UNIVERSITEIT
GENT

FACULTEIT LANDBOUWKUNDIGE EN
TOEGEPASTE BIOLOGISCHE
WETENSCHAPPEN

Academiejaar 2002-2003

GEO-REFERENCED PROBABILISTIC ECOLOGICAL RISK ASSESSMENT

GEOGRAFISCH GEREFEREERDE PROBABILISTISCHE ECOLOGISCHE RISICOANALYSE

door

ir. Frederik VERDONCK

Thesis submitted in fulfillment of the requirements
for the degree of Doctor (Ph.D) in Applied Biological Sciences: Environmental Technology

Proefschrift voorgedragen tot het bekomen van de graad
van Doctor in de Toegepaste Biologische Wetenschappen: Milieutechnologie

op gezag van

Rector: **Prof. Dr. A. De Leenheer**

Decaan:

Prof. Dr. ir. H. VAN LANGENHOVE

Promotoren:

Prof. Dr. ir. P. VANROLLEGHEM
Dr. ir. J. JAWORSKA

GEO-REFERENCED PROBABILISTIC ECOLOGICAL RISK ASSESSMENT

by

ir. Frederik VERDONCK

PhD supervision committee

Prof. Dr. ir. P. VANROLLEGHEM (UGent)
Dr. ir. J. JAWORSKA (P&G)
Dr. ir. O. THAS (UGent)
Prof. Dr. C. JANSSEN (UGent)

Other Jury members

Prof. Dr. ir. J. VIAENE (chair, UGent)
Drs. T. ALDENBERG (RIVM, NL)
Dr. A. HART (CSL, UK)
Prof. Dr. B. GOVAERTS (UCL)
Prof. Dr. ir. W. STEURBAUT (UGent)
Prof. Dr. ir. M. VANMEIRVENNE (UGent)

The research reported in this dissertation was conducted at the Department of Applied Mathematics, Biometrics and Process Control (*BIOMATH*) of Ghent University, Belgium. This research has been funded by a scholarship from the Flemish Institute for the Improvement of Scientific-Technological Research in the Industry (IWT).

Gent, 17 April 2003

De promotoren:

Prof. Dr. ir. Peter Vanrolleghem
Dr. ir. Joanna Jaworska

De auteur:

ir. Frederik Verdonck

"The greatest risk in life is to not take any risk at all"

Author Unknown

Voorwoord

Tijdens het schrijven van mijn eindejaarsthesis aan Biomath in 1999 profileerden mijn perspectieven zich sterk in de richting van de onderzoekswereld. Hierbij was de combinatie toegepaste wiskunde en praktijkgerichte milieutechnologie voor mij een onontbeerlijke component. Mede dankzij mijn kort verblijf bij Procter & Gamble slaagde ik erin een IWT-specialisatiebeurs binnen te halen. Tijdens mijn doctoraatsonderzoek heb ik altijd getracht een evenwicht te handhaven tussen toegepaste statistiek/modellering en praktijkgerichte milieurisicoanalyse. Via networking op nationale en internationale bijeenkomsten probeerde ik me te verrijken in de verschillende disciplines, belangen en perspectieven van de academische wereld, de overheid (maatschappij) en de industrie. De realisatie van dit doctoraat is dan ook te danken aan honderden kleine discussies en gedachtenwisselingen met vele mensen.

Mijn grootste dankwoord gaat naar mijn promoter Peter Vanrolleghem en aan de andere leden van mijn doctoraatbegeleidingscommissie: Joanna Jaworska (Procter & Gamble), Olivier Thas (UGent) en Colin Janssen (UGent) voor het vertrouwen, de waardering, de begeleiding en het sturen van de lange termijnvisie.

Bedankt ook aan mijn thesisstudenten Frederik De Laender en Sofie Vincke, externe contacten Tom Aldenberg (RIVM, Nederland) en Patrick Van Sprang (EURAS) voor hun bijdrage en inzichten. Verder moet ik nog een hele reeks mensen bedanken: Diederik Rousseau, Karel Deschamphelaere, Andy Hart, Bernadette Govaerts, Ad Ragas, Aquafin NV, VMM, EURAS, ...
Al mijn bureaugenoten en collega's van Biomath zijn bedankt voor de aangename, sociale en gemoedelijke sfeer die de vakgroep typeren.

Tenslotte een hartelijke bedankt aan mijn ouders, mijn zus Beatrice, mijn toekomstige schoonfamilie en bovenal mijn verloofde Greet voor de steun en het vertrouwen in mij.

Frederik Verdonck
Gent, 24 juni 2003

Contents

3.2. Uncertainty and Variability Estimation: Comparison of Several Techniques

3.3. Uncertainty and Variability Estimation: Sample Size Issues

3.4. Hierarchical Uncertainty and Variability Estimation

4. Probabilistic Risk Characterisation

4.1. Probabilistic Risk Quotient Method

4.2. Limitations of Current Probabilistic Risk Characterisation Techniques

5. Geo- (and Time-)Referenced PERA

5.1. Geo-Referenced PERA

5.2. Towards Time-Referenced PERA

6. General Conclusions and Further Research

List of Symbols and Abbreviations

AUC	Area Under the Curve
BC_a	Bias Corrected and Accelerated
BLM	Biotic Ligand Model
COD	Chemical Oxygen Demand
cov(X,Y)	Covariance between variables *X* and *Y*
CPP	Cumulative Profile Plot
DOC	Dissolved Organic Carbon
E[x]	Expected value of *x*
EC	Exposure/Environmental Concentration
ECD	Exposure/Environmental Concentration Distribution
EC_x	Effect Concentration at x %
EDF	Empirical Distribution Function
EQS	Environmental Quality Standards
ERA	Ecological Risk Assessment
EPP	Exceedence Profile Plot
EU	European Union
F or *F(x)*	Target (cumulative) distribution function
$\hat{F}$ or $\hat{F}(x)$	Estimated target (cumulative) distribution function
f(x)	Density function of random variable *x*
Geo-	Geo-referenced
GIS	Geographical Information System
GREAT-ER	Geo-referenced Regional Exposure Assessment Tool for European Rivers
HCp	Hazardous Concentration at p %
i	Rank
JPC	Joint Probability Curve
L	Likelihood function
LN	Lognormal distribution
LAS	Linear Alkylbenzene Sulfonate
LCx	Lethal Concentration at x %
MLE	Maximum Likelihood Estimation
N	Normal distribution
n	The number of observations or data points in a sample
NOEC	No Effect Concentration
PEC	Predicted Exposure Concentration
PERA	Probabilistic Ecological Risk Assessment
PNEC	Predicted No Effect Concentration

QSAR	Quantitative Structure Activity Relationships
r	Correlation coefficient
RQ	Risk Quotient
s	Sample standard deviation
s_w	Weighted standard deviation
SS	Species Sensitivity
SSD	Species Sensitivity Distribution
θ	Parameter of a distribution function
TAN	Total Ammoniacal Nitrogen
TGD	Technical Guidance Document
TOC	Total Organic Carbon
w_i	Weight i
WWTP	Waste Water Treatment Plant
X	A random variable
$\bar{x}$	Sample mean
$\bar{x}_w$	Weighted sample mean

Part 1

-

Introduction

Part 1

Introduction

Human pressure on the environment has increased considerably over the twentieth century, mainly due to the rapid growth of human population and its activities. The pollution of the environment with toxic substances has risen to unprecedented levels. Today, according to the World Health Organisation (WHO), there are over 100.000 different man-made chemicals present on the worldwide market. Chemicals are used in all sorts of products in households, industry and agriculture: pesticides, detergents, shampoos, paints, lubricants, medication, cosmetics, batteries… All these chemicals eventually end up, totally or partly, in the environment through a variety of exposure routes. The negative consequences of these developments have become apparent in the deterioration of ecosystems, the extinction of species and numerous human health hazards (e.g. the crisis of dioxin and PCB contamination in chicken products in Belgium in 1999).

This has led governments to develop new laws and regulation that puts constraints on these chemical emissions. These are based on environmental quality standards and environmental/ecological risk assessment. The key question to be answered is: "What is the likelihood (i.e. probability) of adverse effects occurring to exposed ecological systems due to exceedance of a toxicity level by an environmental concentration?". The goal of ecological risk assessment is to estimate the likelihood and the extent of adverse effects occurring to humans and ecological systems due to exposure(s) to substances. It is based on the comparison of a predicted or measured exposure/environmental concentration with a 'no effect concentration' based on a set of (acute or chronic) toxicity test results (i.e. testing species sensitivity).

In the current, deterministic framework, inputs to the exposure and effect prediction models are single values and the risk is calculated as a simple ratio of exposure concentration and effects (see Figure 1). Consequently, there are only two possible answers to the key question: (1) yes, there is potential risk or (2) no risk. Such answers may mislead stakeholders to think that ecological risks

are simple black or white issues. This conventional, empirical approach does insufficiently account for the inherent variability and uncertainty of the environmental concentration and the species sensitivity. While both being represented by distributions, it is important to separate variability and uncertainty. Variability represents inherent heterogeneity or diversity in a well-characterised population. Fundamentally a property of nature, variability is usually not reducible through further measurement or study. Uncertainty represents partial ignorance or lack of perfect information about poorly characterised phenomena or models (e.g. sampling or measurement error), and can partly be reduced through further research.

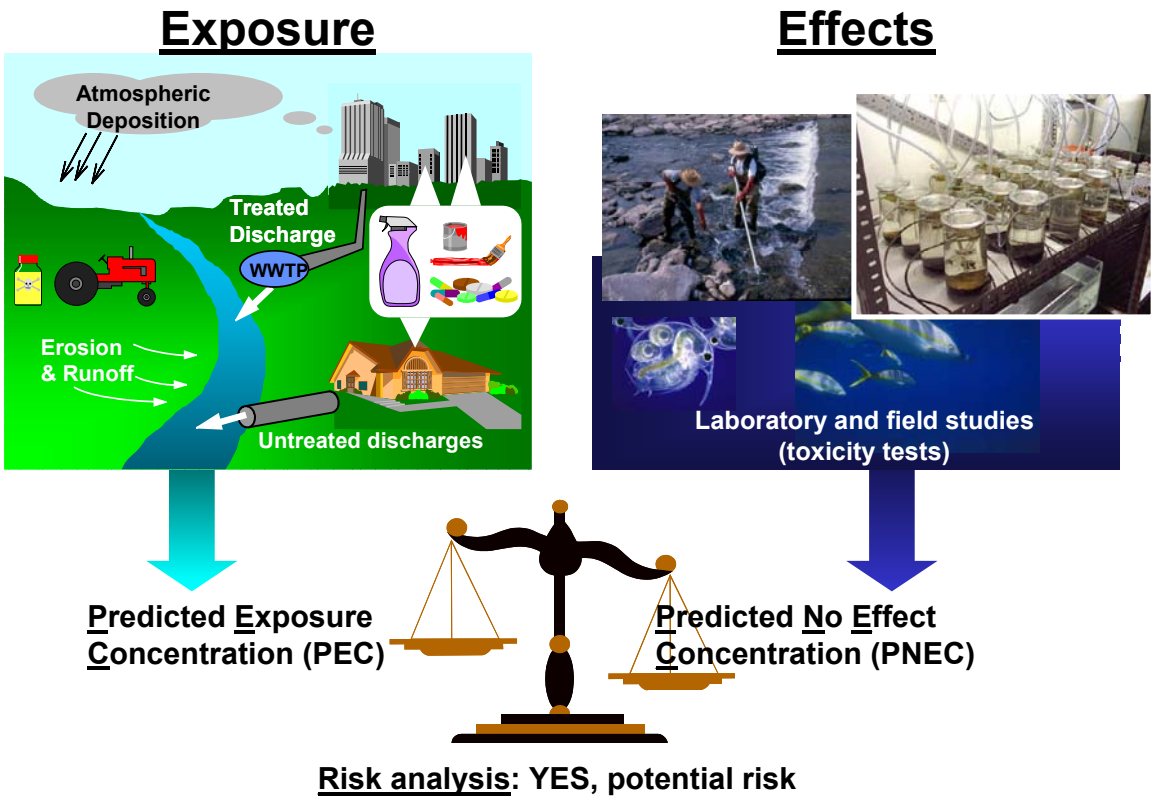


Figure 1: Ecological Risk Assessment

This led to the development of more quantitative and scientifically better funded techniques to estimate probabilistic risks. In a Probabilistic Ecological Risk Assessment (PERA), the exposure concentration and species sensitivity are treated as random variables taken from probability distributions (respectively Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD)) which are combined to give a risk probability. Probabilistic risk assessment therefore delivers a more transparent, realistic and non-conservative approach to estimate risks.

This area is currently a hot topic in the scientific and regulatory field. This is illustrated by the number of recent international meetings on this topic in Table 1. The workshop in 2001 on probabilistic risk assessment for pesticides in Europe (EUPRA) concluded that probabilistic methods *would* improve the environmental evaluation of plant protection products under Directive 91/414/EEC, *if* appropriate action is taken to address their potential weaknesses.

Table 1: Recent workshops on probabilistic ecological risk assessment illustrating the increasing attention and use of such techniques

Date	Place	Topic	Funded by
1999	US	ECOFRAM	EPA
2001	The Netherlands	Probabilistic risk assessment for pesticides in Europe (EUPRA)	EU
2002	London, UK	Statistical extrapolation techniques for environmental effects assessments	ECB
2002	Pensacola, US	Application of uncertainty analysis to ecological risks of pesticides	SETAC
2003	several	EUFRAM	EU

Some of these current (mainly statistical) weaknesses in probabilistic ecological risk assessment are addressed in this dissertation. Most of them deal with misuse of existing techniques (e.g. Monte Carlo analysis, bootstrap), reliability of statistical techniques at small sample size, the lack of consensus on which method or distribution type or what sample size to use, misinterpretation of probability distributions (e.g. output of Monte Carlo analysis), inappropriately or insufficiently dealing with uncertainty or variability (e.g. one- versus two-dimensional Monte Carlo analysis), discussions on how to calculate probabilistic risk… Moreover, all the spatial (and temporal) variability and dependencies is lumped into one probability distribution. Explicitly accounting for these spatial and temporal differences in a respectively geo- and/or time-referenced analysis (or spatial-temporal analysis) could make the risk characterisation more realistic.

The overall objective of this doctoral research is to answer the key question above ("What is the likelihood (i.e. probability) of adverse effects occurring to exposed ecological systems due to exceedance of a toxicity level by an environmental concentration?") with a risk probability and an uncertainty or confidence interval. For this, several statistical methods need to be assessed, implemented and applied in order to characterise or propagate both the inherent variability and uncertainty of the exposure, effects and risk assessment. Note that most of the methodologies presented here are also applicable in other areas (food contamination, agriculture…). In addition, geo-referencing the risk assessment in order to answer the key question more realistically will refine the probabilistic risk assessment. Several case studies are discussed to illustrate all methodologies presented.

The complete outline of this dissertation and the proposal of a geo-referenced probabilistic ecological risk assessment framework can be found at the end of Part 2. This chapter also describes the scope and the state-of-the-art of (probabilistic) ecological risk assessment. The results of this work consist of three major parts: Part 3 describes uncertainty and variability estimation and propagation issues in exposure and effects assessment, Part 4 describes risk characterisation issues and Part 5 describes the geo- (and time-) referencing of risk assessment.

Part 2

-

State-of-the-art

&

Proposal for Improved

Probabilistic Ecological Risk Assessment Framework

A condensed version of the proposal of an improved probabilistic ecological risk assessment framework was published in:

Verdonck F. A. M., Jaworska J., Janssen C. R. & Vanrolleghem P. A. 2002. Probabilistic ecological risk assessment for chemical substances. *Proceedings iEMSs 2002, Integrated Assessment and Decision Support* 1, 144-149. Lugano, Switzerland. 24-27 June 2002.

Part 2

State-of-the-art & Proposal for Improved Probabilistic Ecological Risk Assessment Framework

After defining the scope of this dissertation, a short overview is given of existing frameworks to perform an Ecological Risk Assessment (ERA). Next, the state-of-the-art of the current Probabilistic Ecological Risk Assessment (PERA) is presented. Finally, an improved proposal will be made on how to perform a PERA. This proposal is partly based on the literature study but is a development of the PhD research. This proposal also determined the outline of this whole dissertation.

2.1. Scope

In this section, the scope of this dissertation, along with some (general) concepts and definitions of risk assessment, will be defined.

2.1.1. Hazard and Risk Assessment

Hazard is defined as that object that has the potential for creating undesirable adverse consequences, exposure is the situation of vulnerability to hazards, and risk is the probability or likelihood of an adverse effect due to some hazardous situation. In fact, it is the likelihood to harm as a result of exposure to a hazard, which distinguishes risk from hazard. For example, a toxic chemical that is hazardous to a fish species does not constitute a risk unless fish receptors/populations are exposed to such a substance. Potential risks are estimated by considering the probability or likelihood of occurrence of harm, the intrinsic harmful features or properties of

specified hazards, the population at risk, the exposure scenarios and the extent of expected harm and potential effects (Asante-Duah, 1998).

2.1.2. Ecological Risk Assessment (ERA)

The main application area of risk assessment in this dissertation is "ecology" related. Note that most of the methodologies presented here may also be applicable in other areas. Ecological Risk Assessment (ERA) evaluates the likelihood that adverse ecological effects may occur or are occurring as a result of exposure to one or more agents (Posthuma *et al.*, 2002). In this dissertation, only the effect of one agent at a time will be studied. We believe that the proposed probabilistic framework is flexible enough to extend it for multiple agents.

