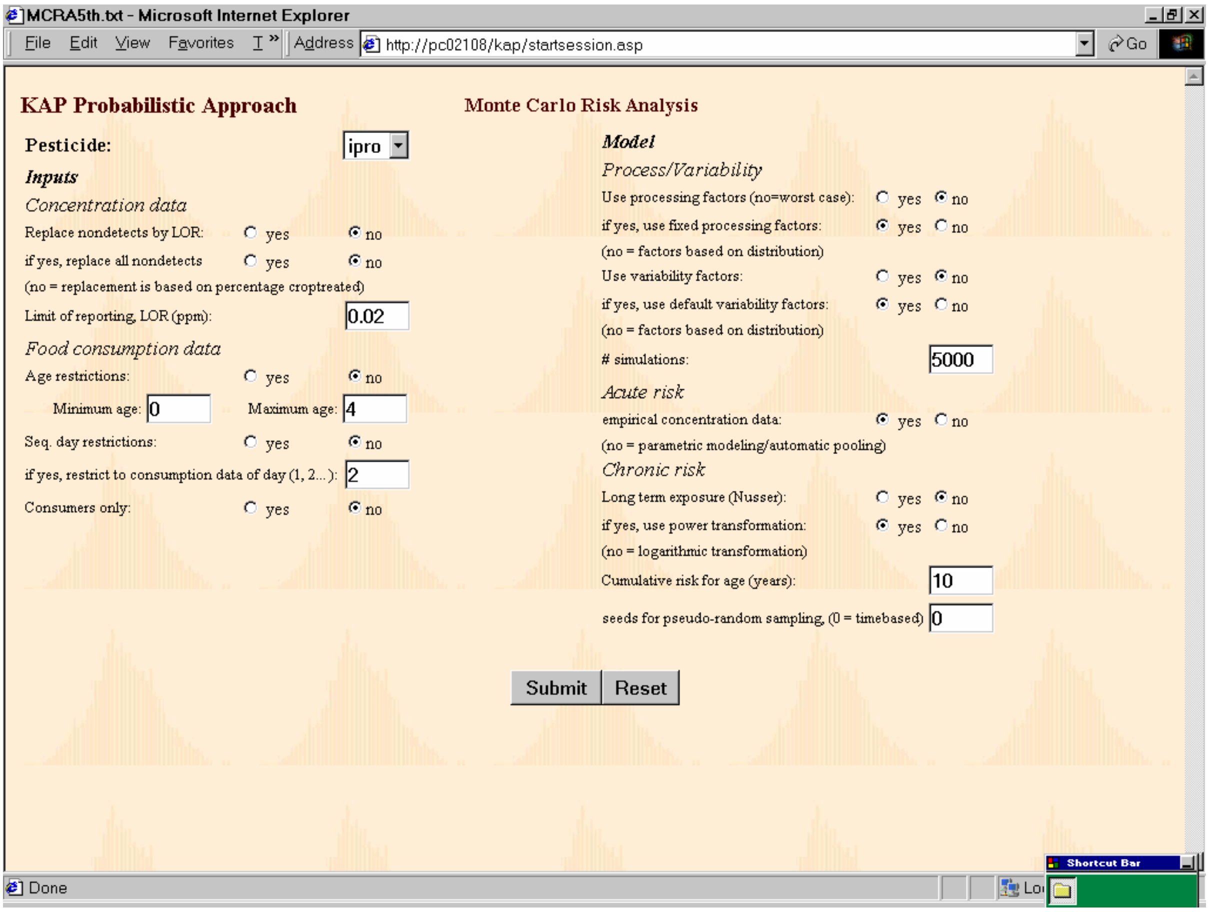
Requested output is specified in the **Output** form. In Figure 3, all output is set to *yes*. To print a summary of the upper quantile, the upper value may be specified. The default value is 95%. Two kinds of graphs are available: the upper tail of the intake distribution for a specified quantile and the total intake distribution of all positive values. Note that the value of the upper quantile of the summary and upper tail of the intake distribution are set independently. All output is written to files tab*x*.lis with 1, 2, 3, 4a, 4b, 5, 6 and 7 replacing *x*.

For chronic risk assessment, a diagnostic graph is plotted and output is written to file nusser.lis.

3. MCRA internet application

The MCRA internet application is basically the same as the stand-alone version and the diagram presented in Figure 2 applies. The major differences occur in block 4 where calls to subroutines are made in order to encrypt generated output to allow communication with ActiveX–aware browsers.
In Figure 10, a diagram of the internet version in its environment and related files is presented. The MCRA homepage is found at mcra.html. This HTML-page calls startsession.asp and input.htm to start an MCRA-session. By entering the on-line input form (see Figure 9), the website is locked to other users and the Monte Carlo model is specified. After filling out the form, all settings are passed to the server by pressing the submit button. Repeatedly pressing the submit-button during processing generates a warning. The servers starts MCRA, performs a risk assessment and after completing the analysis returns an output window (reference.htm and outputprint.htm, see Figure 11). Charts and tables are viewed on the web-page and the client is allowed to manipulate the charts using the mouse buttons.

Each session should be ended by pressing the End Session-icon. This unlocks the website to other users. When the website is currently in use (session locked) and a second user tries to enter it, a warning message is generated. A session is unlocked calling restartsession.asp. A call to quick.asp surpasses the analysis and output that is generated in an earlier session (see Figure 11) is viewed immediately



Address: http://pc02108/kap/startsession.asp

KAP Probabilistic Approach

Monte Carlo Risk Analysis

Pesticide: ipro

Inputs

Concentration data

Replace nondetects by LOR: ○ yes ● no

if yes, replace all nondetects ○ yes ● no

(no = replacement is based on percentage croptreated)

Limit of reporting, LOR (ppm): 0.02

Food consumption data

Age restrictions: ○ yes ● no

Minimum age: 0 Maximum age: 4

Seq. day restrictions: ○ yes ● no

if yes, restrict to consumption data of day (1, 2...): 2

Consumers only: ○ yes ● no

Model

Process/Variability

Use processing factors (no=worst case): ○ yes ● no

if yes, use fixed processing factors: ● yes ○ no

(no = factors based on distribution)

Use variability factors: ○ yes ● no

if yes, use default variability factors: ● yes ○ no

(no = factors based on distribution)

simulations: 5000

Acute risk

empirical concentration data: ● yes ○ no

(no = parametric modeling/automatic pooling)

Chronic risk

Long term exposure (Nusser): ○ yes ● no

if yes, use power transformation: ● yes ○ no

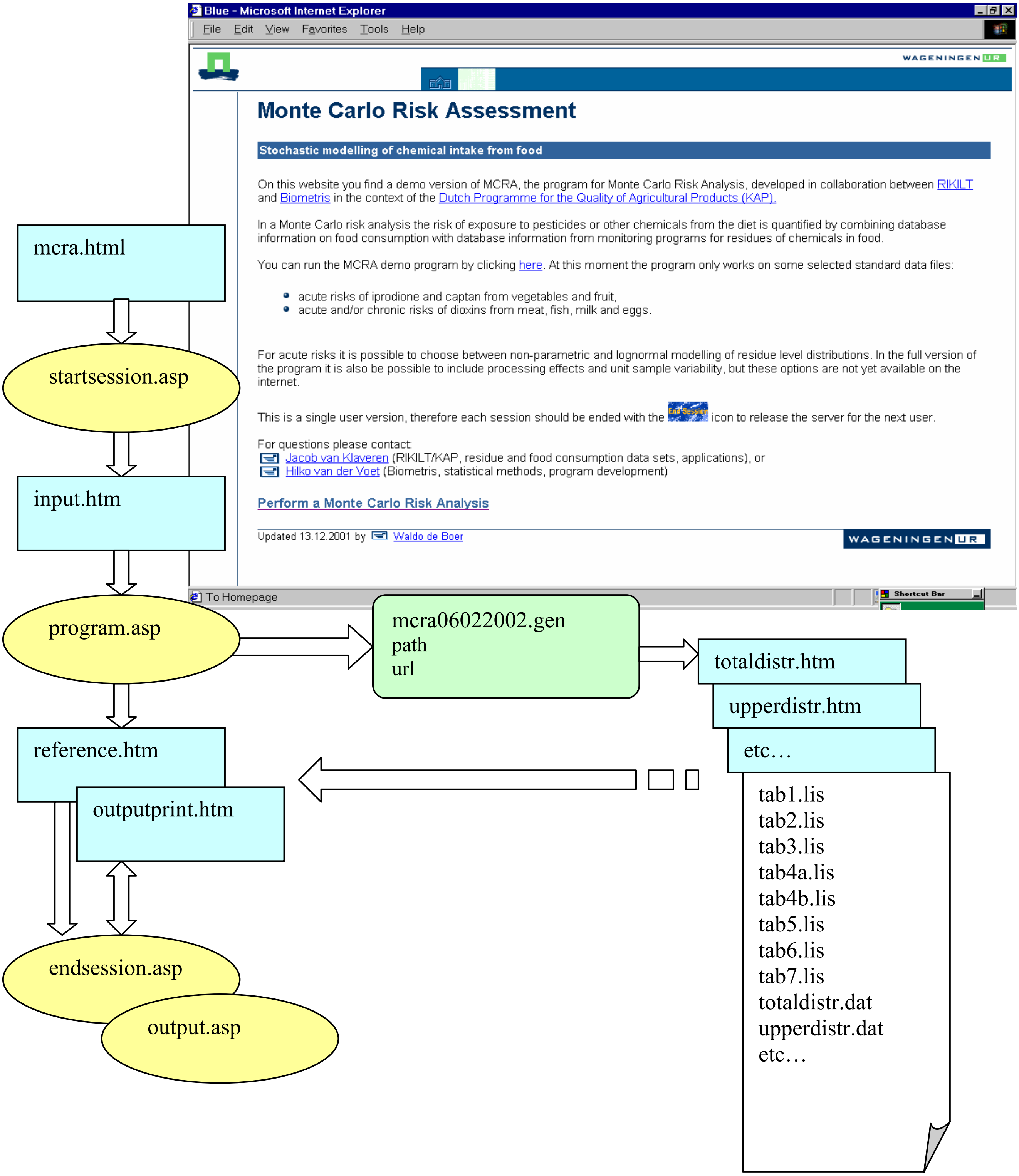
(no = logarithmic transformation)

Cumulative risk for age (years): 10

seeds for pseudo-random sampling, (0 = timebased) 0

Submit Reset

Figure 9: MCRA internet application: on-line input form



Blue - Microsoft Internet Explorer

Monte Carlo Risk Assessment

Stochastic modelling of chemical intake from food

On this website you find a demo version of MCRA, the program for Monte Carlo Risk Analysis, developed in collaboration between RIKILT and Biometris in the context of the Dutch Programme for the Quality of Agricultural Products (KAP).