Ecology is affected by physical changes (such as e.g. channelling of rivers, hydraulic constructions in hydrological systems…) or by chemical changes (i.e. inorganic or organic compounds). Only the effects of these chemical agents are discussed in this dissertation. In practice, protecting ecology against adverse effects of chemical agents often boils down to protect a restricted set of species, mainly because of practical reasons. It is after all not feasible to perform extended ecological studies for thousands and thousands of chemicals that affect the very large number of species that typically constitute an ecosystem. As a result, "ecological" is still often (mis-) used by the scientific community for assessing narrow-minded endpoints (lethality of a set of species). This is also justifiably stressed in literature, e.g. by Forbes *et al.* (2001). Nevertheless, the term ERA is used here because of its common use in literature. In addition, the presented methodologies of this thesis will also hold for other, more relevant endpoints such as demography (i.e. statistical, and mathematical study of ecological populations).

Any risk assessment should be preceded by a problem formulation. This provides a foundation for the entire risk assessment and includes the specification of risk management goals, the selection of assessment endpoints, and the development of a sampling and analysis plan to collect data on measurement endpoints that are needed to support the ERA. However, this phase of defining the problem formulation is not covered in this dissertation.

2.1.3. Probabilistic Ecological Risk Assessment (PERA)

Probabilistic Ecological Risk Assessment (PERA) is an extension of ERA. Several statistical techniques and probability distributions are used to more quantitatively estimate exposure, effects and finally risk. The scope of this dissertation is mainly situated in the use, reliability and improvement of existing techniques and the development of new techniques. For this, a trade-off had continuously to be made between accurate, good but usually more complex statistical

techniques and easy-to-use, -understand and -apply but usually less good statistical techniques. Note that this thesis does not provide guidance on how to decide whether to use probabilistic methods in a particular risk assessment, how to define the outputs that would be required from the probabilistic assessment (before starting it) or how to combine a probabilistic output with other lines of evidence such as field studies.

2.2. Conventional Ecological Risk Assessment

2.2.1. General

Ecological Risk Assessment (ERA) for chemicals seeks to determine whether species, populations, and/or ecosystems are likely to be damaged by chemical inputs from anthropogenic sources. Conventional ERA involves the comparison of a Predicted Exposure Concentration (PEC) with a Predicted No-Effect concentration (PNEC) for a chemical (see Figure 1).

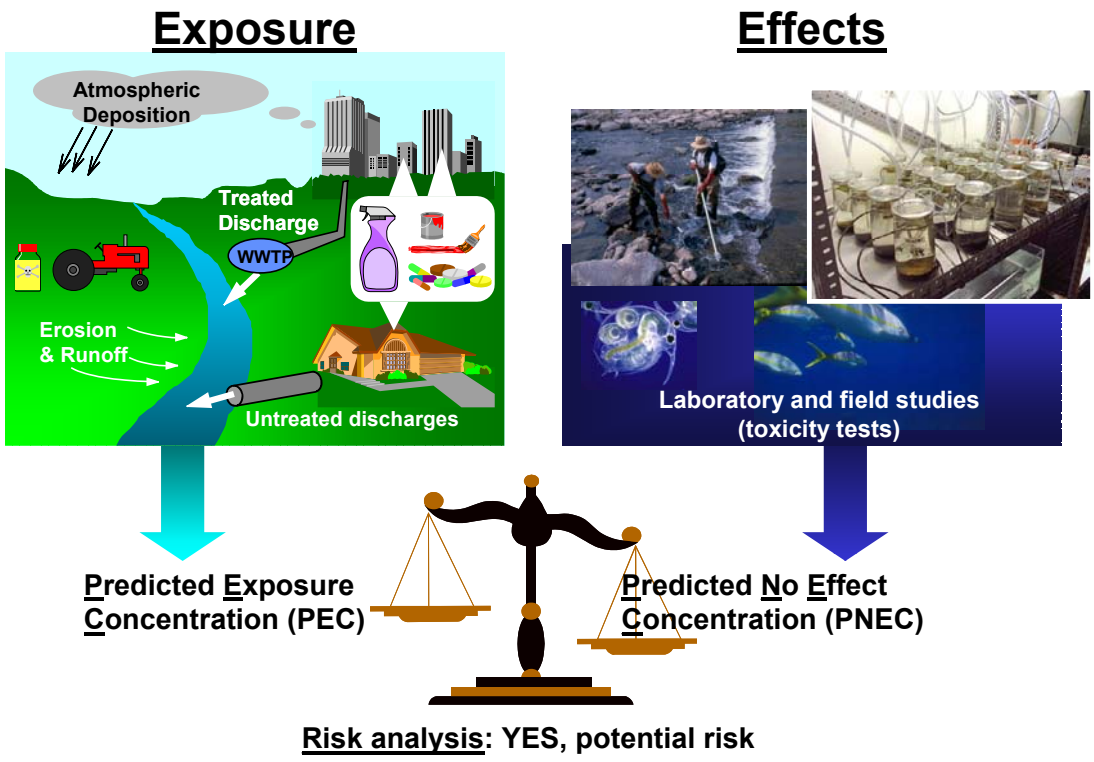


Figure 1: Conventional Ecological Risk Assessment (ERA) framework consisting of an exposure and effects analysis

The risk quotient is defined as the PEC/PNEC ratio. A risk quotient larger or equal to one signifies that there is a potential risk of effects occurring and a large quotient is likely to indicate a high level of risk. A risk quotient smaller than one signifies no risk (considering the conservative assumptions made during the assessments). The need to obtain valid PEC and PNEC values for a particular chemical is therefore fundamental to risk assessment. Much work has taken place over the past few years with the aim of developing procedures that are appropriate for incorporation into risk assessment legislation (Girling *et al.*, 2000).

2.2.2. Derivation of PEC and PNEC in Exposure and Effects Assessment

On the exposure side, a prediction is made of the chemical concentrations in the environmental compartments of concern. Hence, chemical emissions and releases have to be estimated, as well as chemical fate and distribution (see left part of Figure 1). Derivation of a PEC for risk assessment can be subject to different levels of complexity. Further guidance on this matter can be found in, for example, the Technical Guidance Document (TGD) on risk assessment produced by the European Union in support of Commission Directive 93/67/EC on risk assessment for new notified substances and commission regulation (EC) No. 1488/94 on risk assessment for existing substances (EEC, 1996).

On the effects side, selected species are tested on their sensitivity to chemicals (see right part of Figure 1). Several individuals of the same species typically have different sensitivities towards a chemical. This is called intra-species variability. The results are shown in an EC-curve (effect-concentration or dose-response curve). An example is shown in Figure 2. The cumulative probability represents the percentage of individuals affected by the chemical (mortality, reproduction effects, growth effects…).

Next, either the x^{th}-percentile of the EC-curve is taken, resulting in an Effective Concentration EC_x (e.g. EC_{50} for the median) or a No Observed Effect Concentration (NOEC) is calculated based on a significant difference with the lowest observed test concentration. These are the two current approaches mostly used to analyse ecotoxicity test results. In literature, many papers deal with the question whether to use NOEC or EC_x (e.g. Crane & Newman (2000), Scholze *et al.* (2001), Isnard *et al.* (2001), Smit *et al.* (2001), Shieh *et al.* (2001)). The methodologies proposed here can be used both for NOEC or EC_x values and even for other endpoint criteria.

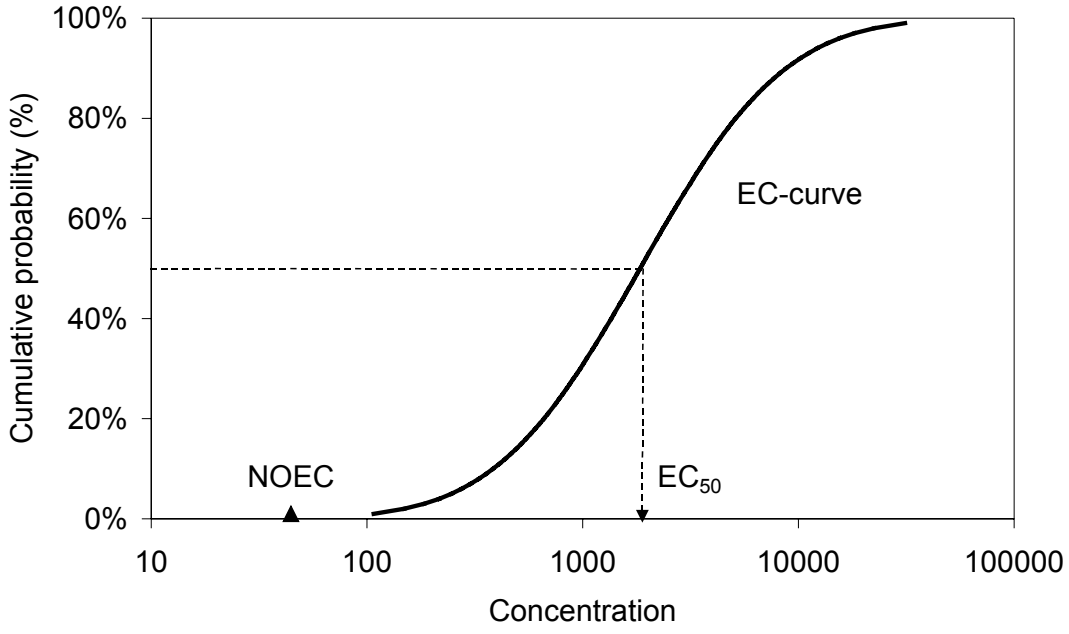


Figure 2: Example of EC-curve (Effect-Concentration curve) and NOEC (No Observed Effect Concentration)

Other species are also tested and they typically have different sensitivities towards a chemical. This is called inter-species variability or species sensitivity. The PNEC is then determined by dividing the single lowest toxicity measure (such as EC_{50} or NOEC values) obtained from single-species tests by an appropriate assessment or safety factor. This method assumes that the assessment factor will be sufficient to ensure that the derived PNEC will protect all species present in a community (i.e. the assessment factor is an arbitrary figure intended to account for extrapolation from single-species laboratory data to natural communities). Assessment factors to derive a PNEC from laboratory test data range from 10 to 1000 depending on the number and kind of data (Girling *et al.*, 2000).

2.2.3. Drawbacks

This conventional method has been criticized many times. The main drawbacks are (based on Hart (2001), EPA (2001), Warren-Hicks & Moore (1995)):

- In the conventional approach, PEC and PNEC are considered as single, crisp values whereas in reality they are characterised by uncertainty and variability (definitions see section 2.3). Accounting for this would avoid problems associated with using worst-case assumptions (e.g. lack of consensus in defining the worst case, and the generation of unrealistically extreme assessments by combining multiple worst case assumptions).

- Results are often viewed as "the answer"; importance of uncertainty is sometimes lost.
- The conventional methods do not efficiently use all available data. For example in the PNEC derivation, only the toxicity to the most sensitive species is used rather than using all available toxicity data to quantify variation between species (inter-species variability).
- The methods also do not encourage further research because uncertainty is inappropriately considered.
- Risk is not expressed as a probability (as it should be, see general risk definition above). Instead, risk is expressed as a ratio, basically a "yes/no risk"-statement.

In short, the conventional methods make use of conservative and insufficiently transparent, realistic, scientific assumptions.

However, these conventional methods are likely to remain the primary tool for lower tiers of risk assessment because they are simple and rapid, and are appropriate for use as screening tool provided they are sufficiently conservative (i.e. over-protective) (Hart, 2001).

2.3. Probabilistic Ecological Risk Assessment

The drawbacks of the conventional, deterministic ecological risk assessment have led to the development of more probabilistic techniques in this area (Hart, 2001). The importance and usefulness of a more probabilistic approach is often stressed in literature (Burmaster, 1997), (Campbell *et al.*, 2000), (Cullen & Frey, 1999), (Jager *et al.*, 2001), (Warren-Hicks & Moore, 1995), (EPA, 2001), (ECOFRAM, 1999). In a Probabilistic Ecological Risk Assessment (PERA), the Exposure Concentration (EC) and Species Sensitivity (SS) are treated as random variables taken from probability distributions (respectively Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD)) which are combined to give a risk probability (see Figure 3). Note in Figure 3 that the interpretation of the cumulative probability is different for the ECD and SSD (how to interpret them will be discussed later).

In these probabilistic types of ecological (and human) risk assessments, the distinction between data uncertainty and variability should be made (Hoffman & Hammonds, 1994), (Cullen & Frey, 1999), (Burmaster & Wilson, 1996), (Rai *et al.*, 1996), (EPA, 2001), (Hart, 2001). The National Academy of Sciences (NRC, 1983) has recommended that the distinction between variability and uncertainty should be maintained rigorously at the level of individual components of a risk assessment (e.g. emissions characterisation, exposure assessment) as well as at the level of an integrated risk assessment. A workshop sponsored by the US Environmental Protection Agency provided recommendations regarding the use of two-dimensional simulations, which were incorporated into a 1997 agency policy document (EPA, 1997).

Variability represents inherent heterogeneity or diversity in a well-characterised population. Fundamentally a property of nature, variability is not reducible through further measurement or study. Temporal and spatial variations of chemical concentrations can be captured in a variability distribution, called ECD. Various species sensitivities towards a chemical can also be captured in a variability distribution called SSD. In Figure 3, the variability distributions are shown by a black line. Uncertainty represents partial ignorance or lack of perfect information about poorly characterised phenomena or models (e.g. sampling or measurement error), and can partly be reduced through further research (Cullen & Frey, 1999). In Figure 3, the uncertainty is shown as a grey band around the cumulative variability distribution function. For each percentile of the variability distribution, an uncertainty or confidence interval can be calculated (i.e. the uncertainty distribution).

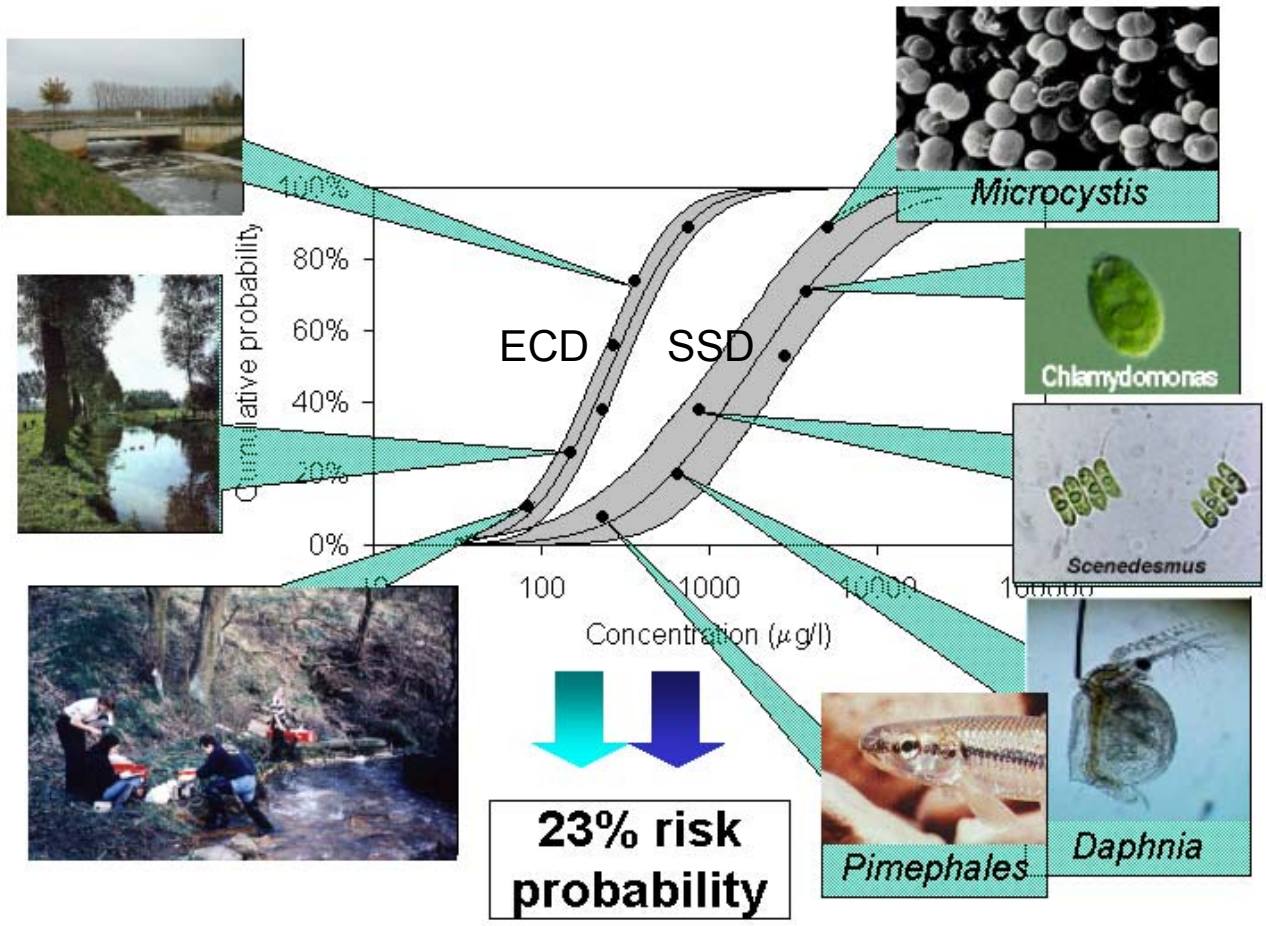


Figure 3: Probabilistic Ecological Risk Assessment (PERA) framework based on Exposure Concentration Distribution (ECD on the left) and Species Sensitivity Distribution (SSD on the right) along with their uncertainty band

Not every assessment requires or warrants a quantitative characterization of variability and uncertainty. For example, it may be unnecessary to perform a probabilistic analysis when screening calculations show exposures or risks to be clearly below levels of concern (and the screening

technique is known to significantly over-estimate exposure). As another example, it may be unnecessary to perform a probabilistic analysis when the costs of remediation are low.
Often, a "tiered approach" may be helpful in deciding whether PERA can add value to the assessment and decision. In a tiered approach, one begins with a simple screening level model and progresses to more sophisticated and realistic (and usually more complex) models only as warranted by the findings and value added to the decision. Throughout each of the steps of a tiered approach, soliciting input from each of the interested parties is recommended (EPA, 1997). PERA can be considered as such a higher tier approach.