In a Monte Carlo risk analysis the risk of exposure to pesticides or other chemicals from the diet is quantified by combining database information on food consumption with database information from monitoring programs for residues of chemicals in food.

You can run the MCRA demo program by clicking here. At this moment the program only works on some selected standard data files:

- acute risks of iprodione and captan from vegetables and fruit,
- acute and/or chronic risks of dioxins from meat, fish, milk and eggs.

For acute risks it is possible to choose between non-parametric and lognormal modelling of residue level distributions. In the full version of the program it is also be possible to include processing effects and unit sample variability, but these options are not yet available on the internet.

This is a single user version, therefore each session should be ended with the End Session icon to release the server for the next user.

For questions please contact:
Jacob van Klaveren (RIKILT/KAP, residue and food consumption data sets, applications), or
Hilko van der Voet (Biometris, statistical methods, program development)

Perform a Monte Carlo Risk Analysis

Updated 13.12.2001 by Waldo de Boer

Figure 10: MCRA internet application: environment and files

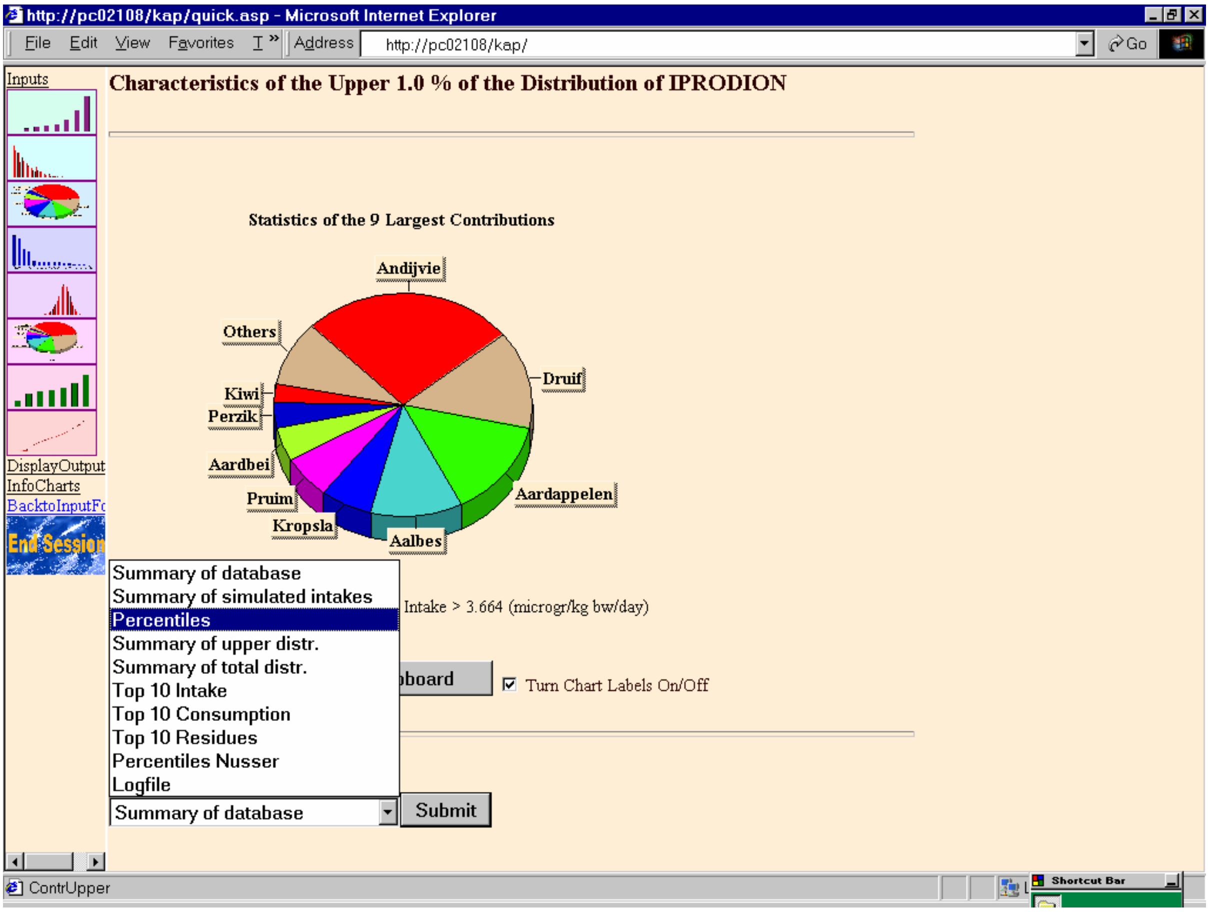


Figure 11:MCRA internet application: piechart of contributions to upper tail of intake distribution

When new data become available files are copied to the directory on the server. New residue codes are added with HTML-code `<option>####</option>` in file input.htm. To run a new version of MCRA, change the version number in program.asp.

4. Model description

4. 1.Introduction

This chapter describes a stochastic (or Monte Carlo) model for the assessment of acute risks due to the intake of pesticides from food. The model combines food consumption survey data and pesticide concentration data from monitoring programs. The model allows for effects of food processing between monitoring and ingestion, it can model unit variability either from available data or using default assumptions, and it uses information on limit of reporting and percent crop treated to check whether non-detects present a source of uncertainty.
The basic model is:

$$y_{ij} = \frac{\sum_{k=1}^{p} x_{ijk}\ c_{ijk}}{w_i}$$

where y_{ij} is the intake by individual i on day j (in µg pesticide per kg body weight), x_{ijk} is the consumption by individual i on day j of food commodity k (in g), c_{ijk} is the concentration of the pesticide in commodity k eaten by individual i on day j (in mg/kg, 'ppm'), and w_i is the body weight of individual i (in kg). Finally, p is the number of food commodities accounted for in the model. Note that the definition of 'commodity' should be flexible: it may represent a raw agricultural commodity (RAC), e.g. 'apple', but the user of the model should have the option to discern processing-related subdivisions, e.g. 'apple, peeled' or 'apple, juiced'.

In the stochastic model the quantities x_{ijk}, w_i and c_{ijk} are assumed to arise from probability distributions for individual food consumption and weight, $p(x_1 ,...,x_p,w),$ and for pesticide concentrations in each food commodity, $p_k(c)$. In principle these probability distributions may be parametric (e.g. completely defined by the specification of some parameter values) or empirical (e.g. only implicitly and roughly defined by the availability of a representative sample).

We restrict our attention to the basic model of the empirical distribution of food consumption and body weight as is provided by the national food consumption surveys. A recipe data base has been used to convert the amounts of food as consumed to amounts of commodities ($x_1 ,...,x_p$) of raw agricultural products which are used in the model. For example, from the Dutch Food Consumption Survey 1997 food consumption patterns ($x_1,...,x_p$), body weight (w) and age (a) are available for 6250 individual persons on 2 consecutive days. Depending on the problem, Monte Carlo samples may be drawn from the complete data base, from a day or age-restricted subset or from consumers only.

Residue concentration data are available from the KAP-data base (Oracle), which stores annually more than 200.000 records of measurements originating from food monitoring programs for meat, fish, dairy products, vegetables and fruit.

Given these probability distributions (or estimates thereof) Monte Carlo simulations can be used to generate an estimate of the probability distribution $p(y_{ij})$ to assess acute risks by intake of the pesticide (see 4. 2). When dietary components are consumed on a nearly daily basis, intake values y_{ij} may be used to estimate the probability distribution $p(y_{i.})$ for chronic risk assessment purposes (see 4. 4).

4. 2.Modelling of pesticide concentrations in consumed food

4.2.1.Distributional assumptions

Residue concentrations in the various food commodities are independent and therefore can be modelled by univariate distributions.

4.2.1.1Non-parametric modelling of residue levels
In the non-parametric approach, residue values are sampled at random from the available data and combined with the consumption data to generate a new distribution of exposure values. To assess the risk-exposure, percentiles of the exposure distribution are estimated.

4.2.1.2Parametric modelling of residue levels
In the parametric approach, residue concentrations per food commodity are sampled from parametric distributions. A special feature of residue data is that the large majority of measured concentrations (often more than 80%) is recorded as zero (non-detects). These values may correspond to true zero concentrations (for example because the substance is never used in the specific product), or they may correspond to low concentrations which are below a pre-established reporting limit (LOR). In any case, the residue concentration distribution is very skew, with a large spike at zero and an extended tail to higher values. For statistical modelling a two-step procedure is chosen. First, the presence of a concentration ≥LOR on food products is modelled with a binomial distribution with a parameter *p* representing the probability of a reported residue level. Probability *p* depends on the pesticide and the product and is estimated as the fraction of detects. Secondly, the non-zero residues are modelled with a parametric distribution. After consideration of several possibilities using the program *BestFit*, the lognormal distribution has been selected as being both theoretically sensible and practically useful. The parameters μ and ● are the mean and standard deviation of the log-transformed non-zero residue concentrations.
In the basic model (see 4. 1)

$$c_{ijk} = I_{ijk} \cdot cpos_{ijk}$$

with I_{ijk} indicating whether a residue concentration is sampled (I_{ijk}=1) or not (I_{ijk}=0), and $cpos_{ijk}$ the residue concentration in the subpopulation of positive values. The probability of I_{ijk} being 1 or 0 depends on the number of detects found for commodity *k* and I_{ijk} is sampled separately for each individual *i* on occasion *j*.

4.2.2.Modelling of processing effects

Concentrations in the consumed food may be different from concentrations in the product as measured in monitoring programs (typically raw product) due to processing, such as peeling, washing, cooking etc.
In general we assume the model:

$$cpos_{ijk} = f_k \cdot cr_{ijk}$$

where cr_{ijk} is the concentration in the raw product, and where f_k is a factor for a specific combination *k* of RAC and processing. Values will typically be between 0 and 1, although occasionally the processing factor may also be >1.
The user of the model will have to specify processing factors for each commodity *k* as defined in the food consumption data base. For this purpose it is advised to maintain a data base of processing factors, indexed by pesticide, RAC and processing type (e.g. washing, peeling, other processing). Before running the model it may then be necessary to specify how the necessary processing factors are derived from the data base entries and/or other information. Example: if there are no processing

factors known for captan in pears, it may be decided to use the corresponding factors for apples instead.

Often the information will be of limited quality, and this may be entered in the Monte Carlo modelling by specification of uncertainties. A practical proposal is to specify for each processing factor two values:

1. $f_{k,nom}$: the nominal value, typically some sort of mean from an experimental study
2. $f_{k,upp}$: an upper 95% confidence limit, which typically will be set by an expert (even if statistical information on variability of the factor is available, there will often be uncertainty due to the appropriateness of the processing study for the population of the risk analysis). The upper limit should be such that experts will easily agree that it is not set too low.

A typical data base entry might thus read:

pesticide	RAC	processing	$f_{k,nom}$	$f_{k,upp}$
captan	apple	washing	0.5	0.7

and, confronted with the need to have processing factors for pears in a specific risk analysis, an expert may decide upon:

pesticide	RAC	processing	$f_{k,nom}$	$f_{k,upp}$
captan	pear	washing	0.5	0.8

In the Monte Carlo modelling, processing factors can be used in either of three ways (for each commodity k to be chosen by the user):

1. (no processing factor) Just take $f_k = 1$. This is in most (though not all) cases a worst-case assumption. No data on processing are needed and therefore this route is useful in a first tier approach.
2. (fixed value) Use $f_k = f_{k,upp}$. Available information on specific processing effects is used, although still in a cautionary way (in accordance with the precautionary principle). Note that $f_{k,nom}$ values need not to be specified.
3. (distribution based) Sample f_k using a normal distribution. Log or logit transformed values of $f_{k,nom}$ and $f_{k,upp}$ are used to define the first two moments of the normal distribution. Two situations are distinguished depending on the type of transformation.
   a) The logarithms of $f_{k,nom}$ and $f_{k,upp}$ are equated to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution thus is specified by a mean $\ln(f_{k,nom})$ and a standard deviation $\{\ln(f_{k,upp}) - \ln(f_{k,nom})\}/1.645$. Values are drawn from this distribution in the Monte Carlo simulations. Processing factors f_k will be nonnegative. Note: $f_{k,upp}$ and $f_{k,nom}$ values equal to 0 are replaced by a low user-specified value (e.g. 0.01); this is useful computationally to avoid problems with logarithms.
   b) The logits of $f_{k,nom}$ and $f_{k,upp}$ are equated to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution thus is specified by a mean $\text{logit}(f_{k,nom})$ and a standard deviation $\{\text{logit}(f_{k,upp}) - \text{logit}(f_{k,nom})\}/1.645$. Values are drawn from this distribution in the Monte Carlo simulations. Processing factors f_k will be between 0 and 1. Note: $f_{k,upp}$ and $f_{k,nom}$ values equal to 0 and 1 are replaced by user-specified values (e.g. 0.01 and 0.99); this is useful computationally to avoid problems with logits.

   The user should keep in mind that, in case of a lognormal distribution, $f_{k,nom}$ defines the median, while $f_{k,upp}$ quantifies skewness. The same holds for the logistic. Usually, a logarithm will be the standard transformation, but for very skew distributions (see Figure 12) occasionally values above 1 are sampled (upper row, 1^{rst}, 3rd and 5th plot). A logit transformation should be considered.

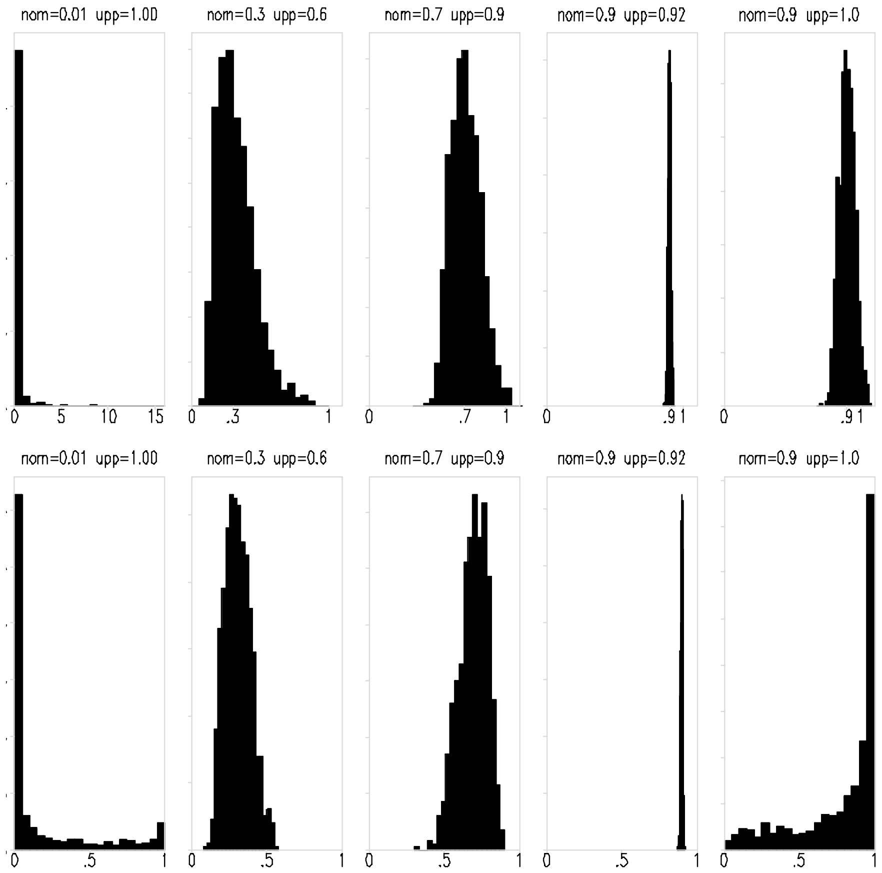


Figure 12: Lognormal (upper row) and logistic (lower row) distributions for various values of $f_{k,nom}$ (=nom) and $f_{k,upp}$ (=upp)

4.2.3.Modelling of sample variability

Monitoring measurements cm_k are typically made on homogenised composite samples. Such a composite sample is composed of nu_k units with nominal unit weight wu_k each. The weight of a composite sample is therefore $wm_k = nu_k \times wu_k$. This weight is often larger than a consumer portion, e.g. a typical composite sample of 20 sweet peppers weighs 3.2 kg, whereas daily consumer portion weights in the Dutch Food Consumption Survey 1997 ranged from 0.08 g to 458 g.

How should monitoring data be used to estimate the raw commodity concentration levels cr_{ijk} in probabilistic acute risk assessment? Although the mean level of cm_k may be a fair estimate of the mean level of cr_{ijk}, the variability of cm_k is not appropriate to estimate the variability of cr_{ijk}. In smaller portions more extreme values may occur more readily, and thus acute risks may be higher than would follow from a direct use of the composite sample data.

In non-probabilistic modelling of acute risks the unit-to-unit variability has been addressed by the definition of a variability factor *v*, which is the ratio between a ‘high’ value (maximum or high percentile, not clearly defined) and the mean (or median) value of residue levels of individual units in a batch of units. Values for *v* can be obtained by measuring individual units. In practice such data are mostly available from field trials, although for risk assessment it would be more appropriate to calculate unit variability in monitoring samples. We therefore advise to use field trial values for *v* only when monitoring sample values for *v* are not available.

If there are insufficient data from measurements on individual units, the FAO/WHO Expert Consultation (FAO/WHO, 1997; Crossley, 2000) recommended to assume (conservatively) that all of the residue in a composite sample would be present in one of the units. Under this assumption *v* equals the number of units in the composite sample. If Codex sampling protocols are used, then the number of units per composite sample is 5 for large crops (unit weights > 250 g) and 10 for medium crops (unit weights 25-250 g). For small crops (< 25 g) a variability factor $v = 1$ was recommended. More recently, it has been proposed to replace the default value 10 with 7. For commodities which are processed in large batches, e.g. juicing, a variability factor $v = 1$ is proposed. To summarise:

unit weight, *wu*	variability factor, *v*
< 25 g	1
25 -250 g	7
> 250 g	5
juicing, marmalade/jam, sauce/puree	1

Table 2: Default variability factors

The lognormal distribution is considered as an appropriate model for many empirical positive residue level distributions. We will also assume a lognormal distribution for unit residue concentrations. Let this distribution be characterised by μ and σ, which are the mean and standard deviation of the log-transformed concentrations *lc*. The variability factor *v* can be converted into the standard deviation σ (see below). Upper-tail percentiles of this lognormal distribution are influenced in two opposing ways by the magnitude of the variability factor:

1. Because of more spread, the percentiles $c_q = e^{\mu + z_q \sigma}$ increase with σ relative to the median e^{μ} (z_q is the 100*q* percent point of the standard normal distribution);
2. However, the median e^{μ} decreases with σ relative to the expected value (mean) *E(c)* according to: $e^{\mu} = E(c) \cdot e^{-\frac{1}{2}\sigma^2}$.