Probabilistic methods also have their weaknesses: more complex assessment, requirement of more data, difficult to communicate with stakeholders, difficult to validate… A large part of these weaknesses is surmountable. Another part is difficult to overcome, but also hold for conventional risk assessment. Probabilistic methods are not the only tool, but should be used in conjunction with other tools and other lines of evidence, such as field studies and incident data (Hart, 2001).

This section depicts the state of the art on the first (probabilistic) steps towards a full PERA. It is divided into four parts: the current state-of-the art of probabilistic exposure, effects and risk assessment and how these can be refined in space (and time). This structure is application-driven. The state-of-the art on statistical, modelling and simulation techniques is not discussed here. Rather, they are spread out over all subsequent chapters (mostly the methods sections) where needed. Because the goals of this thesis are methodological, the structure of the dissertation is methodology-driven.

2.3.1. Probabilistic Exposure Assessment: State-of-the-Art

Probabilistic exposure assessment tries to predict or measure the environmental or exposure concentration of a chemical under study (see Figure 1 left and Figure 3 left) by using probabilistic techniques. The exposure assessment field is mainly characterised by prediction models. Consequently, uncertainty and/or variability propagation simulation techniques are more common in a probabilistic assessment. Exception is for example the uncertainty and variability estimation based on a large monitoring database in Govaerts *et al.* (2001).

There are no real methodological developments since most propagation techniques are already well studied and common in other fields. In a propagation method, information on the input is passed on through an exposure model to the output. The two most popular techniques are the Monte Carlo analysis and the (rather recent) probability bounds analysis. Probability bounds analysis is a combination of probability theory and interval analysis. Every variable is specified by a lower and upper bound instead of a probability distribution as in Monte Carlo analysis. Technical details on

how the techniques work can be found in Chapter 3.1. Several examples are given in Table 1. For each study, it is shown whether uncertainty and/or variability and correlations were considered. Below, some examples are discussed in more detail.

A first example is the Monte Carlo engine used in GREAT-ER (Geo-referenced Regional Exposure Assessment Tool for European Rivers). GREAT-ER is a (aquatic) chemical exposure prediction tool for use within environmental risk assessment schemes (Feijtel *et al.,* 1997). GREAT-ER 1.0 calculates the ECDs of consumer "down-the-drain-chemicals" in surface waters, for individual river stretches as well as for entire catchments. The system uses a Geographical Information System (GIS) for data storage and visualisation, combined with simple mathematical steady-state models for prediction of chemical fate (Schowanek *et al.,* 2001). The Monte Carlo takes into account seasonality of the determinants and parameter uncertainty (Boeije *et al.,* 1997). Seasonality deals with major environmental variation throughout the year(s). Parameter uncertainty deals with the difficulties to estimate model parameters, and with the inherent variability of specific processes. Unfortunately, these variability and uncertainty are not treated separately in the Monte Carlo analysis making the ECD output difficult to interpret.

Table 1: Examples of case studies in probabilistic environmental exposure assessment

Reference	Technique	Variability? (ECD)	Correlations?	Uncertainty?	Second order?
(Boeije *et al.*, 1997)	Monte Carlo	Yes	Yes	Partly*	No
(Ritter *et al.*, 2000)	'Joint probability method'	Yes	No	No	No
(Matthies & Berding, 2001)	Monte Carlo	Yes	Unknown	Partly*	No
(Giri *et al.*, 2001)	Monte Carlo	Partly	No	Partly*	No
(Pawlisz *et al.*, 2001)	- Monte Carlo - Probability bounds analysis	Yes	Unknown	Yes	Yes
(Regan *et al.*, 2002)	- Monte Carlo - Probability bounds analysis	Yes	No	Only Bounds	Only for prob. bounds analysis
(MacLeod *et al.*, 2002)	- Monte Carlo - First order analysis	Partly	No	Partly*	No
(Warren-Hicks *et al.*, 2002a)	Monte Carlo	Yes	Yes	Partly*	No

* Partly means the uncertainty of some input parameters was not considered

Probability bounds analysis and second order Monte Carlo, both uncertainty and variability propagation techniques, were used in a case study on Blue Tits exposed to chlorpyrifos in apple orchards by Pawlisz *et al.* (2001). The case study application indicated that the major strengths of both techniques are their ability to identify the most important and reducible sources of uncertainty and to express analyst confidence about risk predictions.

Regan *et al.* (2002) investigated the exposure uncertainty using Monte Carlo analysis and probability bounds analysis. Unfortunately, they compared apples with oranges: a one-dimensional Monte Carlo (describing variability) with a two-dimensional probability bounds analysis (describing variability and uncertainty). They believe that probability bounds analysis is most useful as a tool to identify the extent of uncertainty in model application. However, they did not compare this with a two-dimensional Monte Carlo analysis.

There is, clearly, an increasing trend in the use of ECDs. Nevertheless, there are still some drawbacks related to the use of uncertainty and variability propagation methods:

- Not all Monte Carlo simulations examples in Table 1 separate uncertainty and variability. Some recognize the need to distinguish both but eventually do not do it. A first order or one-dimensional Monte Carlo simulation can only propagate variability or uncertainty, but not both at the same time without having difficulties with interpreting the output distribution. For this, a second order or two-dimensional Monte Carlo simulation is needed.
- Not all propagation methods account for the correlations between the inputs (see Table 1). Vose's (1996) 'cardinal rule of risk analysis modelling' is "Every iteration of a risk analysis model must be a scenario that could physically occur". If e.g. a high river flow is selected ad random, then a low temperature will be more likely than a large one if the river flow is highly negatively correlated with the temperature. Therefore, one of the restrictions to be placed on the model is to recognise inter-dependencies between its uncertain components. It is possible to simulate jointly distributed random variables in which correlations may exist.
- In practice, Monte Carlo is sometimes wrongly applied or the output is insufficiently interpreted. Let us take GREAT-ER, an exposure model with a Monte Carlo engine, as an example. GREAT-ER predicts ECDs. These distributions are difficult to interpret because probability distributions cannot be assigned to all input variables and parameters (even very sensitive parameters as the chemical consumption rate) due to software limitations and mixing up of uncertainty and variability. This makes the output distribution difficult to interpret.

These issues are addressed in Chapter 3.1.

2.3.2. Probabilistic Effects Assessment: State-of-the-Art

Effect models or ecotoxicity data can be used to construct a Species Sensitivity Distribution (SSD). Some examples of effect models are Debtox (Kooijman & Bedaux, 1996), QSARs (Quantitative Structure Activity Relationships), BLM (Di Toro *et al.*, 2001) … Only few developments can be found on the use of propagation techniques and effect models in order to determine SSDs. One example is Fuchsman *et al.* (1999). They developed a probabilistic model to predict effects threshold concentrations for chlorinated benzenes in sediment.

However, this domain is mainly data driven and consequently characterised by statistical developments and applications in the estimation of the variability and uncertainty based on toxicity data. This section only focuses on the statistical developments. Numerous applications can be found in literature (e.g. Posthuma *et al.* (2002)).

Since SSDs were originally proposed to derive environmental quality standards in the late 1970s and mid-1980s in the United States and Europe, respectively, their importance in ecotoxicity evaluations has steadily grown. Since its origin, intensive discussions have taken place on principles, statistics, assumptions, data limitations, and applications (Posthuma *et al.*, 2002). The history of SSD approaches can be found in Van Straalen & van Leeuwen (2002) and Suter II (2002).

The basic assumption of the SSD concept is that the sensitivities of a set of species can be described by some statistical distribution. The available ecotoxicological data are seen as a sample from this distribution and are used to estimate the parameters of the SSD. In parametric methods, the mean and variance among the test species are used to calculate a concentration expected to be safe for most species of interest, which can be used to set an environmental quality criterion. A more recent application is the use of SSDs in ERA.

A SSD can be visualised as a cumulative distribution function (see Figure 4 and Figure 3 right). This is the integral of an associated probability density function. The cumulative distribution function curve follows the distribution of the sensitivity data obtained from ecotoxicological testing, plotting effect concentrations derived from acute or chronic toxicity tests, for example LC_{50} values and No Observed Effect Concentrations (NOECs), respectively. The number of data to construct SSDs varies widely, between few data at all (for many chemicals) to more than 50 or 100 sensitivity values (for a few chemicals). It is evident that the number of data is highly important for the derivation of the SSD, and for conclusions based on them.

The most common current approach is to derive the Predicted No Effect Concentration (PNEC) from the 5^{th} percentile of SSD (EU-TGD, 1995) as shown in Figure 4. Historically that value is known as Hazardous Concentration at p-protection level or HC_p. A cut-off percentage p is chosen

(to protect 1-p percent of species), and the desired "safe" concentration HC_p is calculated. The 5th percentile of a chronic toxicity distribution has been chosen in the earliest methods as a concentration that is protective for most species in a community (namely 1-p %), but the value of p is a policy decision, not science. In popular use of the method, the complementary value of p has become known as the 95% (100-p) protection criterion.

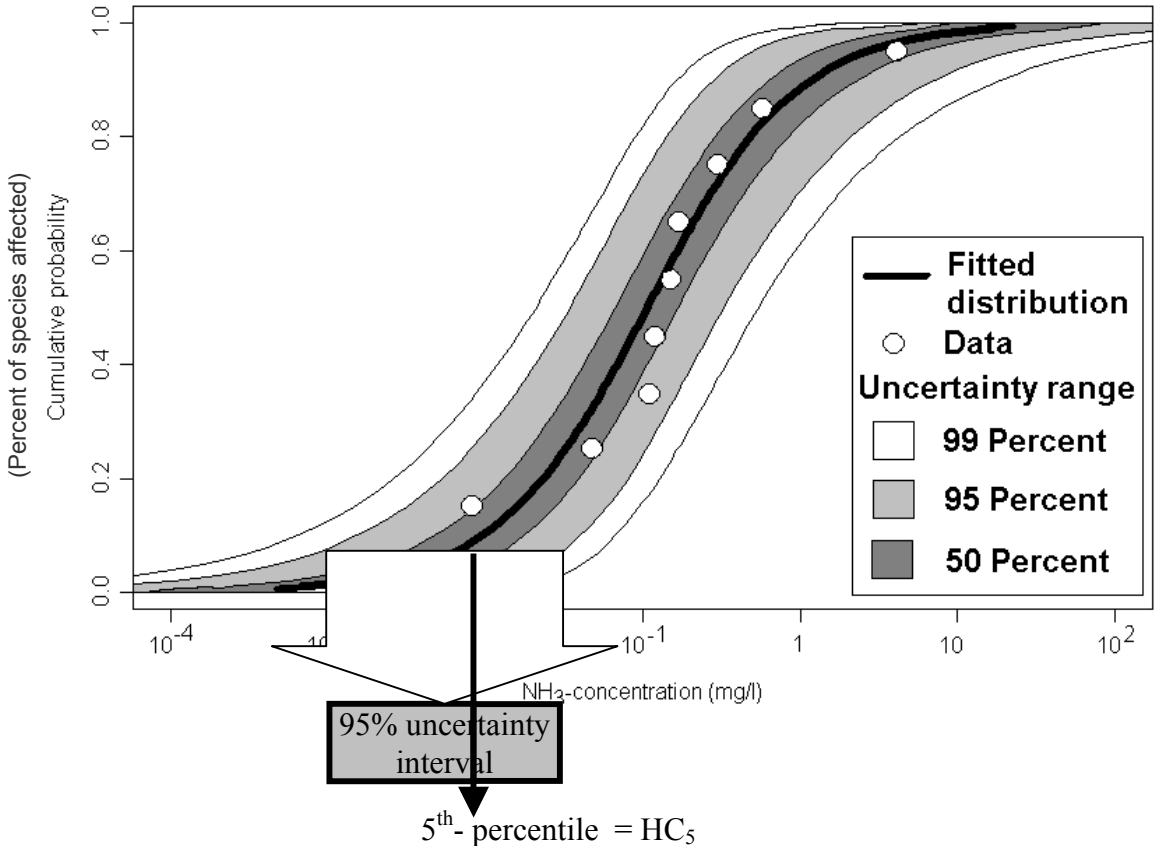


Figure 4: Example of a SSD (Species Sensitivity Distribution - loglogistic distribution) with uncertainty band and its HC_5 (Hazardous Concentration at 5%)

Researchers also started to determine a confidence or uncertainty interval on the HC_5 (see Figure 4). This was mainly done because the median HC_5 is a conservative estimate of the HC_5 calculated without uncertainty (Aldenberg & Slob, 1993). Note that Aldenberg & Jaworska (2000) extended the calculation of uncertainty to both HC_p and p at a given concentration. Until now unfortunately, only few studies report confidence intervals. A confidence or uncertainty interval can quantify the sampling error in the HC_5 estimate. Sampling error is because only a sample of limited size is collected. In theory, one would need to collect an infinite number of samples to obtain the correct estimate of the HC_5.

Several techniques exist to estimate variability and (sampling) uncertainty in a data set. An overview of uncertainty and (inter-species) variability estimation techniques applied in the ecotoxicology field is given in Table 2.

Table 2: Literature overview of SSD case studies with sampling uncertainty in probabilistic ecological effects assessment

Reference	Method for confidence interval estimation	SSD Distribution**	Plotting position ***
(Kooijman, 1987)	Not found	Loglogistic	
(Erickson & Stephan, 1988)	Not found	Triangular (on log-transformed data)	
(Van Straalen & Denneman, 1989)	Not found	Loglogistic	
(Wagner & Løkke, 1991)	Not found	Lognormal	
(Aldenberg & Slob, 1993)	Classical statistics	Loglogistic	
(Jagoe & Newman, 1997)	Nonparametric bootstrapping (resampling)		Mean
(Aldenberg & Jaworska, 2000)	Classical and Bayesian approaches	Lognormal	Hazen
(Newman *et al.*, 2000)	Classical statistics	Gompertz = Weibull	
(Newman *et al.*, 2000)	Nonparametric bootstrapping (resampling)		Mean
(Shao, 2000)	Bootstrapping and maximum likelihood method	Burr type III	
(Van Der Hoeven, 2001)	Nonparametric method		Mean
(Grist *et al.*, 2002)	Nonparametric bootstrapping (resampling) with BC(a) CI*		i/n
(Grist *et al.*, 2002)	Bootstrap regression	Loglogistic	i/n

* BC(a) CI: Bias Corrected (and accelerated) Confidence Interval
** SSD Distribution: this is the assumed probability distribution for parametric
*** Plotting position refers to the way cumulative probabilities are calculated (will be discussed in detail in Chapter 3.2)

This area of quantitative risk analysis is currently an active area of research, but mainly methods from classical statistics, such as bootstrap (Davison & Hinkley, 1997) (Frey & Rhodes, 1998) or maximum likelihood estimation (Frey & Burmaster, 1997), have been applied so far, with an

emphasis on parametric analyses. Parametric bootstrapping and maximum likelihood methods were found to produce similar results (for sample sizes 5, 10 and 20) (Frey & Rhodes, 1998). Aldenberg & Jaworska (2000) compared Bayesian and classical approaches for the Gaussian (normal) distribution (for several sample sizes). Despite largely different numerical schemes, both approaches lead to identical answers. Jagoe & Newman (1997) compared the nonparametric bootstrapping (resampling) with the maximum likelihood method (assuming lognormal distributed data). The parametric method was found to be superior to the resampling, however only in the case of lognormally distributed data. As stated by Newman *et al.* (2000), clear advantages are to be gained by using the non-parametric bootstrap methodology to generate HC_p estimates, because no assumptions have to be made on underlying distributions. However, their insightful article only used the basic bootstrap technique (for sample sizes larger than 20). Grist *et al.* (2002) describe a hybrid bootstrap regression approach. They found that this method can yield a substantially different estimate for the SSD when compared with both the basic nonparametric bootstrap by Newman *et al.* (2000) and the more frequently used parametric approaches. So far, not all techniques have been compared for small data sets (e.g. sample size = 20 or less).