The composite sample measurements cm_k are estimates of *E(c)*. Percentiles of the unit distribution for a batch with expected value (mean) cm_k are therefore equal to:

$$c_q = cm_k \cdot e^{-\frac{1}{2}\sigma^2 + z_q \sigma}$$

The combined influence in this simple case is that c_q increases with σ for high percentiles ($z_q > \sigma$), but decreases with σ for relatively low percentiles ($z_q < \sigma$).

The following approaches to the modelling of sample variability should be incorporated in the model:
1. Use estimated values of *v*
2. Use default (conservative) values of *v*

These approaches are described in more detail below. In both cases we assume that the majority of residue level data is from a representative sample of composite samples. Alternatively, surveys may be available in which residue level data have been collected for large amounts of individual units. These data can be used directly, although care is needed to reflect the structure of between-batch/within-batch variability (Hamey, 2000).

4.2.3.1Use estimated values of the variability factor *v*

In this approach it is essential to discern between-batch variability from within-batch variability. Typically, variability factors are calculated for units from one field trial or commercial batch, although such batches are not always clearly defined. Variability factors describe the variability between units within batches. The proposed approach is as follows:

- If a value for *v* is available that can be interpreted as the ratio between the 97.5 percentile and the median of a lognormal distribution of unit residue levels in one batch, then, with μ and σ representing the mean and standard deviation of the log-transformed concentrations *lc* , we have:

$$v = \frac{e^{\mu + 2\sigma}}{e^{\mu}} = e^{2\sigma}$$

or $\sigma = \frac{1}{2}\ln(v)$

- For each iteration i in the Monte Carlo simulation, obtain for each commodity k a simulated intake x_{ik} , and a simulated composite sample residue concentration cm_{ik} .
- Calculate the number of unit intakes nux_{ik} in x_{ik} (round upwards) and set weights w_{ikl} equal to wu_k, except for the last partial intake, which has weight $w_{ikl} = x_{ik} - (nux_{ik} - 1)wu_k$.
- Draw nux_{ik} simulated log-concentration values lc_{ikl} from a normal distribution with mean $\mu = \ln(cm_{ik}) - \frac{1}{2}\sigma^2$, and standard deviation σ.
- Backtransform and sum to obtain the simulated concentration in the consumed portion:

$$cr_{ik} = \sum_{l=1}^{nux_{ik}} w_{ikl} e^{lc_{ikl}} \bigg/ x_{ik}$$

Note: Variability between units is often quantified with the coefficient of variation (CV) rather than the variability factor v. With v defined as the ratio between 97.5 percentile and median, the relation between these two characteristics in a lognormal distribution is: $CV = \sqrt{v-1}$, or $v = 1 + CV^2$.

4.2.3.2 Use default (conservative) values of the variability factor v

In the absence of reliable data it may be appropriate to use a default value for v (e.g. v = 5 or 7). This approach is almost equal to the previous one. However, in order to stay on the safe side, the variability factor is only invoked to obtain a larger spread of unit concentrations, but not to lower the estimate of the median value μ. This can be interpreted as assuming that composite samples have been obtained from very homogeneous sets of units (with effectively v = 1), although this homogeneity will not be assumed for consumer portions.

Consequently, in this approach the unit log-concentrations are drawn from a normal distribution with mean $\mu = \ln(cm_{ik})$, and is otherwise the same as described above.

4.2.4.Modelling of non-detect levels

Most monitoring measurements of pesticides are non-detects, i.e. no quantitative measurement is reported. The status of the limit of reporting (LOR) used by the laboratory is often not clear. In the food risk world the limit is commonly indicated as LOD (limit of detection) or LOQ (limit of quantification). Unfortunately, only values higher than LOD or LOQ are reported, in spite of official recommendations to always report the numerical values below LOD or LOQ limits if available (IUPAC 1995). When a pesticide can enter the food chain only via crop treatment, and when the percentage of crop treated is (approximately) known to be $100p_{crop\text{-}treated}$, then this knowledge may be used to infer that $100(1\text{-}p_{crop\text{-}treated})$% of the monitoring measurements should be real zeroes, contributing nothing to pesticide intake, whereas other non-detects in the monitoring data could have any value below the limit of reporting. For $100(p_{non\text{-}detect} + p_{crop\text{-}treated} - 100)$% of the monitoring measurements, 0 and LOR represent best-case and worst-case estimates. A simple way (tier 1 approach) to consider the uncertainty associated with non-detects is to compare intake distributions for these best-case and worst-case situations.

4. 3.Specification of model inputs and uncertainty analysis

We distinguish between choices on the model and model inputs.

1. Choices with respect to the model and problem situations.
   Once made, these choices are considered as fixed: they add no uncertainty to model outcomes. The following choices are relevant:
   a. consumer population: total or restricted to a subset of certain ages
   b. days: all or restricted to a specific subset of days
   c. type of risk calculation: acute (daily intakes) or chronic (usual intakes)

   d. type of model for residue data: empirical (non-parametric) or parametric
   e. for parametric models: pooling of parameters over products yes or no
   f. approach to incorporate unit variability
2. Model inputs: these represent the numeric data that enter the model. In general they will have an associated uncertainty, and in order to allow future extensions of the model to evaluate the uncertainty of the model outcomes it is necessary, that something is known about these uncertainties.

Model inputs are:

1. Food consumption data base considered as a representative sample from the relevant population; uncertainty is implicit in the sample, and can be evaluated with resampling procedures (e.g. bootstrap)
2. Pesticide monitoring data base; in the case of empirical modelling resampling procedures can be used to assess the uncertainty; in the case of parametric modelling the uncertainty can be expressed as standard errors of the parameters.
3. Percentage agricultural use of pesticide (% crop treated)
4. Non-detects; in a simple first approach the maximal uncertainty from non-detects is estimated from a comparison of simulations with substitution of 0 and LOR for the non-detect measurements.
5. Variability factors and unit weights (approach 1 or 2)
6. Processing factors to describe the net effect of processing practice on pesticide intake;

Model inputs 3, 5 and 6 can be specified in general (i.e. applicable for all products), or specific values for products can be given.

For inputs 3-6 one should specify either conservative values, or nominal values in connection with information on the uncertainty in these values. In order to make this as practical as possible this information is requested in the form of a limit (either upper or lower), which should be considered conceptually as a one-sided 97.5 % confidence limit. The program will translate the nominal and limit values into a normal uncertainty distribution on an appropriate scale (logistic for factors restricted to the interval [0,1], lognormal for non-negative inputs such as sample weight.

4. 4.Chronic risk assessment

4.4.1.Introduction

In dietary risk assessment, usual intake is defined as the long-run average of daily intakes of a dietary component by an individual. From a statistical point of view, assessing the usual intake can be reduced to the problem of estimating the distribution of a random variable y_i that is measured with error. A model for the relationship between the observations y_{ij} and the true random variable of interest y_i is:

$$y_{ij} = y_i + u_{ij}$$

where u_{ij} is an additive measurement error for individual *i* on day *j*. For independent, normally distributed y_i and u_{ij}, estimation of the distribution of y_i is straightforward. When observations y_{ij} are non-normal and the measurement error variance is heterogeneous across sampling units, estimation is less simple. Nusser *et al.* (1996) describe a procedure for estimating the percentiles of the distribution of long-run average daily intakes using non-normal dietary intake data. Principally, their method consists of three steps:

1. transforming the daily intake data to approximate normality using a combination of a power function and a grafted polynomial function. The polynomial provides some flexibility against power transformed components that are still deviating from normality,
2. estimating the parameters of the usual intake distribution in the transformed scale, and
3. estimating the percentiles of the distribution of usual intakes in the original scale.

The basic ideas of Nusser *et al*. are suited for dietary components that are consumed on a nearly daily basis, e.g. dioxin in fish, meat or diary products.

4.4.2.Modelling long term daily intake

Usually, food consumption data are available for individuals on 2 (or more) consecutive days. Monte Carlo samples are drawn from the data base to generate an estimate of the probability distribution of the intake of residues. For this situation, the model for the usual intake distribution is:

$$y_{ijk} = y_i + u_{ij} + e_{ijk}$$

with y_{ijk} the observed intake of individual *i* on day *j* and residue *k, y_i* is the unobservable usual intake value for individual *i, u_{ij}* is the unobservable measurement error for individual *i* on day *j*, and e_{ijk}, the unobservable error for individual *i* on day *j* for residue *k*. In the normal scale, $y_i \sim \mathrm{N}(\mu, \sigma^2_{cons})$, $u_{ij} \sim \mathrm{N}(0, \sigma^2_{day})$ and $\mathrm{e}_{ijk} \sim \mathrm{N}(0,\sigma^2_0)$.

4.4.2.1Step 1: power transformation and splinefunction

The observations y_{ijk} are transformed close to normality using a power transformation. As indicated by Tukey (1962), the expected value of a normal score $z = (y\text{-}\mu)/\sigma$ can be approximated by the U-score:

$$U_{ijk} = \Phi^{-1}[(r_l - 3/8)/(N + 1/4)]$$

where r_l is the rank of the ijk^{th} observation y_{ijk} and *N*, the total number of observations. The power γ is estimated by minimising the error sum of squares:

$$\sum_{i=1}^{n} \sum_{j=1}^{o} \sum_{k=1}^{p} (u_{(ijk)} - \beta_0 - \beta_1 y^{\gamma}_{(ijk)})^2$$

over a grid of values of γ, where $U_{(ijk)}$ and $y_{(ijk)}$ denote the order statistics of U_{ijk} and y_{ijk}.
The observations are replaced by power transformed observations:

$$z_{ijk} = y^{\gamma}_{ijk}$$

After a power transformation, some components still deviate from normality. To minimise deviations in the Y-direction an integrated B-spline is fitted to the (U_{ijk}, z_{ijk}) pairs. The spline function is enforced to be monotone increasing by constraining the parameters to be nonnegative. The knots of the spline function are placed such that the interval lengths between knots are equal with two data points left to the left knot and two right to the right knot. The number of knots is optional, here K = 7 is taken. In the intervals, a cubic spline of order 3 is fitted, outside the joint left and right knot the spline is linear. Observations that are transformed by a power in combination with a spline function are denoted by $z_{spline,ijk}$. These values are approximate normally distributed.

4.4.2.2Step 2: estimation of parameters of the usual intake distribution

The power transformed daily intakes are transformed having zero mean and unit variance:

$$z^*_{spline,ijk} = (z_{spline,ijk} - \hat{\mu}_{spline})/\hat{\sigma}_{spline}$$

Parameters of the standardised usual intake distribution in the normal scale are estimated assuming the following model:

$$z^*_{spline,ijk} = z_{spline,\,i} + u_{ij} + e_{ijk}$$

with variance components σ^2_{cons} estimating the variability between consumers, σ^2_{day}, estimating the day to day variability within consumers and σ^2_0, estimating the variability between residues within a sampling day within consumers. The variance components are estimated using standard statistical methods (Restricted Maximum Likelihood). Their sum is close to 1 because the transformed data (indicated by the asterisk) have mean 0 and variance 1. Normal equivalent deviates of the usual intake distribution (mean 0 and variance σ^2_{cons}) are calculated using:

$$q_{usual} = \hat{\sigma}_{between}\Phi^{-1}(p/100)$$

with p a percentage and $\hat{\sigma}_{between} = \hat{\sigma}_{cons}$.

4.4.2.3Step 3: backtransformation and estimation of usual intake

Percentiles in the original scale are estimated by a linear interpolation using the $(U_{ijk}, z^*_{spline,ijk})$ pairs: q_{usual} specifies the values for which interpolated z^*-values are required. The interpolated standardised values, say $z^*_{usual,\ spline}$, are transformed to the original scale by the inverse of the power and correcting for the variance and the mean of the original variable:

$$z_{usual,\ spline} = (z^*_{usual,\ spline} * \hat{\sigma}_{spline} + \hat{\mu}_{spline})^{1/\gamma}$$

4. 5.How to deal with limited information

In the probabilistic model, a distribution of food consumption data as well as a distribution of residue data are used. For both components of the model, a choice can be made between a parametric (see 4.2.1.2) or a non-parametric (see 4.2.1.1) approach. In a *parametric approach* the data are modelled with an appropriate distributional form (e.g. lognormal with parameters σ and μ). In a *non-parametric approach* the empirical distribution is used to sample from directly. Obviously, the latter approach requires more data to obtain a satisfying representation of the full distribution. Therefore, parametric modelling becomes important in data-scarce situations.

4.5.1.The choice between a parametric and non-parametric approach

How many residue data are required for a sensible calculation of upper-tail percentiles in the exposure distribution based on a non-parametric approach? The rule of thumb can be used that the chosen percentile should be contained directly in the data. For example, at least 20 measurements are needed to estimate the 95th percentile and at least 100 measurements to estimate the 99th percentile. More generally, the number of measurements per food commodity (n) should at least equal 1/(1-p%/100) to allow a rough empirical estimate of the p^{th} percentile of the residue concentration distribution to be made. Of course, the risk assessment is only coarse with this minimum amount of data and larger sample sizes per food commodity are certainly worthwhile.

In situations where the number of measurements becomes a problem, an appropriate risk analysis should be based on further modelling. Essentially, the lack of data is compensated by *a priori* assumptions. Assuming a simple distributional form for the residue data, the number of measurements can be smaller in principle (at least 10, say). However, non-detect measurements provide no information about variability, and therefore we should now count the number of positive measurements. Figure 13 shows which approach could be best used depending on the total number of measurements and the number of non-zero measurements. In principle, such a choice could be made separately for each food commodity.

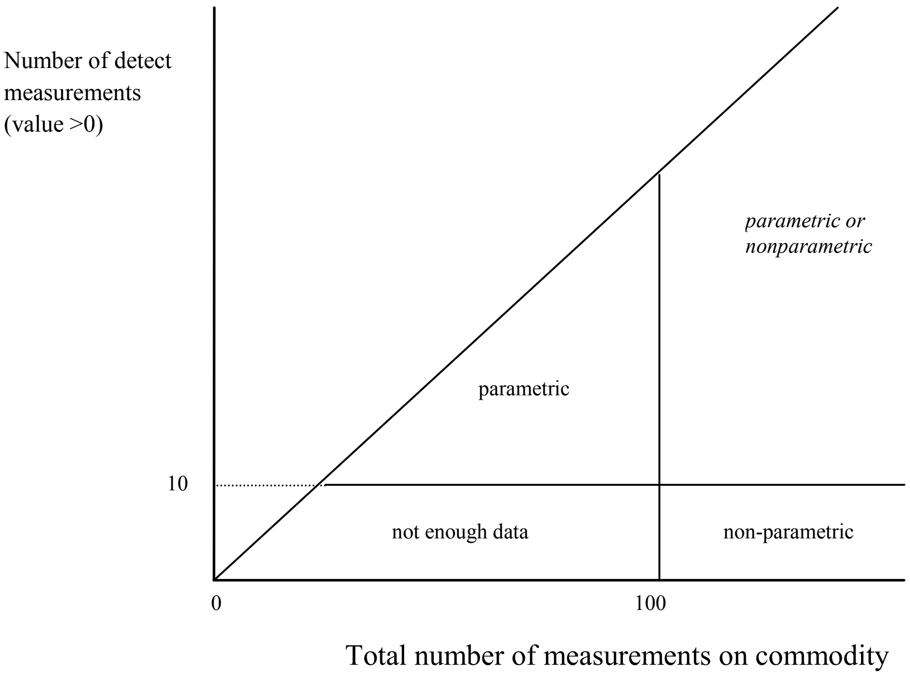

Number of detect measurements (value >0)
10
0
100
Total number of measurements on commodity
parametric or nonparametric
parametric
not enough data
non-parametric

Figure 13: Use of non-parametric or parametric modelling for estimating 99 % exposure percentile in relation to sample size and number of positive measurements.

4.5.2.Grouping of products

When data are limited, the parametric approach has some potential. The distributional form for the residue data is modelled with the lognormal with parameters σ and μ. However, estimation of the sample variance and/or mean are often hampered because data on residues in specific food commodities are sparse or even missing. In those cases, grouping of products into productgroups enlarges the number of measurements per group and may give sufficient data to base estimates upon. We must assume that residue distributions are the same for the grouped products. A second related question is the reliability of estimates, based on a few number of degrees of freedom. The following procedure is designed to cope with the above problems.

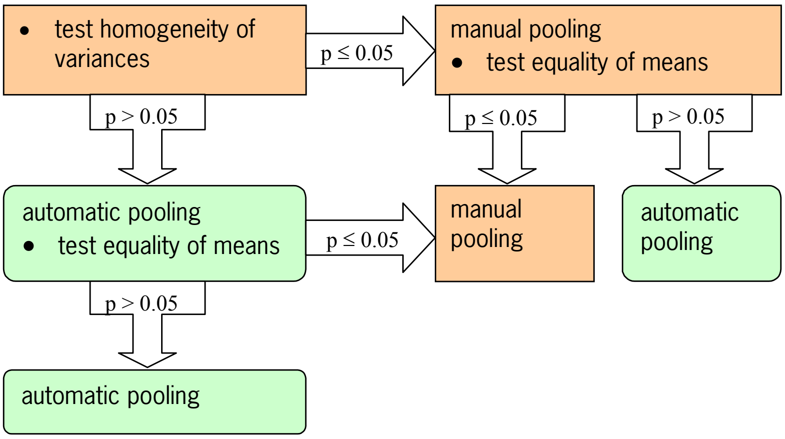
1. Step 1 (see Figure 14). For each product the variance σ^2 and mean μ is estimated. Then, products are assigned to productgroups which are composed of related products, e.g. productgroups consisted of cabbages or all kind of berries. The homogeneity of variances in different productgroups can be assessed using Bartlett's test (Snedecor & Cochran, 1980). The test statistic determines whether variances are to be pooled automatically ($p > 0.05$) or not (p • 0.05). In the latter case, products are assigned to subgroups (within productgroups) manually and the homogeneity of variances is tested again. For homogeneous groups, variances are pooled within productgroups. This process of assigning products to subgroups is repeated until all groups have homogeneous variances. After pooling the variances, an overall test for differences of means is performed, based on analysis of variance. Means are pooled automatically if the probability $p > 0.05$. If not, manual pooling is performed. In Figure 14, steps on the right side require a manual assignment of products into productgroups before variances are pooled. This manual step may be considered optional. This means that when it is decided not to do so, all original variances and means are maintained.
2. Estimates of variances based on less than 10 df are considered not very reliable. Therefore, in step 2 of Figure 14, variances based on < 10 df are compared to the overall variance (pooled over all products except the tested product itself, i.c. corrected) and tested for equality. Variances are replaced by the overall variance (uncorrected) whenever the hypothesis of equality of variances is not rejected; if rejected, the original variances are maintained. If the variance is replaced for (sub)groups with two or more members, a test for differences of means is performed. Means are pooled automatically if $p > 0.05$, if not, the original means are maintained.