There is, clearly, an increasing trend in the use of SSDs and their methods to estimate uncertainty. This is already a large improvement compared with the deterministic effects assessment, which can still be a useful, screening tool. Nevertheless, the current methods for estimation of uncertainty and inter-species variability still have statistically related drawbacks. Note that only the drawbacks relevant for this thesis are listed and most of them were also described by Hart (2001):

- As is clear from Table 2, several techniques can be used to estimate variability and uncertainty: bootstrapping, classical and Bayesian approaches. Most of these techniques or their properties are asymptotic, i.e. they are valid if the sample size tends to infinity. In practice, data sets on toxicity tests are scarce and if available often only at small sample sizes. Consequently, this raises the question: "Given small sample sizes, which techniques are most suitable and should be used for estimating the 5^{th} percentile (or HC_5) from SSDs?"
- As is clear from Table 2, both parametric and nonparametric methods have been proposed and used. Both types give different results for the same estimator (HC_5 in this case). This raises the question: "Should parametric (and if so which distribution type) or nonparametric techniques be used for SSDs at small sample sizes?"
- In addition, none of the papers in Table 2 (except Jagoe & Newman (1997) and Shao (2000)) performed a simulation study in which important concepts as coverage and bias were tested to assess the reliability of their proposed method (both concepts will be explained in Chapter 3.2).
- Both mean and Hazen plotting have been proposed and used (see Table 2). Again, both give different results. Which plotting is the most accurate and should be used?
- In Table 2, only inter-species variability and sampling uncertainty due to selecting species from a community is considered. There are more sources of uncertainty and variability. Until now, no real proposals are made in literature to include inter-laboratory variability,

intra- and inter-species variability into one single SSD reflecting all inherent natural heterogeneities (i.e. hierarchical variability). In addition, one might be interested in an uncertainty estimate as well. For this, all sources of sampling error need to be included in an uncertainty band around the SSD.

These issues are addressed in Chapter 3.2 and 3.4. Not all drawbacks of the use of SSDs are described here. Only the more statistically related problems are dealt with. Other drawbacks can be found in e.g. Forbes *et al.* (2001) and Posthuma *et al.* (2002).

One of the other remaining issues in SSD determination is the choice of sample size. The choice of an appropriate sample size is an essential component of any experimental design. Two important considerations need to be made to determine the appropriate sample size. First, the accuracy and scientific reliability of the method to estimate the 5^{th} percentile should be assessed (relevance of parameters, sampling strategies, availability and representativeness of toxicity data both in terms of number and kind of species and taxonomic groups). Some methods cannot be applied at small sample sizes (say < 20). Second, the desired level of precision should be defined and assessed. Several papers (see Table 3) have already been published on the determination of a minimum or optimal sample size for SSDs. However, each researcher uses his/her own considerations.

Table 3: Overview of proposals on minimum or optimal sample size for SSDs (Species Sensitivity Distributions), together with the criterion considered to obtain that size (consideration 1: the level of scientific reliability, consideration 2: the desired level of precision in the estimation)

Reference	Consideration	Minimum/optimal sample size
(Stephan *et al.*, 1985)	Unknown	8
(van Leeuwen, 1990)	2	5
(Baker, 1994)	Unknown	4-8
(Cowan *et al.*, 1995)	1	20
(Solomon, 1996)	1	9
(Roman *et al.*, 1999)	Both	No proposal
(Vega *et al.*, 1999)	1	10
(Newman *et al.*, 2000)	2	15-55
(Van Der Hoeven, 2001)	2	No proposal
(Wheeler *et al.*, 2002)	1	10

Consequently, different sample sizes are proposed depending on the used criterion and method (in an often incomplete analysis). This makes it extremely difficult to compare several proposals on sample size. In addition, a whole range of the minimum/optimal sample sizes is proposed for SSDs (see Table 3). There is a need for a proper, standardised and scientifically sound procedure to determine a minimum sample size. This issue is addressed in Chapter 3.3.

2.3.3. Probabilistic Risk Characterisation: State-of-the-Art

First, the methodological developments of probabilistic risk characterisation in literature will be discussed. Second, some applications in literature will be mentioned. Finally, the drawbacks of the current methodological and practical developments are discussed.

2.3.3.1. Probabilistic Risk Characterisation Methods: State-of-the-Art

The characterisation of the risk of toxicants to species, when both EC and SS are uncertain and variable is the central issue in PERA (Aldenberg *et al.*, 2002). The methodology focuses on cumulative distribution function-type probability plots of both the ECD and the SSD and is well developed (Cardwell *et al.*, 1999), (Solomon *et al.*, 1996), (Solomon *et al.*, 2000), (Solomon & Takacs, 2002), (Giesy *et al.*, 1999), (Giddings *et al.*, 2000), (Warren-Hicks *et al.*, 2002). The calculation of a probabilistic risk can be done in many ways. The overlap between the EC and SS probability density functions, as well as between the respective cumulative distribution functions, have both been suggested as a measure of this risk (cf. Solomon *et al.* (2000)). However, such graphical measures of risk can be defined exactly.

Cardwell *et al.* (1993) plotted cumulative distribution functions of the ECD and of SSD for chronic and acute toxicity over log concentration. The cumulative probability of the SSD expresses the percentage of species affected, and the cumulative probabilities of the ECD are converted to probabilities (%) of exceeding certain log concentrations. An example of an ECD and a SSD is shown in the left panel of Figure 5.

Because risk assessment considers both likelihood of EC and likelihood of SS, risk can be expressed as a joint probability, for example, that n% of species will be affected x% of the time or in y% of the locations, depending on the type of exposure data collected. These probabilities can be expressed as the probability of exceeding a fixed criterion on the SSD, such as, for example, the 10^{th} percentile of the distribution of all species or a distribution of inherently more sensitive species. Another method of presenting these joint probabilities is in the form of a Joint Probability Curve (JPC). This format was suggested in the AERA program (The Cadmus Group, 1996), recommended for displaying risks by the Aquatic Working Group of ECOFRAM (ECOFRAM, 1999). The derivation of the JPC is relatively simple and offers a useful tool for communication of risks as it allows what-if questions to be addressed and gives the risk assessor and risk manager a method for assessing the effects of changes in assumptions, such as the choice of a different percentile from the species affected.
The plotting of joint probabilities come in graphs called risk distributions or risk distribution functions (Warren-Hicks *et al.*, 2002), and joint probability curves, namely exceedance profile plots (ECOFRAM, 1999), (Giesy *et al.*, 1999), (Solomon *et al.*, 2000), (Giddings *et al.*, 2000), (Solomon

& Takacs, 2002) or cumulative profile plots (Aldenberg *et al.*, 2002). Therefore, JPCs come in two forms: either as a graph of ECD exceedance against fraction of species affected (i.e. cumulative probabilities of SS), or as a graph of fraction of species affected against cumulative probabilities of EC. The first is called an Exceedance Profile Plot (EPP) (Giesy *et al.*, 1999), and involves plotting one minus the cumulative probability of the ECD against the cumulative probability of the SSD for any given concentration. The second JPC curve results from plotting the cumulative probability of the SSD on the ordinate against the cumulative probability of the ECD on the abscissa for any given concentration. The latter JPC plots are called Cumulative Profile Plots (CPP). Both JPCs represent the same risk curves; they are just different ways of visualisation. The EPP is more common but the CPP is probably easier to draw and interpret than the EPP, since it only involves cumulative probabilities (Aldenberg *et al.*, 2002). An example of a JPC - CPP is plotted in the right panel of Figure 5.

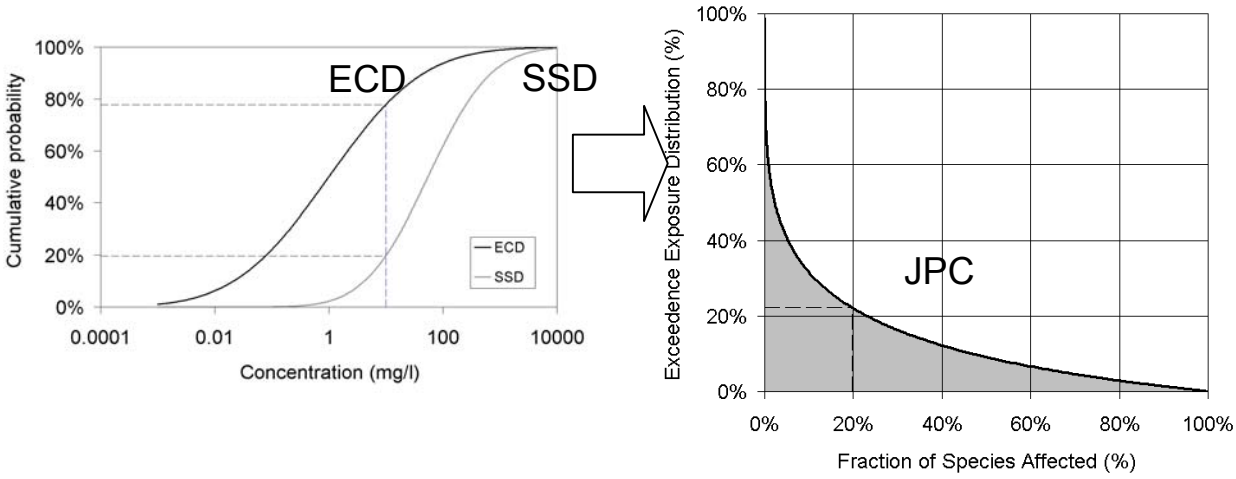


Figure 5: Exposure Concentration Distribution (ECD), Species Sensitivity Distribution (SSD) and Joint Probability Curve (JPC) (type EPP: Exceedance Profile Plot) with its Area Under the Curve (AUC)

The Area Under the Curve (AUC) of a JPC is considered as a numerical measure of the risk of the toxicant to species (Solomon *et al.*, 2000), (Solomon & Takacs, 2002), which a risk manager wants to minimise.

There are also other methods for risk characterisation. Aldenberg *et al.* (2002) compared different methods mathematically and concluded that the discrete summation for the expected risk of Cardwell *et al.* (1999), Van Straalen's ecological risk (1990), the numerical integration of risk distribution curves in the WERF methodology (Solomon & Takacs, 2002) (Warren-Hicks *et al.*, 2002), as well as the AUC of JPCs are all numerically equal to, and may be interpreted as, *the risk of some log EC to exceed some log SS*, as originally implemented by the probability of failure in reliability engineering. The graphical interpretation of this risk is the AUC of the product of the

ECD cumulative distribution with the SSD probability density function, or alternatively, the AUC of the product of the ECD probability density function with the SSD cumulative function.

In addition, Aldenberg *et al.* (2002) developed a probabilistic risk look-up table when both ECDs and SSDs are (log)normally distributed (Table 5.3 in Aldenberg *et al.* (2002)). To avoid tabulating four parameters to determine the probabilistic risk (mean and standard deviation of EC and SS), the SSD can be standardised and the ECD can be scaled to the SSD. In this way, a two-parameter dependent probabilistic risk is obtained.

2.3.3.2. Some Probabilistic Risk Characterisation Applications

The above developments in PERA increase in attention and are more frequently used. To illustrate this, Table 4 gives an overview of some applications found in literature. For each case study, it was checked what kind of risk characterisation method is used and if an uncertainty or confidence interval is calculated. In studies where multiple methods/tiers are presented or used, the most advanced one is described. Note that in most references it was not the main purpose to assess the risk characterisation method itself.

Table 4: Non-exhaustive literature overview of PERA case studies

Reference	Risk characterisation method	Risk Uncertainty?
(Van Straalen, 1990)	Quantitative risk calculation	No
(Warren-Hicks & Moore, 1995)	Semi-quantitative	Semi-quantitative
(Solomon *et al.*, 1996)	Quantitative overlap	No
(Manz *et al.*, 1999)	Quantitative overlap	No
(ECOFRAM, 1999)	JPC	No
(Cardwell *et al.*, 1999)	Some type of JPC	Qualitative
(Moore *et al.*, 1999a/b)	Risk function	Qualitative
(Giddings *et al.*, 2000)	JPC	Qualitative
(Campbell *et al.*, 2000)	Probabilistic risk quotient	No
(Solomon, 2000)	JPC	No
(Duvall & Barron, 2000)	Probabilistic risk quotient	Semi-quantitative
(EPA, 2001)	Probabilistic risk quotient/ JPC	Yes
(Maund *et al.*, 2001)	Probabilistic risk quotient but based on SSD and an ECD point estimate	No
(Aldenberg *et al.*, 2002)	Mathematical Risk framework	Qualitative
(Poletika *et al.*, 2002)	Qualitative overlap	No
(Regan *et al.*, 2002)	Probabilistic hazard quotient but based on ECD and a ‘SSD’ point estimate	Only bounds
(Schwacke *et al.*, 2002)	Risk = P(EC) * P(SS|EC)	Yes

2.3.3.3. Drawbacks of the Current Probabilistic Risk Characterisation Methods

There is, clearly, an increasing trend in the use of more quantitative, realistic risk characterisation methods. This is already a large improvement compared to the deterministic risk quotient method, which is still a useful screening tool. Nevertheless, the current probabilistic risk characterisation methods still have drawbacks.

- Sometimes (especially in human health risk assessment), only ECD (or only SSD) is considered and both the variability of the EC and SS are not accounted for fully.
- The probabilistic risk lookup table of Aldenberg *et al.* (2002) is, unfortunately, only valid for (log)normal distributions. It is shown above that other distribution types (such as Pareto) may be needed for risk assessment purposes.
- Nonparametric methods have not yet been used in the risk characterisation although it is shown above that they can be useful as well. Moreover, not all risk characterisation methods presented above are capable of dealing with nonparametric distributions.
- The JPC and AUC methodology, developed in ECOFRAM (1999) can be applied to many types of distributions but is unfortunately, although relatively easy to construct and calculate, sometimes difficult to understand and interpret by decision-makers and risk managers. This was my personal experience at a SETAC Pellston workshop on the application of uncertainty analysis to ecological risks of pesticides (24 February until 1 March 2002, Pensacola, Florida, USA).
- Some attempt to calculate a probabilistic risk quotient. In short, a probabilistic risk quotient is (as in the conventional risk quotient) the ratio of the EC and SS, but EC and SS are now treated as probability distributions. Maund *et al.* (2001), for example, only consider the SS as a probability distribution in the probabilistic risk quotient. Although available, the EC is reduced to a point estimate. Campbell *et al.* (2000) and EPA (2001) on the other hand successfully calculated a complete probabilistic risk quotient.
- Almost no literature sources calculate an uncertainty or confidence interval on their risk estimate although many acknowledge the need to distinguish between uncertainty and variability.
- Consequently, little attention is given to the visualisation of the risk and its uncertainty interval.

Note that only the drawbacks relevant for this thesis are listed. A solution for these problems will be investigated in Part 4.

2.3.4. Refining the PERA in Space (and Time)

Currently, risk assessments, especially those for regulatory decisions are done for generic situations determined by a set of default values. However, the spatial and temporal variability of exposure and effect, for instance, can be quite high. Spatial variability in chemical concentrations arises from many factors, including the mechanism of contamination, physical and chemical dilution and transformation processes, and physical characteristics of the site (Cullen & Frey, 1999). Temporal variability in chemical concentrations may arise from for example wind erosion, leaching and bioaccumulation, which may result in concentrations in predatory fish that increase with time. For example in Belgium alone, atrazine concentrations in surface water range from 50 ng/l (detection limit) to more than 1 mg/l (Vandenbroele *et al.*, 2000). This is a range of five orders of magnitude. The temporal variation of atrazine in Susquehanna River fall line (USA) is ranging from about 20 to 250 ng/l (Foster *et al.*, 2000). This can already be very high for some narrow tolerance range of aquatic organisms. Both diurnal and seasonal variability of Linear Alkylbenzenesulfonates (LAS) concentrations in river basins in Japan are well documented from previous studies (Takada *et al.*, 1992), (Takada *et al.*, 1994). The lowest concentration (at noon) in the river Nogawa was a factor 4 lower than the highest concentration (in the early morning). Similarly, LAS concentrations in the river Tamagawa in winter months are about 10 times higher than those in summer are.

Contamination of surface waters from pesticide typically occurs in single or repeated pulses due to agricultural runoff, spray drift, or intermittent urban and domestic use. These input patterns typically result in a period of high concentration followed by a decline in concentration due to hydrological dilution, degradation, or partitioning from water to air or sediments. A second pulse may follow the first in a matter of days, or pulses may be separated by as much as a year or more. Standard laboratory toxicity tests using constant exposure concentration typically do not investigate the toxicity of time-varying or repeated exposures. The difficulty of estimating effects of realistic time-varying exposures from measurements made under constant exposure is often an important source of uncertainty in ecological risk assessment of pesticides (ECOFRAM, 1999).

Consequently, incorporating spatial and temporal characteristics of the receiving environment could further increase realism. Geographical Information Systems (GIS) and dynamic assessments are useful tools to account for the spatial and temporal variability of the ECD and SSD.

In literature, several examples can be found on exposure, effects and/or risk characterisation combined with GIS and/or dynamic analysis. GREAT-ER, for example, has already built in the idea of refining the exposure assessment by explicitly accounting for the spatial variability (geo-referencing the ECD) (Feijtel *et al.,* 1997). Instead of having one lumped ECD for an entire catchment (representing spatial and temporal variability), each river stretch has its own ECD (only representing temporal variability). A Monte Carlo analysis propagates the temporal variability of

the input parameters (such as the river flow). Other examples on GIS applications can be found in Röpke *et al.* (2002), Di Mauro *et al.* (2000), Giri *et al.* (2001), Havens *et al.* (1998)…
For dynamic exposure modelling and monitoring, examples are given in Deksissa & Vanrolleghem (2001), Karman & Reerink (1998), ECOFRAM (1999). Examples on effects monitoring are given in Mancini (1983), Reinert *et al.* (2002), ECOFRAM (1999), Bonnomet *et al.* (2002), Milne *et al.* (2000), Karman & Reerink (1998) and examples on effects modelling can be found in Kooijman & Bedaux (1996), ECOFRAM (1999).

Clearly, there is an increasing trend in the use of more GIS and dynamic tools in the risk assessment area. Nevertheless, these developments are just the first steps towards a full geo- or time-referenced PERA.

- No study was found that presented a lumped exposure or effects distribution and then actually refined the spatial component of that distribution by geo-referencing the analysis. Nevertheless, many studies were found that include GIS in their exposure or effects analysis, albeit without this purpose of refinement.
- Geo-referencing exposure, as in the above GREAT-ER example, is already a refinement of the exposure estimate. However, the real benefit of geo-referencing would be to geo-reference the effects as well. In this way, risk will be geo-referenced and should therefore be more refined and realistic.
- While GIS systems provide powerful tools for spatial analysis, their capabilities for complex and dynamic analysis are limited. Traditional models, on the other hand, are powerful tools for complex and dynamic situations but they often lack the intuitive visualisation and spatial analysis functions that GIS offers. Obviously, the integration of GIS and simulation models, together with the necessary databases and expert systems should make more powerful, easy-to-use and easy-to-understand risk information systems (Fedra, 1998).

The main focus in this thesis is to prove the usefulness of refining the PERA by explicitly accounting for the spatial (and temporal) characteristic of the exposure, effects and risk. This idea is further explored in Part 5.

2.4. Proposal Improved Probabilistic Ecological Risk Assessment Framework & Outline of this Thesis

Based on the conventional ERA and the first developments in the PERA, new overviews of geo-referenced PERA are proposed in this section. This section is also the guide that links all chapters of this dissertation. First, an overview of the PERA framework is given. Second, geo-referenced PERA is situated as one particular level of detail in a tiered framework. These overviews are developed to provide clarity and structure in the current set of several statistical methods and mathematical models in PERA literature.

Two different approaches can be used to determine the Exposure Concentration Distribution (ECD) and the Species Sensitivity Distribution (SSD). Data from either measurements in the environment or toxicity tests can be used directly (see Figure 6, right side). The alternative is to use prediction or extrapolation models, especially in case of new chemicals (see Figure 6, left side). Examples of exposure models are GREAT-ER (model for point-source emissions), PRZM-EXAMS (model for pesticide fate), E-USES… Examples of effect models are QSARs (Quantitative Structure Activity Relationships), DebTox, BLM (predicts toxicity based on bioavailability)… In practice, exposure models are more common for ECD determination and effect data are more common for SSD determination. Obviously, this may shift in the future.

A distinction is made between statistical methods for

- Propagating uncertainty and variability through mathematical models (open arrows in Figure 6): used in the effects and exposure modelling (discussed in Chapter 3.1) but also in the risk characterisation (discussed in Chapter 4.1).
- Characterising data uncertainty and variability (full arrows in Figure 6): used in the ECD and SSD estimation based on measured data (discussed in Chapters 3.2, 3.3 and 3.4) but also for estimating variability and uncertainty of input parameters and variables needed for exposure and effect modelling.

In literature, researchers focus too much on either ECD or SSD determination in a PERA. This is also reflected by the structure of the previous sections. However, the same statistical techniques are used in the ECD or SSD determination. Therefore, the structure of this dissertation will be methodology driven. It is true, however, that most studies on probabilistic exposure assessment deal with uncertainty and variability propagation because models are used more often (see Figure 6, top left). And most studies on probabilistic effects assessment deal with uncertainty and variability estimation (see Figure 6, bottom right).

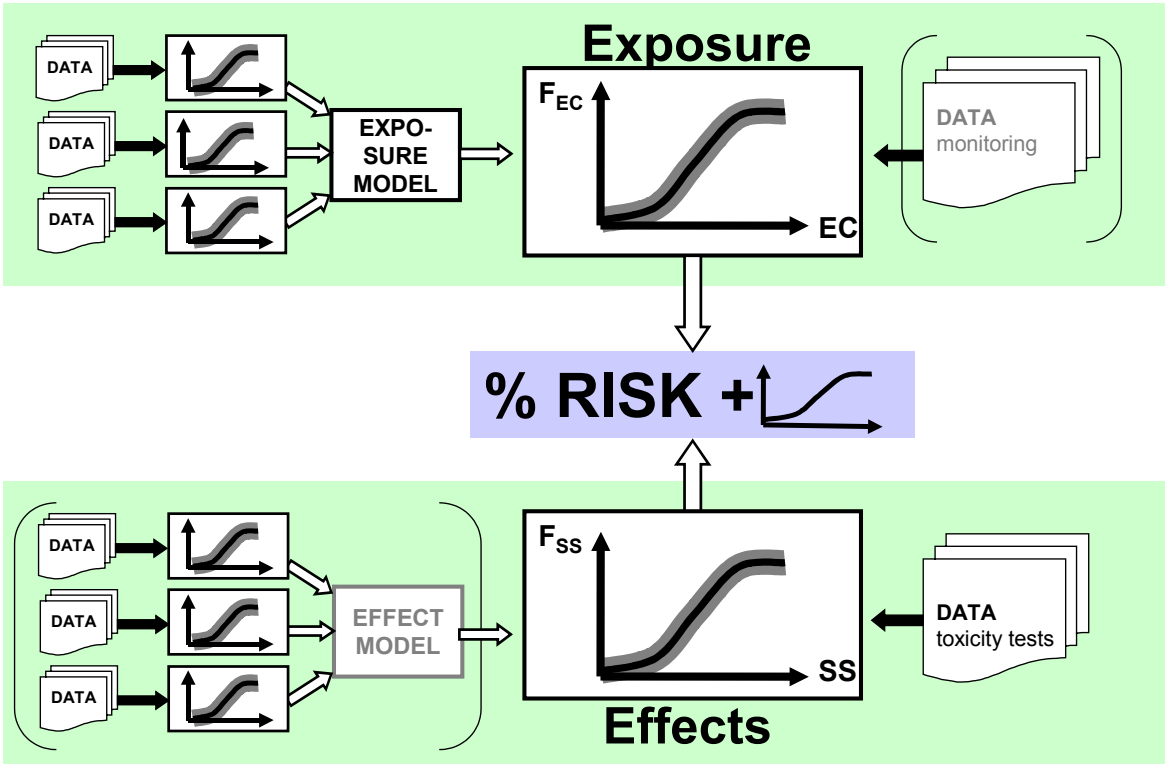


Figure 6: Proposed framework of probabilistic ecological risk assessment (EC: Exposure Concentration, SS: Species Sensitivity)

To provide clarity and structure, Figure 7 shows an overview of several tiers (of different level of detail) of PERA. In the top panel, the conventional ERA is shown. A random variable (be it the exposure concentration or the species sensitivity) is considered as a crisp value. Uncertainty is partly ignored, partly considered through safety or assessment factors. The second panel represents the PERA. PERA is an extension of the conventional approach since both the inherent variability and uncertainty (shown as a grey band) is explicitly quantified and assessed. However, all types of variability are eventually lumped in a single distribution.

In the next panel/tier, the spatial variability is explicitly accounted for. The random variable X is considered for every spatial location (called here geo-referencing). As a result, the variability distribution no longer includes spatial variability but only temporal and other types of variability. This leads to a large number of geo-referenced distributions but with smaller variances. The geo-referencing tier is dealt with in Chapter 5.1.

Time-referencing would further increase the level of detail and realism, as time-specific information would be accounted for. This is represented in the lower panel of Figure 7. Time related information could be formatted in two ways in an attempt to capture the temporal variability. First, time series can be used as such or, second time series can be summarised into concentration-duration-frequency surfaces. These surfaces are three-dimensional plots with on the three axes the concentration, the

duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. Time-referencing or dynamic ERA is not part of the aim of this thesis. Rather, some preliminary developments are discussed and explored in Chapter 5.2.

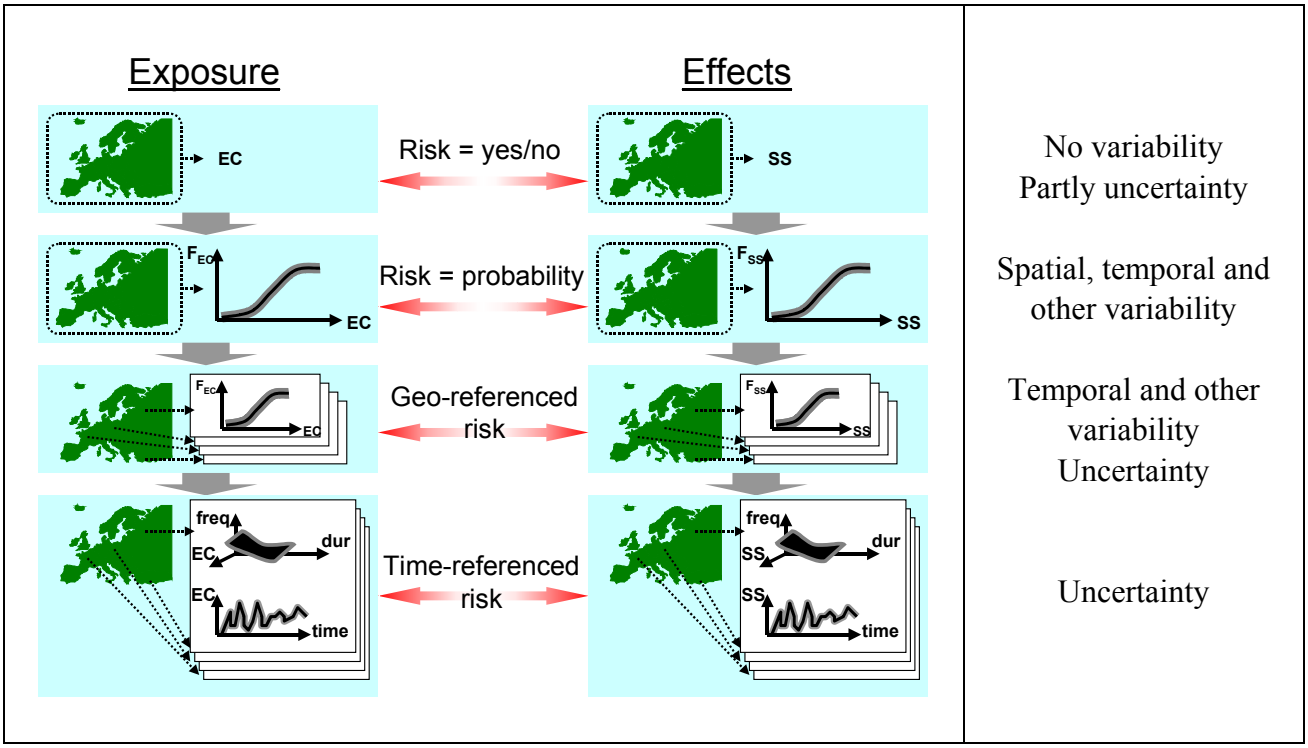


Figure 7: Different tiers in ERA (Ecological Risk Assessment), (EC: Exposure Concentration, SS: Species Sensitivity)

Part 3

-

Uncertainty and Variability Propagation and Estimation

Chapter 3.1

-

Uncertainty & Variability Propagation

Part of this chapter was published in:

Rousseau D., Verdonck F. A. M., Moerman O., Carrette R., Thoeye C., Meirlaen J. & Vanrolleghem P. A. 2001. Development of a risk assessment based technique for design/retrofitting of WWTPs. *Water Science & Technology*, 43(7), 287-294.

Chapter 3.1

Uncertainty & Variability Propagation

In Probabilistic Ecological Risk Assessment (PERA), propagation of uncertainty and variability is needed when modelling is part of the assessment. The open arrows in Figure 6 of Part 2 indicate where propagation is needed: to model the Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD) and also for risk characterisation. In a propagation method, information on the input uncertainty and variability is passed on through a model to the output.

3.1.1. Selection of the Propagation Technique

There are a variety of ways to propagate information about variability or uncertainty through a model. Here, three main analytical and simulation (approximation) methods are given: first-order error analysis, probability bounds analysis and Monte Carlo analysis. A good reference with an extensive overview of techniques is Cullen & Frey (1999).

First-order error analysis is a method for propagating uncertainty or variability in the random parameters of a model into the model predictions. The method is based on first (and higher) order approximations of the model by Taylor series expansions of the model's equations. These methods are also called methods of moments because they utilise the mean, variance and higher moments of probability distributions. Moment methods, especially the first-order kind, are well suited to simple linear models, or ones that can be linearised around operating points without substantial errors. The accuracy of the method decreases as the model becomes more nonlinear. Complicated models that consist of a large number of equations (like large exposure models) cannot be evaluated using first-order error analysis.
The advantage of using first-order error analysis is that the relative contribution of each uncertain variable to the output uncertainty is known. This relative contribution can be used to prioritise data collection efforts to reduce uncertainty in the parameter. Another advantage is that the exact input

distribution is not required by statistical theory. The variances are combined through the equation, regardless of the input distribution. Therefore, the investigator does not need to spend much time worrying about the exact input distribution. The investigator only needs to know the variance of the random parameter (in the proper units), for the underlying theory to hold. However, the fact that only information regarding central moments is considered is also a drawback of these methods. Information regarding the tails of each input distribution, for example, is not specifically considered. Therefore, the selection of a probability distribution based upon the moments of the model output may not properly capture effects at the tails of the distribution, although it may be adequate for characterization of the central tendencies Cullen & Frey (1999).

Probability bounds analysis is a related strategy for making probabilistic inferences in the face of uncertainty. It is a method for computing bounds on the distribution of a sum, product, or arbitrary mathematical expression, given only bounds on the distributions. Probability bounds analysis gives the same answer as interval analysis does when only range information is available. It also gives the same answers as Monte Carlo simulation does when information is abundant enough to precisely specify input distributions and their dependencies. Thus, it is a generalization of both interval analysis and probability theory (Ferson *et al.*, in press).
It is often possible to obtain bounds on a quantity. Moreover, bounds are often easier to compute than approximate estimates, which, in contrast, routinely require the solving of integrals. This simplicity of calculation extends to the combination of bounds. Some limitations of probability bounds analysis are that (1) bounds on a distribution cannot show what distribution is most likely within the bounds, (2) maintaining the optimality of answers may be hard when there are repeated variables or when there is a lot of empirical information about complex dependencies among the variables and (3) all outputs must be expressed in terms of cumulative probability (Ferson *et al.*, in press).

Finally, sampling methods are a good alternative. A very common sampling method for propagating variability or uncertainty is Monte Carlo simulation, in which random samples of parameters are selected according to their respective assigned distributions. Sampling methods have a high computational effort; but, with computers being more sophisticated and more easily affordable nowadays, this is hardly a critical issue anymore. The Monte Carlo method has distinct advantages because it is not limited to any particular set of assumptions about the nature of errors or their magnitude. The Monte Carlo simulation may not always be an efficient method for estimating error bounds on a prediction, but it may well be the most effective approach for exploring the mechanisms involved in propagating uncertainty and the factors involved in minimizing and controlling these uncertainties. Depending on whether variability, uncertainty or both variability and uncertainty need to be propagated, a first or respectively second order Monte Carlo simulation will be used. In this chapter, both techniques will be discussed and applied to some cases in the environmental risk assessment field. Since Monte Carlo is a well-known and studied technique, the innovating aspect will be the specific applications.

3.1.2. First Order Monte Carlo Simulation

In this first section, the first order or one-dimensional Monte Carlo is discussed. It propagates either uncertainty or variability if interpretation problems need to be avoided.

3.1.2.1. Introduction

Not all current exposure and effect modelling account for uncertainty or variability. The added value of accounting for uncertainty and variability is discussed in detail in Part 2. Sometimes, Monte Carlo analysis is used to propagate uncertainty or variability. In practice, Monte Carlo is sometimes wrongly applied or the output is insufficiently interpreted. Let us take GREAT-ER, an exposure model with a Monte Carlo engine, as an example. More information on the GREAT-ER model and its applications can be found in Part 2 and Chapter 5.1. GREAT-ER predicts exposure concentration distributions (ECDs). These distributions are difficult to interpret because in the software, probability distributions cannot be assigned to all input variables and parameters (even very sensitive parameters as the chemical consumption rate) due to software limitations. Furthermore, uncertainty and variability are mixed up. This last issue is further discussed in 3.1.3.

In this chapter, two case studies were performed that illustrate Monte Carlo applications in the risk assessment field. The first one can be situated on the exposure side, more in particular modelling of the uncertainty of the ECD of the effluent of a Waste Water Treatment Plant (WWTP). Here, there is a need to determine a confidence interval on the probability of exceeding the legal effluent standards of a WWTP. The second one can be situated at the effects side, more in particular on the modelling of the temporal variability of the bioavailability of a metal determining the species sensitivity.

The goal of this section is two-fold. First, the Monte Carlo technique needs to be programmed into a usable set of flexible software tools. For this, a proper methodology to account for correlations and a proper sampling technique needs to be selected. And second, the tool needs to be applied to the two discussed case studies. For this, it will also be shown how to correctly interpret the Monte Carlo output.

3.1.2.2. Monte Carlo Simulation

In the next paragraphs, the Monte Carlo methodology is described in detail. Specific attention is given to the pseudorandom number generator, how to simulate the probability distributions, several

sampling schemes and several ways of modelling correlations. The Monte Carlo was programmed in C++.