3. After carrying out step 1 and 2, there may still remain products with less than 10 df. These products are considered again. The variances are judged visually and assigned by hand to one or more of the products with approximately the same value for the (pooled) variance, After testing the variances, the variances are pooled again, replacing the variance based on < 10 df with the pooled one. Testing for differences of means is performed and for those cases where $p > 0.05$, means are also pooled.
4. Finally, those cases where variances are pooled but means not, are considered again. The products may be rearranged into (sub)productgroups based on similarity of their means. Then, pooled means are calculated replacing the original ones. This last pooling step is optional and not indicated in the figure.

Once decided on performing a parametric risk assessment, rearrangement of products into (sub)groups to estimate the necessary parameters is almost inevitable. Therefore, it is not possible to compare results of a non-parametric risk assessment with a parametric one as such, because nearly always some form of pooling has preceded estimation.

Step 1:

- Calculate variances and means for each product
- Classify products into groups
- Test homogeneity of variances and equality of means within groups of products. Results are: not significant ($p > 0.05$) or significant ($p \leq 0.05$).



Step 2:

- Take products(-groups) with df < 10
- Compare variance with overall variance (corrected). Replace variance with overall variance (uncorrected) for non-significant test results.

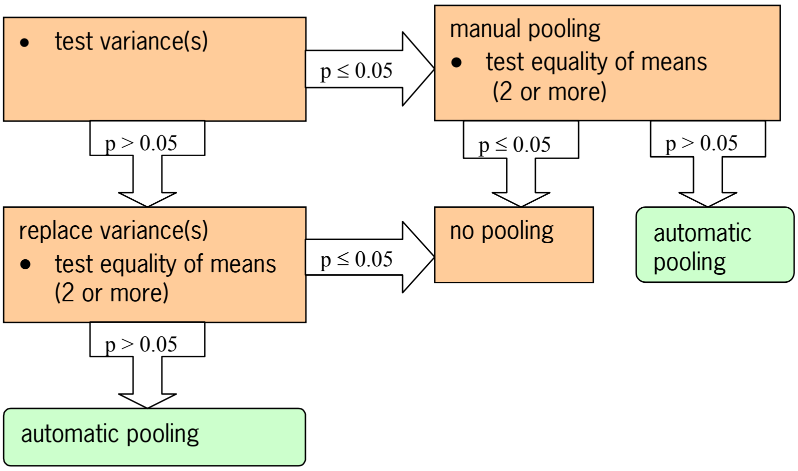


Figure 14: Step 1 and 2, schematic outline of grouping of products into (sub)productgroups when the number of available data on residue levels is limited.

4.5.3. Estimation based on summary data or histogram data

In EU reporting, residue data are sometimes reported in a tabulated (histogram) form. For histogram data, the parameters of the lognormal distribution can then be obtained by fitting normal distributions to a set of observations or counts. Statistics are $n_1 \ldots n_{k,}$ the number of counts in k classes. The group limits are logtransformed and a normal distribution is fitted to standardised normal probabilities based on group limits and the numbers $n_1 \ldots n_k$. Parameters μ and σ are estimated. Group limits c_k are given with c_1 = LOR.

Occasionally, data are reported in a very condensed form. Summary statistics are used to describe characteristics of the underlying residue distributions. These statistics may be used to base estimates of μ and σ upon. De Boer and Van der Voet (2000) describe a procedure to deal with data-scarce situations that seems to work rather satisfying.

5. Appendix

5. 1.Procedures in the MCRA program

PRODLABREAD: reads commodity-labels and levels (PNRLEV). Reads indicator value if residue is allowed or not on a commodity from file xxxx_pro.lis. Prints number of commodities (NPNR).

CMPLABREAD: reads residue-label (compound) from file xxxx_cmp.lis. Prints residue code and label.

INDIVIDUALSREAD: reads consumer characteristics (PERSNR, PLEEF, PGEWI) from file individuals.lis. Calculates the total number of consumers contained in the data base (PERSTAL) and the minimum and maximum age (MINLEEF, MAXLEEF).

CONSUDATREAD: reads consumption data and processing codes from file xxxx_con.lis and stores data in backingstorefile xxxx_con.bac. The procedure checks if a backingstorefile is present. If present, data are directly read from this binary file which speeds up runtime considerably. Note that existing backingstorefiles should be removed or deleted when new data become available. In general, older consumer codes (RESP) do not match the newer ones in the consumption data file. Forms*[2] a subset containing consumers only and prints a warning message.

HDATREAD: loads histogram data on residues from xxxx_histo.xls. Calculates parameters μ and σ of the lognormal distribution. [option summary data = yes][3].

SDATREAD: loads residue summary data from xxxx_sum.dat. Calculates parameters μ and σ of the lognormal distribution. [option histogram data = yes].

FDATREAD: reads residue concentration data from file xxxx_res.lis and the total number of samples taken on each commodity from file xxxx_nos.lis. Calculates mean residues, fraction of positive values (detects) and number of zero residues (non-detects). Restricts the total set to a subset of commodities on which residue concentration data are available. [option empirical concentration data or full data = yes]

LOGNPARA: calculates parameters μ and σ of the lognormal distribution for the full data approach. [option empirical concentration data = no].

VFREAD: loads variability factors and unit weights from file xxxx_varf.xls. Replaces missing unit weights by value 9999. [option use variability factors = yes].

NPNRPRO: calculates new number of commodities or number of commodities/processing type combinations.

PRESENT: generates a variate (PRES) indicating which commodities are present and sets a scalar (SUBSET).

%PRESENT: forms a subset according to the value of scalar SUBSET.

PFLABREAD: loads processing codes and labels from proccode.xls and makes new labels which are combinations of commodity and type of processing. Checks if codes for consumed processed commodities (xxxx_con.lis) are present in proccode.xls. If not, a warning message is printed. Calculates the total number of commodities and processing type combinations (NPNR) and replaces the old value of NPNR (total number of commodities) by the new value. Forms* new variates for unit weights and variability factors by expanding the old structures according to the number of times each commodity is processed. Replaces* unit weights and variability factors of processing types 9, 11 and 13 by default values 9999 and 1, respectively. Contains %PRESENT. [option use processing factors = yes].

DAY%CONSUDATREAD: calculates consumption data matrix for (un)processed commodities. Checks if all labels for consumed commodities are present and prints a warning if unknown commodities are present. Checks the number of days, levels of day and restricts days according to the specified day (to the first day if restricted day does not exist). Prints a warning message if consumptions on just one day are available. Applies* age restrictions and performs pre-processing for a printed summary of the data. [option day restrictions = yes].

[2] * optional, see Figure 3: MCRA stand-alone version: input form MCRA-input.xls.

[3] For procedures that are optional the relevant option is mentioned within brackets [####].

%CONSUDATREAD: calculates consumption data matrix for unprocessed or processed commodities. Checks if all labels for consumed commodities are present and prints a warning if unknown commodities are present. Creates variate with respondent and daynumbers for option Consumers only. Note that the consumption data matrix contains all available days. Applies* age restrictions and performs pre-processing for a printed summary of the data. [option day restrictions = no].

NVHOMOGE: new version of VHOMOGENEITY. Tests homogeneity of variance. [option empirical concentration data = no].

NMHOMOGE: tests homogeneity of means and performs automatically pooling for p >0.05. [option empirical concentration data = no].

POOLING: pools variances and means manually or automatically*. Pooling is performed in a three 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 and

test homogeneity of means within groups
if means are homogeneous,
pool means.

All groups with a significant test result (p < 0.05) are heterogene and the user has the choice to assign manually* commodities to a new subgroup. Means and variances in a subgroup are tested again and the process of pooling, assigning and testing is repeated unless it is decided otherwise.

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.

3. Manual pooling of means within groups with homogene (pooled) variances (result b)

Means are assigned to subgroups. This step is not followed by a test of homogeneity.

The next pop-up menu's are used to guide the user through the pooling procedure:

'Option MANUAL pooling is chosen'
'Create subgroups (file= xxxx_var.xls) or choose automatic pooling'

'Variances and means are pooled manually'
'Variances are heterogene, Ignore heterogeneity?'
'Yes (= pooling)'
'No (= stop program)'
'Cancel (= no manual pooling)'

'Option MANUAL pooling is chosen'
'Create subgroups (file= xxxx_var.xls) or ignore heterogeneity'

'Option MANUAL pooling is chosen'
'Some groups with variances with df < 10 are heterogene!'

'Variances and means for df < 10 are pooled manually'
'Variances are heterogene, Ignore heterogeneity?'
'Yes (= pooling)'
'No (= stop program)'
'Cancel (= no manual pooling for df < 10)'

'Option MANUAL pooling of heterogene means is chosen'
'Create subgroups (file=xxxx.mu.xls) or set manual pooling is no (no pooling)'

[option empirical concentration data = no].

TABPOOLING: prints a summary of the data after the pooling procedure (number of detects, non-detects, fraction of detects, pooled parameters μ and σ of the lognormal distribution, the original parameters on logscale before pooling, number of degrees of freedom of sigma after pooling, productgroups e.g. groups of commodities arranged on common characteristics in combination with allowance of the use of a residue on a commodity). [option empirical concentration data = no].

NOPOOLING: prints a summary of the data (number of detects, non-detects, fraction of detects, parameters μ and σ of the lognormal distribution). For all commodities, parameters μ and σ need to be present because parametric modelling is set without pooling. When some variances are missing, the job is abandoned and a warning message is printed. [option empirical concentration data = no].

CRTRREAD: loads data on percent crop treated.