Note that this general Monte Carlo methodology is in essence also the basis for other simulation techniques such as the bootstrap (this technique estimates uncertainty and variability in a data set) that will be discussed in upcoming chapters.

3.1.2.2.1. The Monte Carlo Simulation Technique

The Monte Carlo approach was developed by Stanislaw Ulam and John von Neumann to simulate probabilistic events for military purposes in 1946 (Frey & Li, 2001). The Monte Carlo principle is illustrated in Figure 1. The method has been extensively described in literature ((Cullen & Frey, 1999), (Vose, 1996), (Hammersley & Morton, 1964)). For each model input that is considered to be a random variable, a probability distribution is specified. One random sample from each input distribution is selected, and the set of samples is entered into the deterministic model. The model is then solved, as it would be for any deterministic analysis. The model results are stored and the process is repeated until the specified number of model iterations (called here shots) is completed. Using Monte Carlo techniques, it is therefore possible to represent uncertainty in the output of a model by generating sample values for the model inputs, and running the model repetitively. Instead of obtaining a discrete number for model outputs as in a deterministic simulation, a set of output samples is obtained (Cullen & Frey, 1999).

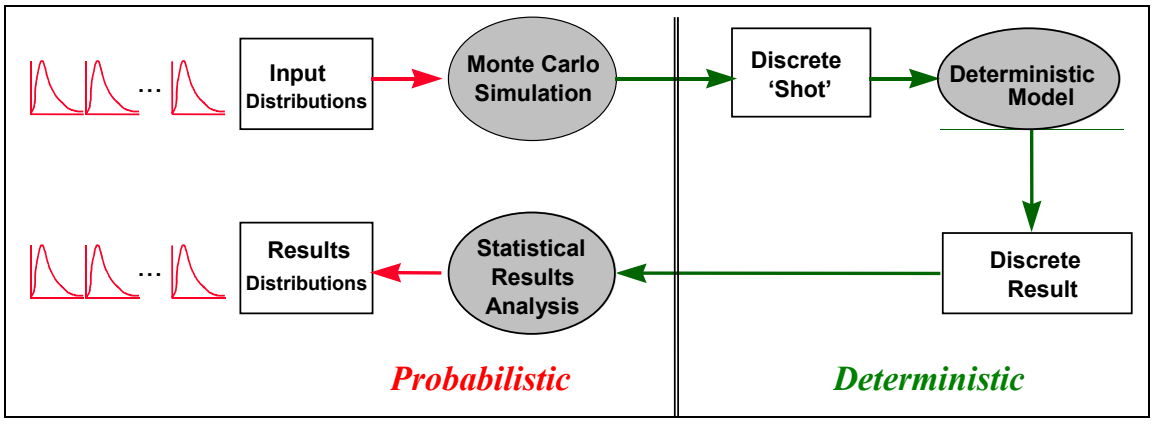


Figure 1: Principle of Monte Carlo simulation

Some guiding principles of good practice for use of the Monte Carlo simulation can be found in Vose (1996), Burmaster & Anderson (1994) and EPA (1997).

3.1.2.2.2. Pseudorandom Number Generator

Simulation methods for dealing with uncertainty are often based upon the use of a random number generator. The random number generator should come as close as possible to the ideal of generating a series of truly independent random numbers, i.e. there must not be any correlation between successive random numbers. Computers essentially generate "pseudo"- random numbers, which explains why we would be able to predict a random number if given the preceding random numbers. In reality, therefore, these computer-generated pseudo-random numbers are not truly random, but they can be treated as such.

Pseudorandom number generators typically have a 'random seed' or 'starting value' as an input. By changing the seed, one can change the sequence of random numbers obtained. However, if the same seed is used in two or more analyses, then the same set and sequence of random numbers would be obtained. This is particularly useful for checking the implementation of the Monte Carlo and the model output.

The random number generator of Law & Kelton (1991) will be used here. Law & Kelton describe several tests to which a random number generator can be subjected to ascertain how well the generated numbers do (or can) resemble values of true independent and identically uniformly distributed random variates (between zero and one). There are two quite different kinds of tests. Empirical tests are the usual kinds of statistical tests and are based on the actual numbers produced by a generator. Theoretical tests are not tests in the statistical sense, but use the numerical parameters of a generator to assess it globally without actually generating any numbers at all.

3.1.2.2.3. Simulating Probability Distributions

Random samples of model input parameters are selected according to their respective assigned probability distributions. This can be done using various methods. Here, the inverse transform method was used. Monte Carlo simulation requires the generation of uniformly distributed random numbers between 0 and 1. This uniform distribution must be transformed into the assigned probability distribution of the input parameter. This is achieved by using an inverse cumulative distribution function. Indeed, for any given probability distribution (as in the top left plot in Figure 2), it is possible to construct a cumulative distribution function (as in the top right plot in Figure 2). The inverse cumulative distribution function (as in the bottom right plot in Figure 2) has an abscissa with values ranging from zero to one, and an ordinate with values representing possible outcomes for the random variable of interest. Thus, uniformly distributed random numbers may be used to represent the percentile of the random variable for which a sample is to be generated. The sample values for the random variables are calculated using the inverse cumulative distribution function transformations based on the randomly generated samples. For the developed C++ tool, inverse

cumulative distribution functions were programmed for the uniform, triangular and the generalised extreme value distribution.

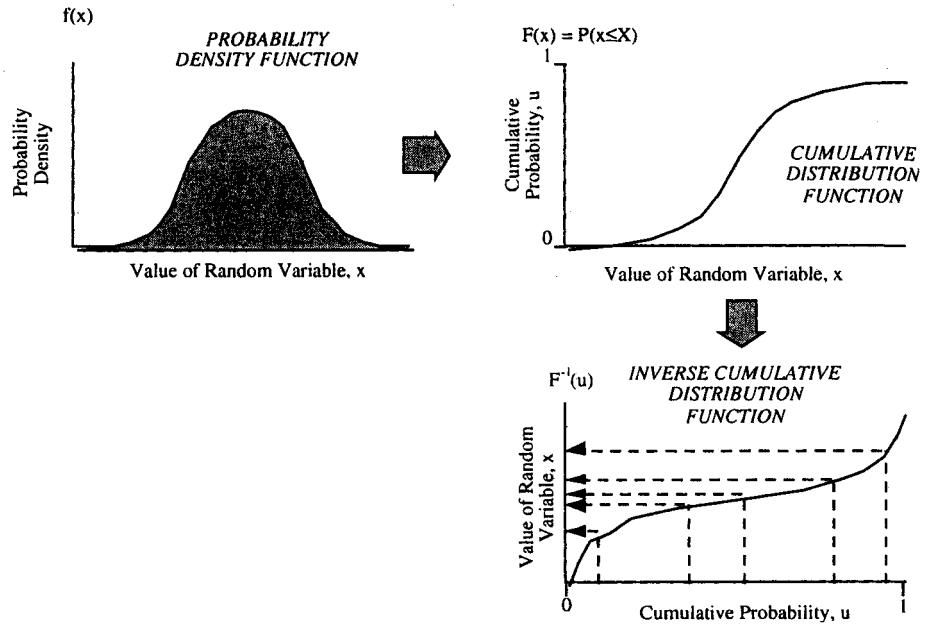


Figure 2: Sampling probability distributions using the inverse transform (Cullen & Frey, 1999)

Another method was used to generate random numbers for the normal distribution. The method can be found in Milton & Stegun (1970). Random samples of a lognormal distribution are obtained by transforming the lognormal distribution to the corresponding normal distribution, taking a normal sample from the latter, and then converting the obtained value back to the lognormal distribution. The transformation formulas between the parameters of a normal and a lognormal distribution are:

$$mean_{\log} = \exp(\mu + \frac{\sigma^2}{2})$$

$$stddev_{\log} = mean_{\log} \cdot \sqrt{\exp(\sigma^2) - 1}$$

$$\mu = \ln(mean_{\log}) - \frac{1}{2} \cdot \ln\left(\frac{stdev_{\log}^2}{mean_{\log}^2} + 1\right)$$

$$\sigma = \sqrt{\ln\left(\frac{stdev_{\log}^2}{mean_{\log}^2} + 1\right)}$$

where	$mean_{\log}$	the mean of the (lognormally distributed) data
	$stdev_{\log}$	the standard deviation of the (lognormally distributed) data
	μ	the mean of the log-transformed data
	σ	the standard deviation of the log-transformed data

Most statistical software packages have built in random number generators for a number of distribution types.

3.1.2.2.4. Sampling Schemes

A short overview of the most frequent sampling techniques is given here ((Cullen & Frey, 1999); (Vose, 1996)): crude Monte Carlo sampling, Latin hypercube stratified sampling, directional simulation with importance sampling and quasi-Monte Carlo sampling.

Crude Monte Carlo sampling

Crude Monte Carlo simulation is based on random sampling from the joint frequency distribution of the input variables as described in 3.1.2.2.3. To obtain a relatively close approximation of the theoretical frequency distribution, it is therefore necessary to take a large (huge) number of shots (e.g. 10000) from the input distribution. This technique results in very inefficient computations but if the number of shots is large enough, this technique will give accurate results.

Latin Hypercube stratified sampling

To reduce the number of Monte Carlo simulations, so-called stratified sampling techniques have been developed. These methods allow for a more efficient exploration of the input parameter space by dividing the input distributions into intervals, and sampling from these intervals rather than from the whole frequency distribution (see Figure 3). The most commonly used technique is known as the Latin Hypercube sampling. A much lower number of shots are needed to obtain the same accuracy as the crude Monte Carlo.

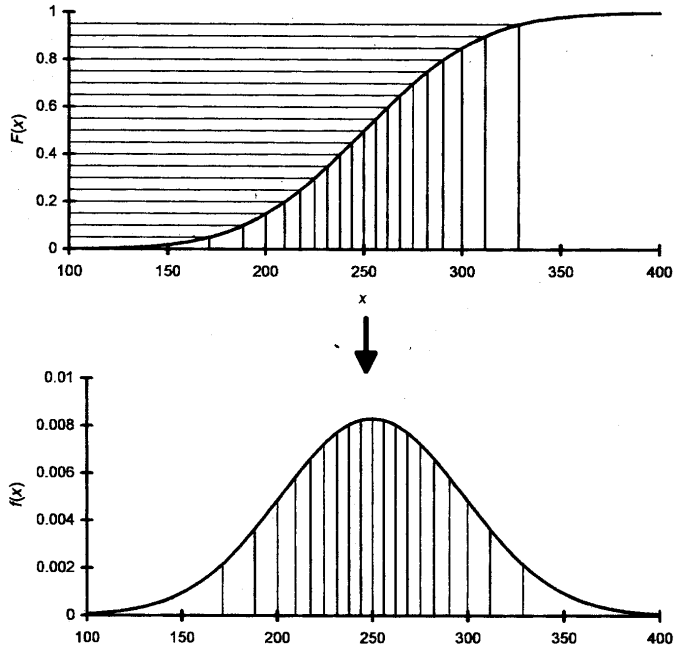


Figure 3: Latin Hypercube Sampling

But as with crude Monte Carlo sampling, the accuracy of the output is a function of the number of samples and no rules are available for choosing the number of shots. Only some suggestions (e.g. number of shots equal to 4/3 or twice the number of uncertain input parameters) are available and,

therefore, each modeller must check for convergence for the model and problem in question (Melching, 1995) (see also 3.1.2.2.6).

Directional simulation with importance sampling

Directional simulation with importance sampling, sometimes called weighted sampling, is mainly used when one is interested in a particular part of the distribution, for instance long tails (see e.g. Portielje *et al.* (2000)). In probabilistic risk assessment, distribution tails are very important. The main idea is to execute a heavier sampling in the regions that draw particular attention. This can be done in several ways. One way is to use Latin hypercube sampling with equiprobable intervals but to take a large number of samples from intervals of interest (e.g. intervals that represent the high percentiles of the input frequency distribution). After doing so the output distribution has to be corrected with a weight factor, but the prediction of the part of interest in the distribution is much more accurate.

Quasi-Monte Carlo sampling

Quasi-Monte Carlo techniques replace pseudo-random sequences with low-discrepancy sequences which have a more uniform behaviour (Fox, 1999). In a certain sense, Quasi-Monte Carlo methods combine the advantages of Monte Carlo and uniform lattice methods. In particular, fewer quasi-random samples are needed to achieve a similar level of accuracy as obtained by using pseudo-random sequences.

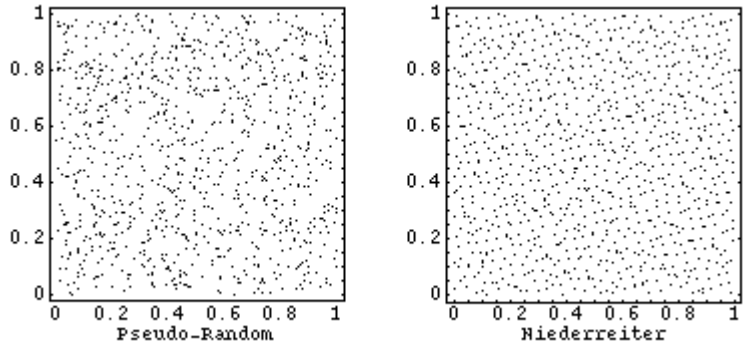


Figure 4: Pseudo-Random Monte Carlo sampling (left) versus Quasi Monte Carlo sampling (right)

3.1.2.2.5. Simulation of Correlations

Vose's (1996) 'cardinal rule of risk analysis modelling' is "Every iteration of a risk analysis model must be a scenario that could physically occur". If e.g. a high river flow is selected ad random, then

a low temperature will be more likely than a large one if the river flow is highly negatively correlated with the temperature. Therefore, one of the restrictions that must be placed on the model is to recognise inter-dependencies between its uncertain components. It is possible to simulate jointly distributed random variables in which correlations may exist.

Before going any further, it is worth explaining dependency, correlation and regression (Vose, 1996). Consider *X* and *Y* as two correlated random variables.

A dependency relationship in risk analysis modelling is where the sampled value from one variable (called the independent) has a statistical relationship that approximately determines the value that will be generated for the other variable (called the dependent). Its chief difference to correlation is that it presumes a causal relationship.

Correlation is a statistic used to describe the degree to which one variable is related to another. Pearson's correlation coefficient is given by:

$$r = \frac{\mathrm{cov}(X,Y)}{s(X)\cdot s(Y)}$$

where cov(*X*,*Y*) is the covariance between random variables *X* and *Y* and *s(X)* and *s(Y)* are the sample standard deviations. Correlation can be considered to be a normalised covariance between the two data sets: dividing by the standard deviation of each data set produces an unitless index between -1 and 1. Correlation is frequently used alongside a regression analysis to measure how well the regression line explains the observed variations of the dependent variable.
This correlation statistic should not be confused with Spearman's rank order correlation coefficient also used in this context. It uses the ranking of the data, i.e. what position (rank) the data points take in an ordered list from the minimum to the maximum values, rather than the actual data themselves. It is therefore independent of the distribution shapes of the data sets and allows the integrity of the input distributions to be maintained. This makes the Spearman's rank order correlation more interesting than Pearson's correlation. Thus, rank correlation is a measure of the strength of the monotonic relationship between two variables. Spearman's *r* is calculated as:

$$r = 1 - \left(\frac{6\sum(\Delta R)^2}{n(n^2-1)} \right)$$

where *n* is the number of data pairs and ΔR is the difference in the ranks between data within a pair. A disadvantage of rank order correlations is the difficulty in selecting the appropriate correlation coefficient in case no observations are available.
There are mathematical constraints associated with correlations. For instance, one variable cannot be strongly positively correlated with each of two variables that are themselves strongly negatively

correlated. Such constraints can be summarised by saying the matrix of correlations must always be a positive (semi) definite matrix. Many homegrown and even some commercially available software packages for Monte Carlo simulation do not check that the user's input satisfies this condition. If the correlations entered are the result of coherent empirical studies, this will not be a problem though. However, if results from different studies are mixed or hypothetical values for correlations are used, it may be important to check that the input corresponds to a feasible correlation matrix (Warren-Hicks & Moore, 1995).

Regression is a mathematical technique used to determine the equation that relates the independent and dependent variables with the least margin of error. A line that passes as closely as possible through the data points represents a regression equation. The most common technique is that of least squares linear regression. This objectively determines the straight line ($Y = aX + b$) such that the sum of the squares of the vertical deviations of the data points from the line is a minimum.

The presence of moderate to strong correlations will have little effect on the central portions of the output distributions but may have larger effects on the tails of the output distributions (Burmaster & Anderson, 1994). In probabilistic risk assessment, distribution tails are very important. Therefore, a proper methodology to account for correlations needs to be selected. Several techniques exist for modelling correlations in Monte Carlo simulation. Three of them were investigated in more detail. Method 1 is an exact correlation corrected sampling method. Method 2 is based on rank order correlation and correlation matrices and is an approximation technique. The third method offers more accurate but correspondingly more time-consuming and data-consuming techniques for dependency modelling. An example will be used to illustrate the 3 methods. Two variables Y and X are normally distributed with a mean and a standard deviation of respectively 200 and 160 for variable Y, and 5 and 2 for variable X. X and Y are linearly correlated according to the regression equation $Y = 80*X - 200$ with a correlation coefficient r of 80% ($r^2 = 0.64$).

Method 1: Restricted Pairing Techniques (Cullen & Frey, 1999)

Conditional distributions are important in many model applications to environmental systems. Sampling of a dependent input variable Y from a conditional distribution is based on the Rosenblatt-transformation. First, the independent variable is sampled. From a realization X' of the independent variable X, the conditional distribution function (given that $X = X'$) of a dependent variable Y is calculated. From this conditional distribution function, the dependent variable is now sampled. Conditional distribution functions can be estimated from subsets of data of the dependent variable for each of a number of classes of the independent variable.

If the marginal distribution of X is normally distributed, then the conditional distribution of Y has a mean (expected value) and a standard deviation of Cullen & Frey (1999):

$$E(Y \mid x) = \mu_Y + r\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$\sigma_{Y|x} = \sigma_Y\sqrt{1 - r^2}$$

where $E(Y|x)$ is the expected value of Y conditional on x
μ_Y is the mean of Y
r is the correlation coefficient
σ_Y is the standard deviation of Y
σ_X is the standard deviation of X
$\sigma_{Y|x}$ is the standard deviation of Y conditional on x

In the example, a random shot for X is 7.2528. This is the 87th percentile. The conditional distribution of Y has a mean (expected value) and a standard deviation:

$$E(Y \mid x) = 200 + 0.64 \cdot \frac{160}{2} \cdot (7.2528 - 200) = 361.1972$$
$$\sigma_{Y|x} = 160 \cdot \sqrt{1 - 0.8} = 71.55418$$

A random sample from this conditional distribution with mean $E(Y|x)$ and standard deviation $\sigma_{Y|x}$ is then 317.3632.

Note that this exact method can easily be generalized to three or more normal variables X, Y, Z, … The correlated random variables are then found as:

$$\begin{pmatrix} X \\ Y \\ Z \\ \dots \end{pmatrix} = V \cdot \Lambda^{1/2} \cdot Q$$

where V are the eigenvectors of the variance-covariance matrix
Λ is a diagonal matrix containing the eigenvalues of the variance-covariance matrix
Q is a vector of standard normal random numbers ($Q \sim N(0,I)$)

Method 2: Method of Iman & Conover (1982)

This method of Iman & Conover (1982), used to generate rank order correlated input distributions, is often applied in literature (Cullen & Frey, 1999), (Janssens *et al.*, 1992), (Vose, 1996), (Warren-

Hicks & Moore, 1995). It is not an exact but an approximative method. The main advantage of this method is that a combination of several distribution types can be used. The technique is a two-step process. First of all, a set of n 'scores' is generated for each distribution to be correlated, where n is the number of shots that are to be run. Then these 'scores' are rearranged together so their ranks produce the desired rank order correlation. In the second step, the distributions to be correlated are all sampled n times and these sampled values are ranked. The ranks are then matched to the 'score' ranks from the first step to produce the sets of values that will be used for each shot in the simulation. Random samples are not determined per shot, but as a whole set.

For the example, since random samples are not determined per shot, a number of Monte Carlo shots have to be specified, let's say 5. Several technical steps need to be followed (not discussed here in detail): sampling N values randomly from the standard multivariate normal distribution (uncorrelated)

$$\begin{pmatrix} 0 & -0.29 & 0.99 & -1.69 & 0.39 \\ 3.85 & 3.42 & 1.52 & 1.82 & 2.53 \end{pmatrix}^T,$$

transform these uncorrelated multivariate normal samples to correlated samples by multiplying with the transformation-matrix found from a Cholesky-decomposition of the correlation matrix

$$\begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0{,}8 \\ 0 & 0{,}6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0{,}8 \\ 0 & 0{,}6 \end{pmatrix}^T,$$

resulting in

$$\begin{pmatrix} 0 & -0.29 & 0.99 & -1.69 & 0.37 \\ 2.31 & 1.82 & 1.70 & -0.26 & 1.83 \end{pmatrix}^T,$$

replacing the samples by their rankings, i.e. the smallest value gets rank 1, the largest gets rank N etc resulting in (this matrix is in fact part of an N x N permutation matrix)

$$\begin{pmatrix} 3 & 2 & 5 & 1 & 4 \\ 5 & 3 & 2 & 1 & 4 \end{pmatrix}^T,$$

sampling N values randomly from the uniform distribution (uncorrelated)

$$\begin{pmatrix} 0.17 & 0.52 & 0.79 & 0.88 & 0.96 \\ 0.25 & 0.27 & 0.73 & 0.87 & 0.98 \end{pmatrix}^T$$

and rearrange the sorted uniform samples, on basis of the random permutation matrix

$$\begin{pmatrix} 0.79 & 0.52 & 0.96 & 0.17 & 0.88 \\ 0.98 & 0.74 & 0.27 & 0.25 & 0.88 \end{pmatrix}^T.$$

These are 5 random shots from a uniform distribution, corrected according to their correlation. Following the cumulative normal distribution of *X* and *Y*, the 5 random correlated samples for *X* and *Y* can now be calculated by taking the inverse cumulative function of *X* for the first row and the inverse cumulative function of *Y* for the second row (as discussed in 3.1.2.2.3).

Method 3: The Envelope Method (Vose, 1996)

The envelope method offers a more flexible way to model dependencies that is both intuitive and easy to control. It models the logic that the value of the independent variable is actually statistically determining the value of the dependent variable (see Figure 5). Its shortfall is that it requires considerably more effort than the two previous methods and is, therefore, only really used when the dependency relationship is going to produce a significant effect on the outcome of the model.

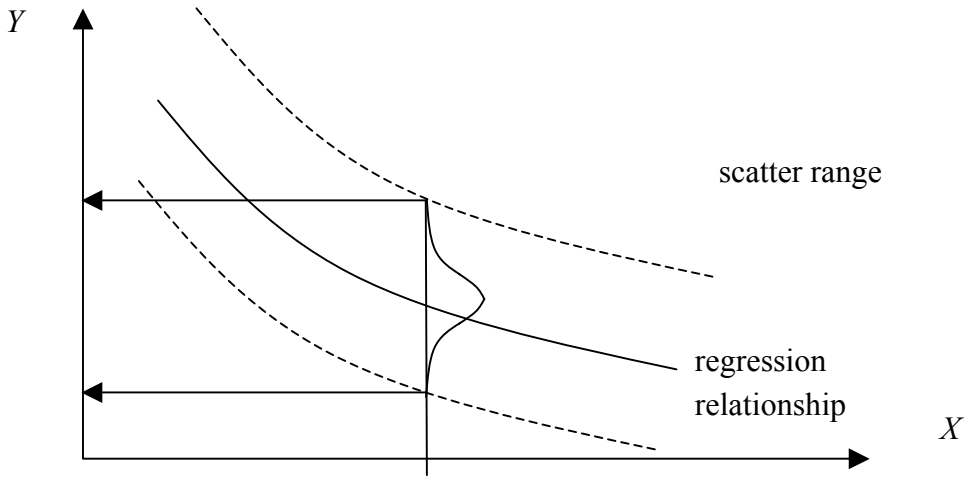


Figure 5: Regression relationship between two variables with uncertainty band

The regression relationship can be linear, logarithmic, polynomial, power, exponential… The scatter range can be described by a distribution, e.g. normal, lognormal, uniform, and triangular… This distribution type can be different from the distribution of *X*.

The method works as follows: a random sample of *X* is taken. *Y* is sampled from the scatter distribution around the regression in *X* (see arrows in Figure 5). Note that for multi-correlated variables, this method requires a tree structure to simulate correlations. Suppose for example, there are four input parameters *X*, *Y*, *Z* and *Q*, which all are correlated with each other. Figure 6 presents one possible way to simulate their correlations. *X* is correlated with *Y* and *Z*. And, *Z* is correlated with *Q*. But this does not mean that *X* and *Q* are uncorrelated. There are no general rules to determine such a correlation tree. This is considered as a disadvantage of the envelope method. It requires expert judgement to know which tree is most suited. In the case studies, it was strived to put the strongest correlations to the top and the weaker correlations near

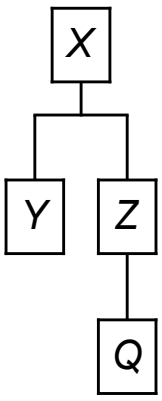


Figure 6: Example of correlation tree

the bottom of the tree.

The possibility of implementing a causal relationship is an advantage when a particular well-known relation exists between two variables. But, when the relationship is not causal or when the relationship is difficult to determine, samples below the minimum or above the maximum (extrapolation) can induce severe errors. In this case, it would be safer just to use a rank correlation coefficient as in method 1 and 2 instead of a doubtful regression equation.

In the example, a random sample for *X* is 7.2528. A sample for *Y* is now determined as a random sample from a normal distribution with mean 7.2528*80 - 200 = 380.224 and standard deviation 120. The sample is 263.050.

Methods comparison

1000 shots were generated for each method. The results are shown in Figure 7. All methods result in about the same scatter band. Method 2 will be used in the second case study and method 3 will be used in the first case study.

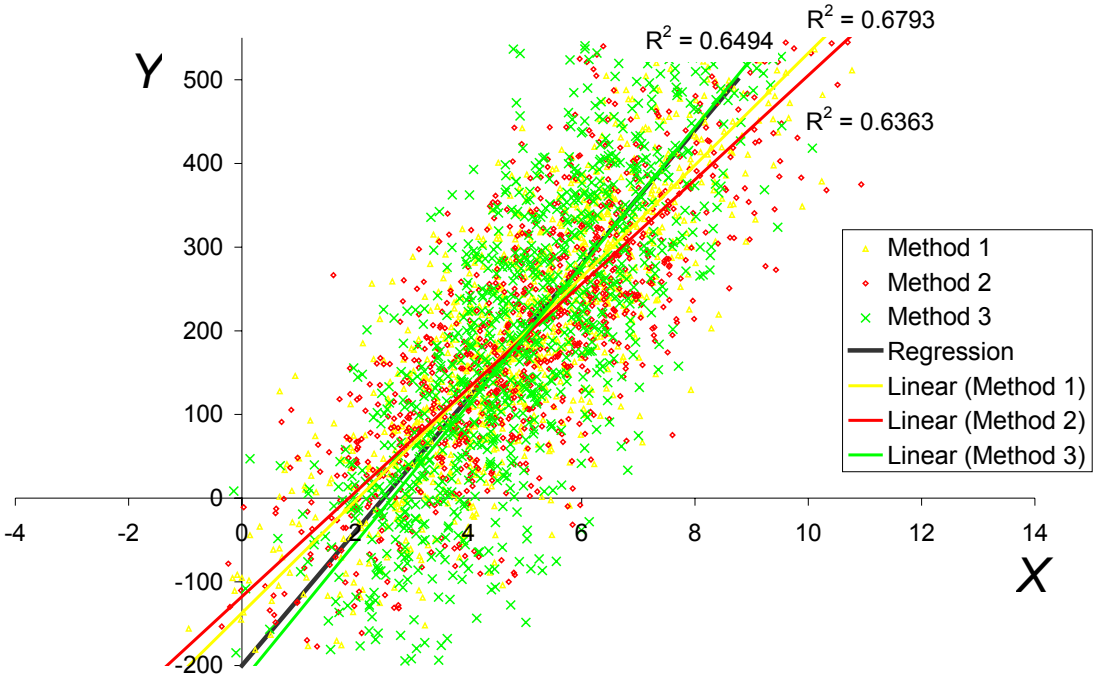


Figure 7: 1000 shots for two correlated variables according to 3 correlation sampling methods (r = 0.80)

3.1.2.2.6. Verification of the Necessary Number of Monte Carlo Shots

There are a number of approaches that can be used to select the number of shots to reach certain accuracy. One approach is to plot the Monte Carlo output versus the number of shots. Figure 8 shows such a graph for a Monte Carlo output (in this case the total number of exceedances of a standard). Only 100 shots are needed to stabilise the simulation outputs. The small instabilities at larger shots have a numerical cause.

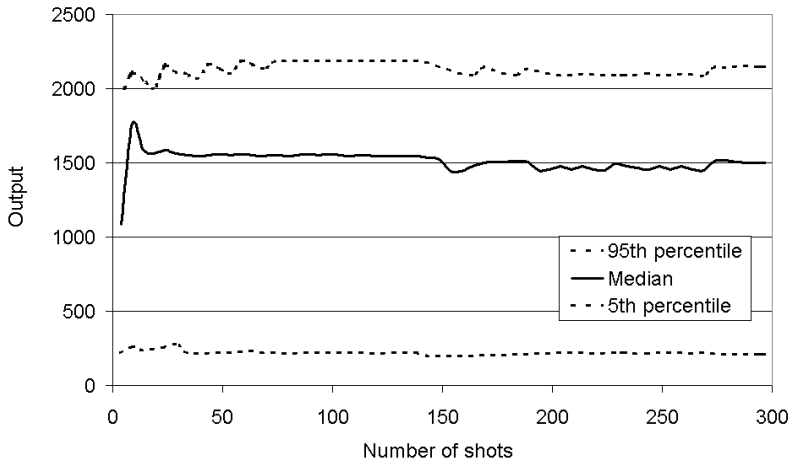


Figure 8: Convergence of the Monte Carlo output (for the exceedance frequency of the first case study)

These plots were made in each Monte Carlo study of this PhD dissertation to assess the minimum number of shots.

3.1.2.3. Case Studies

In this section, two case studies demonstrate the use of uncertainty or variability propagation using Monte Carlo simulation. The first case study on uncertainty propagation was part of a project funded by Aquafin NV, the company responsible for the design, construction, operation and financing of the necessary infrastructure for sewage treatment in Flanders (Belgium). The second case study on variability propagation was part of the MSc. thesis of De Laender (2003).

The main focus in both case studies is to illustrate the Monte Carlo simulation and the interpretation of the output. Minor attention is given to the accuracy of the model and its input parameters or variables.

3.1.2.3.1. Uncertainty Propagation in Waste Water Treatment Plant Modelling

River water quality in Flanders (Belgium) has been dramatically bad during the past decades, because of the high degree of urbanisation, the industrial and agricultural pollution and insufficient basic treatment infrastructure. Almost no watercourses even met the lowest criteria set out in river master plans. One of the challenges Aquafin was facing is to upgrade the patrimonium of old municipal Waste Water Treatment Plants (WWTPs). These plants need to be retrofitted towards strict phosphorus and nitrogen removal consents. In 1991, when the European Directive for urban waste water treatment 271/91 for sensitive areas to eutrophication was introduced, only one-quarter of the Flemish waste water was treated in a WWTP. Moreover, the existing WWTPs did not comply with the present norms.

Within the currently followed design/retrofit-procedure, deterministic dynamic models are used to evaluate different renovation scenarios on their merits. One of the remaining issues when dealing with these deterministic models is the degree of uncertainty linked to their predictions. In other words to what extent can model predictions be taken for reality? The combination of probabilistic modelling techniques with the currently available deterministic models (steady state or dynamic models) could provide the answer needed. By building a probabilistic shell around the deterministic models one could quantify the output uncertainty due to input uncertainty.

The concrete goal of this case study is to determine the probability of exceeding the legal effluent standards of a WWTP. This percentage of exceedance should be accompanied by a confidence interval indicating the inherent uncertainty of influent characteristics and model parameters. Estimation of uncertainty allows WWTP managers (and operators) to choose whether to actively take measures or to conduct additional research.

In literature, others also reported on the design of WWTP by combining a Monte Carlo simulation with a WWTP model. An overview is given in Table 1. Most authors don't distinguish uncertainty and variability or account insufficiently for correlations.

In this case study, the model consists of two parts (see Figure 9). The first part is a WWTP model. A denitrifying WWTP model inspired by the benchmark model described by Spanjers *et al.* (1998) was implemented in the WEST modelling and simulation software (Vanhooren *et al.*, 2003). A more detailed description of the WWTP design can be found in Rousseau *et al.* (2001) and Rousseau *et al.* (2000). The biological treatment was simulated by means of the Activated Sludge Model N° 1 (ASM1) of Henze *et al.* (1987). Secondary sedimentation was simulated by a pointsettler. The second part of the model is a statistical analysis system in which the time series output of the WWTP model is interpreted. The statistical analysis results in concentration-duration-frequency curves from which the time percentage of effluent standard exceedance is derived. This will be further discussed in Chapter 5.2.