CHCKS: determines the optimal storage capacity, e.g. chunksize of each cycle within a Monte Carlo simulation. In general, the capacity of the internal memory is too small to process very large simulations in one time. Therefore, a simulation is performed in cycles. The total number of iterations is subdivided in smaller parts and in each cycle a risk assessment is simulated. The chunksize depends on the total number (NPNR) of commodities or commodities and processing type combinations: the higher the number, the lower the chunksize to avoid storage problems of the internal memory. Chunksizes for NPNR <25, for 25 ≤ NPNR ≤ 50 and for NPNR > 50 are set tot 15.000, 10.000 and 5.000 records, respectively. Calculates the number of cycles (LOOP) as the integer value of the total number of simulations divided by storage capacity. The value of the upper quantile of the intake distribution needed for the summary report of the upper tail is specified by the user. This value may conflict with the storage capacity (S) and simulation size (N).

The following rules apply: a constant Q_{max} is defined as S/N*100*2. If the user supplied value is smaller or equal than 100 – Q_{max}, then the current value is replaced by 100 – Q_{max}, and the percentage of the upper tail equals Q_{max}. This rule applies when the user value is set too low (upper tail is too large). On the other hand, when the supplied upper quantile is set too high compared to the total simulation size, the upper quantile is reset to the default value 99.0%, which usually is sufficient.

PFDATREAD: loads available information on processing factors (f_k) from xxxx_proc.xls and the type of transformation for distribution based factors. If $f_{k,upp}$ and $f_{k,nom}$ are both missing, a value 1 is inserted; if $f_{k,upp}$ is missing, it is replaced by $f_{k,nom}$ and vice versa; if $f_{k,nom} > f_{k,upp}$ both values are interchanged. For fixed processing factors, $f_k = f_{k,upp}$ for commodities on which processing information is available. Otherwise a default value 1 is inserted. For distribution based factors, means (= log or logit transformed $f_{k,nom}$) and variances (based on $f_{k,upp}$ and $f_{k,nom}$) are calculated. A warning is printed when distribution based factors f_k cannot be sampled because $f_{k,nom}$ is missing. For those commodities fixed values are taken instead. [option use processing factors = yes].

PRPFLAB: generates print information, e.g. labels for those commodities that are processed in two or more ways. [option use processing factors = yes].

CNSMPTNSIMU: simulates consumption matrix e.g. selects randomly consumers for a specified day [option day restrictions = yes] or selects randomly consumers irrespective of day. Samples available days for option Consumers only. Calculates total consumption of each commodity and number of consumption occasions. Runs within a For-Loop.

PFSIMU: calculates matrix with fixed processing factors or factors based on a normal distribution: for each consumption occasion a processing factor is simulated. Backtransforms values according to applied transformation, e.g. logarithm or logit. Runs within For-Loop. [option use processing factors = yes].

VARFAC: calculates the number of units in a consumption and standard deviation based on variability factors. Calculates the maximum number of units (VMAX) of all consumptions irrespective of commodity. Generates print information about the commodity with the maximum number of units found. Runs within For-Loop. [option use variability factors = yes].

LORREPLACE: replaces missing values by LOR. All values are replaced or replacement is based on the percent crop treated. In the latter case, the sum of the percentage of non-zero's (detects) and number of LORs (replaced missing values) only approximately equals the percent crop treated because in assigning LORs to zeros a randomisation step is involved. Runs within For-Loop.[option replace non-detects by LOR = yes].

E_SIMU: simulates a residue matrix based on empirical data. Residues are simulated for each consumption. When variability is incorporated in the model, VMAX times a new residue matrix is simulated using the sampled value for a consumption and multiplied with consumer unit portions. If option use variability is *no,* VMAX is set to 1. Prints* a message about variability factors. For processed commodities* an expanded matrix is simulated with the number of columns equal to the number of combinations of commodities and processing types. Missing values* are replaced by LOR. Residues* are multiplied with processing factors. Calculates the total sum of the processing factors and the total number of consumption occasions in order to calculate an mean processing factor. The intake is calculated and the total number of positive residues. Contains VARFAC, LORREPLACE. Runs within For-Loop. [option empirical concentration data = yes].

P_SIMU: simulates a residue matrix based on parametric modelling. Residues are simulated for each consumption. When variability is incorporated in the model, VMAX times a new residue matrix is simulated using the sampled value for a consumption and multiplied with consumer unit portions. If option use variability is *no,* VMAX is set to 1. Prints* a message about variability factors. For processed commodities* an expanded matrix is simulated with the number of columns equal to the number of combinations of commodities and processing types. Missing values* are replaced by LOR. Residues* are multiplied with processing factors. Calculates the total sum of the processing factors and the total number of consumption occasions in order to calculate an mean processing factor. The intake is calculated and the total number of positive residues. Contains VARFAC, LORREPLACE. Runs within For-Loop. [option empirical concentration data = no].

RESICALC: generates summary statistics for output. Runs within for-loop.

COLLECT: collects intakes. Collects* day numbers, consumer codes and ages. Runs within For-Loop.

T4ACALC: performs data processing in each cycle to generate the upper quantile of the intake distribution and consumer characteristics of the top 10 intake. Simulation results of two successive cycles are collected in new structures with two times the length of a chunksize. Then, calculations are performed and various data structures with double length needed to produce output are sorted. The intake results needed to summarise the upper quantile of the intake distribution are saved in structures with the same length as a chunk. In the next cycle, these sorted results and new simulation results are collected again and all calculations are repeated. Note that the process of simulating in cycles restricts the value of the upper quantile. Specifying a too large upper tail may supersede the user supplied value. See also procedure CHCKS. Runs within For-Loop.

T4BCALC: performs calculations to summarise the total intake distribution. Runs within For-Loop.

%COLLECT: stores intakes. Stores* day numbers, consumer codes and ages.

GAMMA: performs power transformation on intake distribution. [Long term exposure = yes].

VCREML: estimates variance components [Long term exposure = yes].

%INTERPOLATE: interpolates backtransformed chronic percentiles according to Nusser. [Long term exposure = yes].

NUSSER: Estimates long term exposure based on power or logtransformed intakes using a grafted polynomial. Long term exposure is estimated for the number of non-detects smaller than 1%. Prints percentiles and cumulative percentiles for a specified value (years). Contains GAMMA, VCREML, %INTERPOLATE, HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT (nusser.dat, nusserdiag.htm, percentilesnusser.dat, percentilesnusser.htm)[4] [Long term exposure = yes].

TAB1PRINT: prints a summary of the data used for simulating consumptions and residues. Mean consumptions are averaged after day* and/or age* restrictions. Printed output is on commodity, average consumption for all consumers and consumers only, number of consumer occasions, the average residue (corrected for processing* and after missing values have been replaced by the LOR*), the number of non-zero residues and the total number of samples (non-zero and zero residues). The same information is printed for commodities which are processed* in more than one way.

TAB2PRINT: prints a summary of the simulation results. Printed output, see TAB1PRINT. Three columns are added: the first describes the difference (%) compared to the average consumption of the data and the second the difference (%) compared to the average residue of the data, the last gives the average of the processing factors per commodity corrected for consumption ratio's*. This table is used to compare the simulation results with the summarised data. Large discrepancies between both tables indicate that simulation results are variable.

TAB3PRINT: prints percentiles, the maximum and average intake. Contains HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT (percentiles.htm, percentiles.dat).

T4APRINT: prints characteristics per commodity of the upper quantile of the intake distribution with the corresponding intake. Printed output is relative contribution per commodity, average concentration per commodity, percentage of each commodity with a residue and the average concentration on commodities with a residue. The same information is printed for commodities which are processed* in two or more ways. Contains HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT, HTM1MOUSE, HTM2MOUSE (averconccomres.htm, averconccomres.dat, uppersens.htm).

T4BPRINT: prints characteristics per commodity of the total intake distribution. Printed output, see T4APRINT. Contains HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT, HTM1MOUSE, HTM2MOUSE (totalsens.htm).

TAB5PRINT: prints the intake per commodity of the 10 consumers with the highest total intake and bodyweight and age. The same information is printed for commodities which are processed in two or more ways*.

TAB6PRINT: prints the consumption per commodity of the 10 consumers with the highest total intake. The same information is printed for commodities which are processed in two or more ways*.

TAB7PRINT: prints residue levels per commodity of the 10 consumers with the highest total intake. The same information is printed for commodities which are processed in two or more ways*.

PLTOTDISTR: plots a graph of the total distribution of positive intakes. Contains HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT (totaldistr.htm, totaldistr.dat).

PLUPDISTR: plots a graph of the upper tail of the intake distribution. Contains HTMHEAD, HTMCODE, HTMGEN, HTMBUT, HTMPRINT (upperdistr.htm, upperdistr.dat).

5. 2.Website related procedures in MCRA

By setting the textstructure WEB in MCRA*ddmmyyyy*.gen to 'yes', the program may be used as an internet application. The program generates different types of output that can be viewed on a number of browsers. To communicate with the browser, these output is written as HTML-script. For the webversion of the program some special procedures are written in order to support functionality at the client-side.

WARNING: warning message 'Fatal error occurred, see logfile'

HTMPRINT: pop-up menu to request output

HTMHEAD: header and definitions HTML-pages

HTMCODE: definitions cabinet file and linkage package ComponentOne ActiveX controls

HTMGEN: definitions chartarea ComponentOne ActiveX controls

HTM1MOUSE: definitions mouse control ComponentOne ActiveX controls

[4] Files in parentheses e.g. ####.htm and ####.dat are generated for internetapplications of the program.

HTM2MOUSE: definitions mouse control checkbox ComponentOne ActiveX controls
HTMBUT: definitions button onclick ComponentOne ActiveX controls

5. 3.Specification of inputfiles

In the next paragraphs the format of files needed for a Monte Carlo risk assessment are described.

5.3.1.Basic files

MCRA*ddmmyyyy*.gen: Monte Carlo Risk Analysis program, for GenStat release 5.4.2, 5th edition. Updates releases of the program are identified by the date string *ddmmyyyy*.

MCRALIB: backingstorefile containing MCRA-procedure-library .

MCRA-input.xls: specification spreadsheet for inputs, model and output.

individuals.lis: data on consumer characteristics (personal no., age, weight)
Example:

```
100       4  10
101       3   9
102       2   8
104       3   9
105       3   8
```

5.3.2.Standard data files

Replace '####' by code for residue e.g. IPRO_res.lis for Iprodione concentration data. Note: missing values are indicated by an '*' (files with extension .lis only).

####_con.lis: consumption data and processing code (personal no., day, commodity code (5 columns), consumption, processing code)
Example:

```
100 1 1 07 01 011 01    50.16000   3
101 1 1 08 05 002 01    80.96400  13
102 1 1 08 05 004 01   131.54400   3
102 2 1 09 03 005 02    60.20200  15
104 1 1 08 04 002 02      .36000   5
```

####_res.lis: positive residue concentration on each commodity (commodity code (5 columns), concentration)
Example:

```
1 07 01 010 01  .140
1 07 01 010 01  .190
1 07 01 011 01  .030
1 07 01 011 01  .140
1 08 01 001 01  .600
```

####_nos.lis: total number of samples (detects and non-detects) on each commodity (commodity code (5 columns), no. of samples)
Example:

```
 1 07 01 010 01   101
 1 08 01 001 01   161
 1 07 01 011 01   105
 1 08 01 002 01     9
 1 08 01 005 01   280
```

####_pro.lis: commodity code, labels and a column indicating whether the residue is allowed (1) or not (0) for each commodity (commodity code (5 columns), indicator, label). Note: the last two columns (indicator and labels) are separated by one space (obligatory).

Example:

```
 1 07 01 010 01  0␣BOON, (PRONK/SLA/SNIJBOON)
 1 07 01 011 01  0␣SPERZIEBOON
 1 08 01 001 01  0␣WITLOF
 1 08 01 002 01  0␣ANDIJVIE
 1 08 01 005 01  0␣KROPSLA, BINDSLA
```

Each commodity is characterised by a commodity code built hierarchically from 5 numbers:
1 - productfile number, 1 character
2 - productgroup number, max. 2 characters
3 - productsubgroup number, max. 2 characters
4 - productnumber, max. 3 characters
5 - productquality number, max. 2 characters
####_cmp.lis: residue label (code (3 columns), label)
Example:

```
11 5 1 IPRODION (=GLYCOFEEN)
```

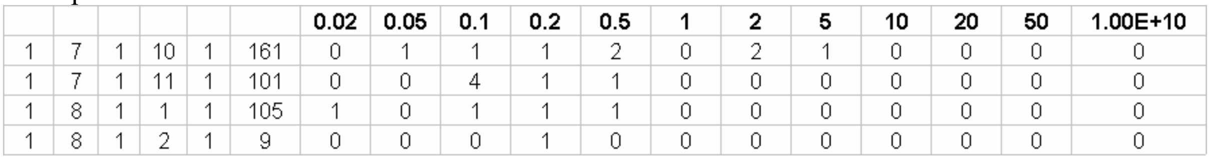
Each residue is characterised by a code built hierarchically from 3 numbers
1 – residue group number
2 – residue subgroup number
3 – residue number

For more details, see Van der Voet et al. (1999).

5.3.3.Optional data files

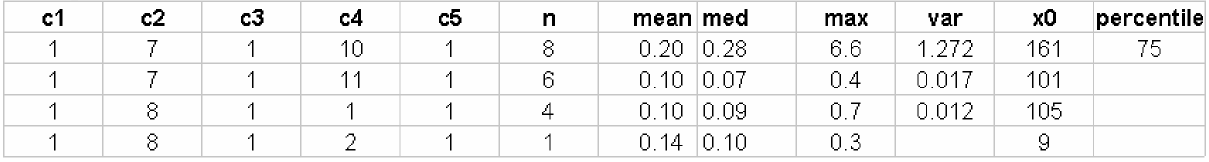
Replace '####' by code for residue. Note: in Excel (.xls), missing values are indicated by a space or an '*'.

####_histo.xls: histogram data
Example:



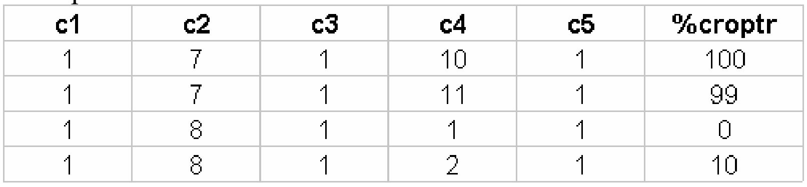
						0.02	0.05	0.1	0.2	0.5	1	2	5	10	20	50	1.00E+10
1	7	1	10	1	161	0	1	1	1	2	0	2	1	0	0	0	0
1	7	1	11	1	101	0	0	4	1	1	0	0	0	0	0	0	0
1	8	1	1	1	105	1	0	1	1	1	0	0	0	0	0	0	0
1	8	1	2	1	9	0	0	0	1	0	0	0	0	0	0	0	0

####_sum.xls: summary data
Example:



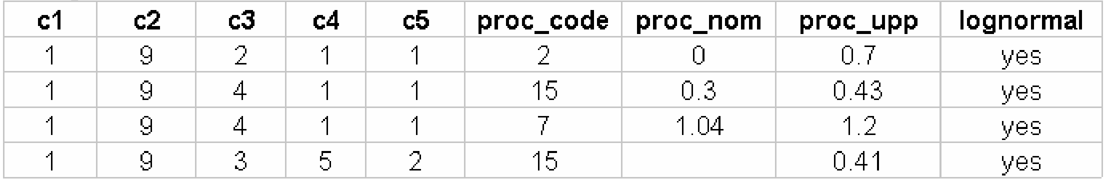
c1	c2	c3	c4	c5	n	mean	med	max	var	x0	percentile
1	7	1	10	1	8	0.20	0.28	6.6	1.272	161	75
1	7	1	11	1	6	0.10	0.07	0.4	0.017	101	
1	8	1	1	1	4	0.10	0.09	0.7	0.012	105	
1	8	1	2	1	1	0.14	0.10	0.3		9	

####_crtr.xls: data on percent crop treated
Example:



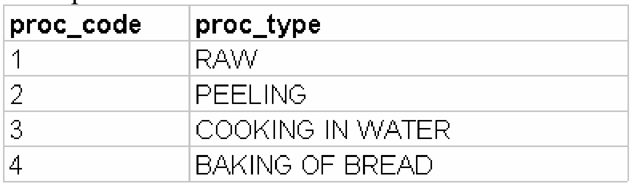
c1	c2	c3	c4	c5	%croptr
1	7	1	10	1	100
1	7	1	11	1	99
1	8	1	1	1	0
1	8	1	2	1	10

####_proc.xls: nominal and upper values for processing factors
Example:



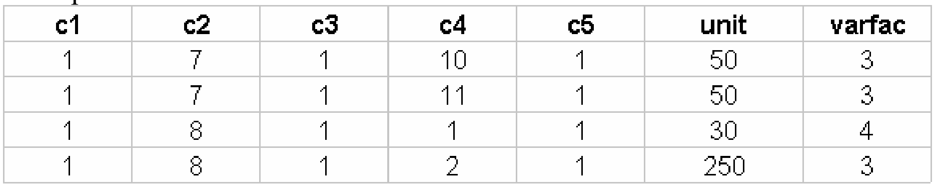
c1	c2	c3	c4	c5	proc_code	proc_nom	proc_upp	lognormal
1	9	2	1	1	2	0	0.7	yes
1	9	4	1	1	15	0.3	0.43	yes
1	9	4	1	1	7	1.04	1.2	yes
1	9	3	5	2	15		0.41	yes

proccode.xls: information on processing codes and names
Example:



proc_code	proc_type
1	RAW
2	PEELING
3	COOKING IN WATER
4	BAKING OF BREAD

####_varf.xls: variability factors
Example:



c1	c2	c3	c4	c5	unit	varfac
1	7	1	10	1	50	3
1	7	1	11	1	50	3
1	8	1	1	1	30	4
1	8	1	2	1	250	3

Files ####_crtr.xls and ####_proc.xls may contain codes of commodities that are not consumed or not present in file ####_pro.lis. MCRA checks which values are present and replaces missing values according to a worst case scenario.

References

- @Risk (1996). Advanced risk analysis for spreadsheets, Windows version. Pallisade Corporation, Newfield, NY, USA.
- Bestfit.(1997). Probability distribution fitting for Windows. Pallisade Corporation, Newfield, NY, USA.
- Blom, G. (1958). Statistical estimates and transformed beta-variables. Wiley, New York.
- Crossley, S.J. (2000). Joint FAO/WHO Geneva consultation – acute dietary intake methodology. *Food Additives and Contaminants*, 17: 557-562.
- David, H.A. (1970). Order statistics. John Wiley & Sons, New York.
- 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.
- GenStat (2000). GenStat for Windows. Release 4.2. Fifth 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.
- Harter, H.L. Expected values of normal order statistics. Biometrika 48: 151-165.
- IUPAC (1995). Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry* 67: 1699-1723.
- Kistemaker C., Bouman M. and Hulshof K. F. A. M. (1998). De consumptie van afzonderlijke producten door Nederlandse bevolkingsgroepen - Voedselconsumptiepeiling 1997-1998. Zeist, TNO-Voeding (Report No: 98.812).
- 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.
- Snedecor, G.W. & Cochran, W.G. (1980). Statistical Methods (7th edition). Iowa State University Press, Ames, Iowa.
- 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 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).