Table 1: Literature overview of Monte Carlo applications on WWTP modelling

Reference	Type WWTP	Type model	Correlations based on:	Distinction uncertainty/ variability?
(von Sperling, 1996)	Facultative ponds	Steady-state/ 'grey box'	Not mentioned	No
(Kops & Vanrolleghem, 1996)	For use in activated sludge plant	Dynamic but incomplete model ("only Monod growth")	Not mentioned	Yes*
(Dunn *et al.*, 1998)	Sewage treatment works	Steady-state/ 'black box'	Rank coefficients	No
(Haas & Trussell, 1998)	Water reclamation plant	Steady-state/ 'grey box'	Correlation coefficients	No
(Cox, 2000)	Activated sludge plant	Steady-state/ 'grey box'	Correlation coefficients	No
This case study	Activated sludge plant	Dynamic/ 'white box'	Regression relationships	Yes*

* because temporal variability is captured by the dynamic model and uncertainty is captured by Monte Carlo

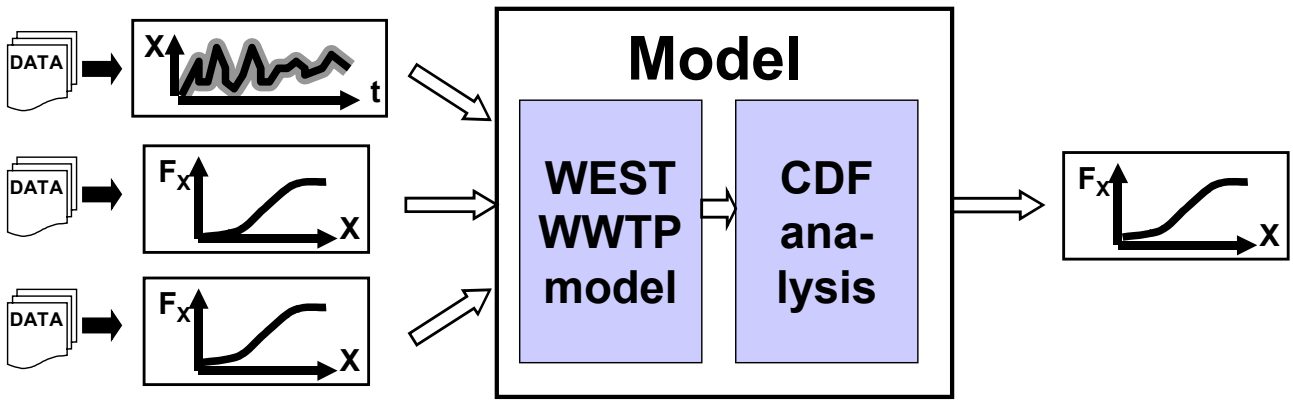


Figure 9: Uncertainty propagation in waste water treatment plant modelling

There are two input types: variables and parameters that are both uncertain (see Figure 9). The uncertainty distributions were estimated based on expert knowledge and literature data.

Parameters are constant during one deterministic simulation. All parameters (heterotrophic and autotrophic growth rates and decay constants, hydrolysis rate, half saturation coefficient for

hydrolysis of slowly biodegradable substrate…) were described by a distribution (mostly triangular and truncated normal distributions). The truncation was necessary to avoid unrealistic negative values and was set at 0.00001. More details can be found in Rousseau *et al.* (2001). The distributions are based on Reichert (1997). The question may arise whether these parameter distributions should be interpreted as uncertainty or variability. They cannot be interpreted as spatial variability because the case study is performed at one particular WWTP. They cannot be interpreted as temporal variability because if the parameter's variations were temporal, then this parameter should not be constant in the dynamic simulation. Instead, it should then be considered as a variable. Consequently, the parameter distributions should be considered as uncertainty. Therefore, there is no need for a second order Monte Carlo that would simulate variability and uncertainty in two loops, as illustrated in Grum & Aalderink (1999) and section 3.1.3.

Variables change during one simulation. Here, the variables are time series. The variables were correlated with the influent flow as this was the only variable for which sufficiently detailed time series were available to feed the deterministic model. Aquafin provided an extensive dataset on influent water quality measurements of several WWTPs from which relationships between input variables could be derived. The regression equations for medium-strength waste water are shown in Table 2. One relationship for COD (Chemical Oxygen Demand) versus flow is shown in Figure 10.

Table 2: Relationships between inputs (flow Q is expressed as m³/day, COD: Chemical Oxygen Demand)

	Minimum	**Average**	**Maximum**
COD (g COD/m³)	$y = 22066\ Q^{-0.6838}$	$y = 135478\ Q^{-0.8199}$	$y = 3.33E+06\ Q^{-1.0994}$
Kjeldahl N (g N/m³)	$y = 103791\ Q^{-1.0836}$	$y = 34100\ Q^{-0.9071}$	$y = 54032\ Q^{-0.9045}$
Nitrate N (g N/m³)	y = 3E-05 Q - 0.3266	y = 0.00013 Q + 3.3301	y = 0.0001 Q + 0.1388

For every component, a triangular distribution was imposed between the minimum, maximum and average values calculated according to the regression in Table 2. In every Monte Carlo shot, time series of e.g. COD were generated based on a given flow time series. The 6-month flow series was generated based on the work of Bauwens *et al.* (1996), after rescaling.

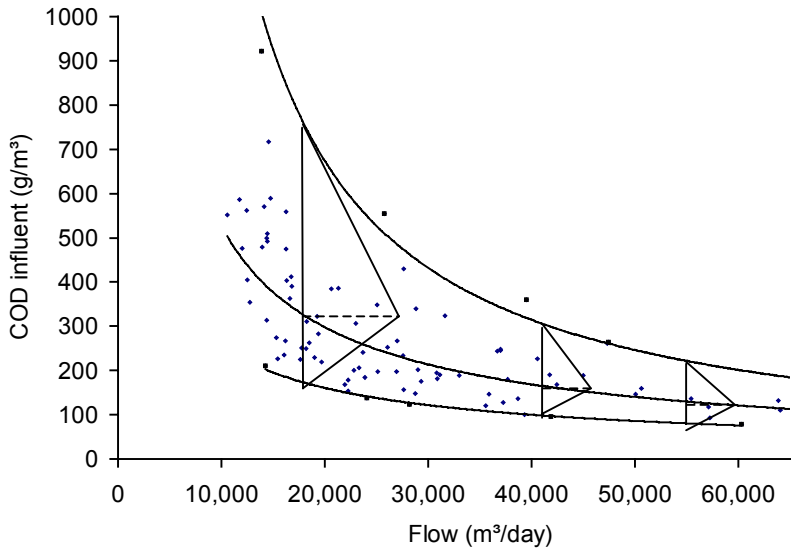


Figure 10: Relation between influent COD-concentrations and flow for medium strength waste water + indication of minimum and maximum concentrations as given in Table 2

In order to cover the entire temperature range in one year, simulations were done over a period of 180 days, starting in the winter period and ending in the summer period. For this case study, 300 Monte Carlo shots were simulated on a Pentium III - 650 MHz based PC. The effluent series were analysed for nitrate-N, ammonium-N and total-N with the effluent standards set to 10 mg N/L, 4 mg N/L and 18 mg N/L respectively. Concentrations were first time-averaged over a two-hour period as imposed by environmental legislations in several countries.

The most important model output is the time percentage of effluent standard exceedance. The Monte Carlo simulation propagated the inherent uncertainty of the inputs and results in a confidence interval or statement for the output. For ammonium-N for instance, the conclusion is that there is 95% certainty that the effluent limit will be exceeded less than 19% of the time. The nitrate-N limit will be exceeded less than 48% of the time (95% certainty) and the total-N limit will be exceeded less than 50% of the time (95% certainty). The European legal standards state that an installation may not exceed the effluent standards more than 5% of the time. Interpreting the output distribution in another way, we are only 43%, 25% and 5% certain that the effluent concentrations of respectively NH_4-N, NO_3-N and TotN comply with this standard. The uncertainty distribution of the effluent standard exceedance for nitrate-N is given in Figure 11. Other than 95% certainty levels can be derived from this distribution. For nitrate-N, the distribution has a lognormal shape.

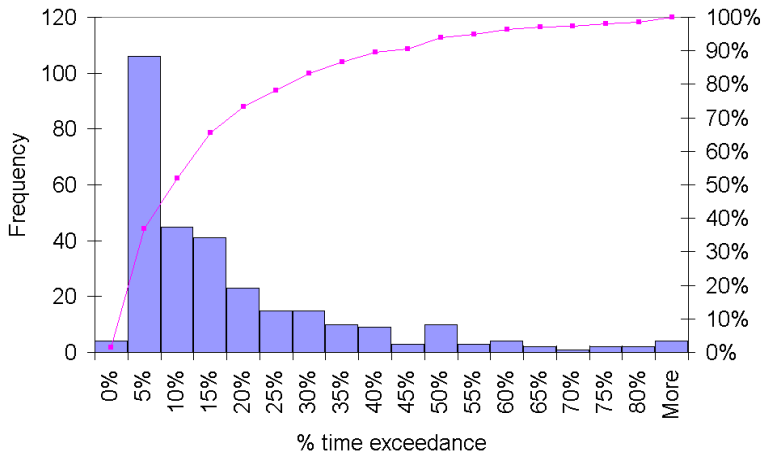


Figure 11: Uncertainty distribution of effluent standard time exceedance for nitrate-N

The developed Monte Carlo software tool is already being used in practice by the private company Aquafin in the retrofitting of old municipal WWTPs in order to comply towards strict phosphorus and nitrogen removal consents (Rousseau *et al.,* 2001), (Bixio *et al.*, 2001a), (Bixio *et al.*, 2001b), (Bixio *et al.*, 2001c). The results show that the decision-making process can be supported under uncertainty conditions and enhance the likelihood of meeting effluent standards not entailing above-normal capital investments. For a particular treatment plant upgrade, the analysis led to reducing the capital investment by 43%, producing savings of more than 1,2 million € (Bixio *et al.*, 2001b).

3.1.2.3.2. Variability Propagation in Biotic Ligand Model

Until recently, environmental water quality standards and risk assessment procedures for metals in surface waters were predominantly based on total and/or dissolved metal concentrations (Janssen *et al.*, 2000). However, the importance of bioavailability and toxicity modifying factors like pH, hardness and Dissolved Organic Carbon (DOC), is increasingly being recognized and is a major contribution to geo-referencing Species Sensitivity (SS). The development of Biotic Ligand Models (BLMs) that predict toxicity of metals to fish, invertebrates and algae (e.g. Di Toro *et al.* (2001), De Schamphelaere *et al.* (2002), Heijerick *et al.* (2002)) can be considered as an important step towards a scientifically sound protection of freshwater environments. Hence, the possible use of these models for regulatory purposes is gaining increased interest in both the scientific and the regulatory community.

It is however still unclear how to deal with the temporal variability of the toxicity modifying factors. This input variability also leads to temporal variability of the metal toxicity. If the yearly average of all input parameters is taken, then the resulting output, the NOEC, will not be the most

conservative estimate i.e. about half of the time, the NOEC will be lower. The use of averages is therefore less interesting from a risk assessment point of view. A lower percentile of the temporal variation of the NOEC is more useful, but it is not straightforward which combinations of inputs lead to a lower percentile in the output.

A Monte Carlo analysis could propagate the temporal variability of the input parameters of the BLM to the output of the BLM. This is also the goal of this case study.

For the simulation of the BLM, several input parameters are needed: total or dissolved organic carbon (TOC or DOC), pH, alkalinity, temperature, Ca, Mg, Na, K, Cl and SO_4-concentrations. Databases of Swedish surface water characteristics, which fulfill these requirements, were obtained from the Institute of Environmental Analysis of the Swedish University of Agricultural Sciences (SLU, http://info1.ma.slu.se). Here, the *Daphnia magna*-BLM (the most advanced chronic BLM, De Schamphelaere and Janssen, unpublished) for Copper was used. The most important outputs of the BLM are No Observed Effect Concentrations (NOECs) for *Daphnia magna*. A schematic overview of the Monte Carlo simulation can be found in Figure 12.

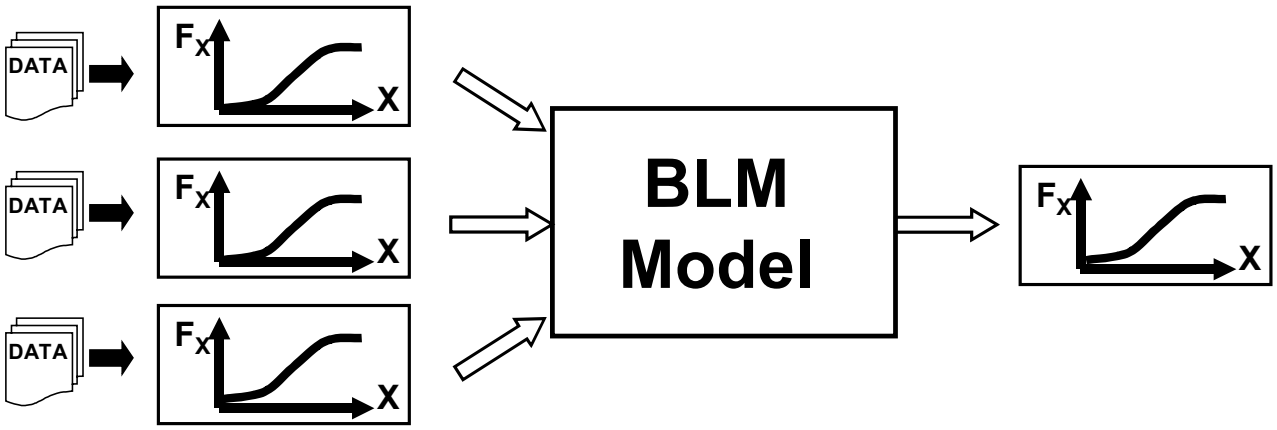


Figure 12: Variability propagation in Biotic Ligand Model (BLM)

Only those parameters that are subject to temporal variability were considered in the Monte Carlo simulation. These input parameters (Total or Dissolved Organic Carbon (TOC or DOC), pH, alkalinity, temperature, Ca, Mg, Na, K, Cl and SO_4) were described by normal and lognormal distributions. More details can be found in De Laender (2003). The correlations between the input parameters were described by means of their correlation coefficient and simulated according to the method of Iman & Conover (1982). The most significant relationships were those involving the Ca concentration. DOC, one of the most sensitive input parameters, was not correlated with any of the other input parameters.

The resulting NOEC output distribution is shown in Figure 13. This distribution represents the temporal variability of the NOEC. Point A in Figure 13 indicates that the NOEC in a particular

Swedish lake will exceed 75 μg/l for 80% of the time. Note that the interpretation does not contain an uncertainty statement. The output distribution given here is only due to variability.

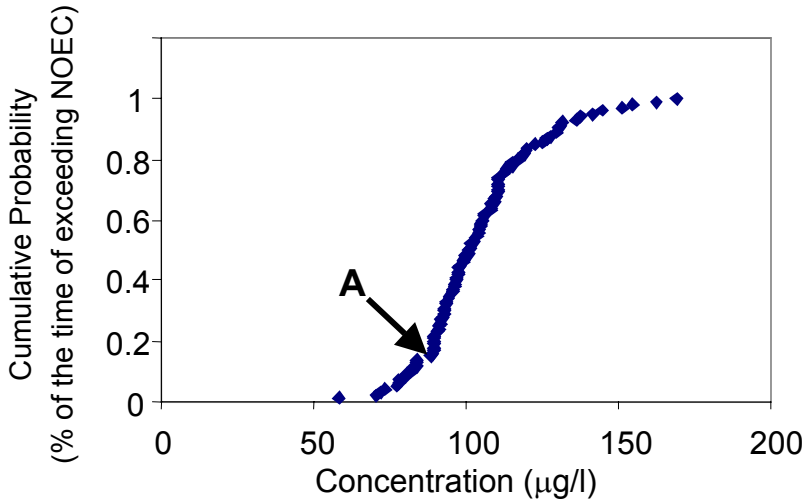


Figure 13: Temporal variability distribution of NOEC for Cu and Daphnia Magna

3.1.2.4. Discussion

A first order Monte Carlo simulation was developed in flexible software tools. First order Monte Carlo is an accurate and easy-to-use technique to propagate variability or uncertainty. Specific attention needs to be devoted to the correlations between inputs, and the interpretation of the output uncertainty or variability.

The case studies illustrated the first order Monte Carlo simulation for uncertainty or variability propagation. The correlations were explicitly considered and the crude Monte Carlo sampling scheme was sufficiently accurate.

In the first case study, the probability of exceeding the effluent limits of a WWTP is accompanied with a confidence interval resulting from the inherent uncertainty of influent characteristics and model parameters as propagated by the Monte Carlo simulation. This uncertainty or confidence interval allows decision-makers to choose whether to adjust the proposed design or to decide on another scenario.
In the second case study, the temporal variability of the NOEC of a heavy metal is estimated by propagating the temporal variability of the compounds that determine its bioavailability. From this variability distribution, some lower percentile may be derived as a protective standard for most of the time.

3.1.3. Second Order (or Two-Dimensional) Monte Carlo Simulation

3.1.3.1. Introduction

A first order or one-dimensional Monte Carlo simulation can only propagate variability or uncertainty, but not both at the same time without having difficulties with interpreting the output distribution. Variability represents heterogeneity or diversity, which is not reducible through further measurement or study. Uncertainty represents ignorance about a poorly characterised phenomenon, which is sometimes reducible through further measurement or study. It is already shown in Part 2 that variability and uncertainty should be treated separately. For this, among other propagation techniques, a second order or two-dimensional Monte Carlo simulation could be applied. The goal of this section is to present and apply second order Monte Carlo simulation in a case study.

3.1.3.2. Second Order Monte Carlo Simulation

In this section, the actual second order Monte Carlo methodology and the sampling of the second order random variables and the simulation of correlations will be discussed. The random number generation and the verification of the Monte Carlo output is the same as in the previous section on first order Monte Carlo simulation.

3.1.3.2.1. The Second Order Monte Carlo Simulation Technique

Let X_1, X_2, X_i…, X_n be n random input variables of a mathematical model. Here, the n random variables are all describing true heterogeneities of environmental processes and can thus be interpreted as variability. Each random variable can therefore be characterised by a distribution and depends on some set of parameters θ. This is noted as $X_i \sim F(\theta_i)$. Often, the parameters θ_i cannot be estimated accurately because for instance, only a limited data set is available. This leads to uncertainty of the parameters θ. Random variables that are uncertain and variable at the same time are called second order random variables (Burmaster & Wilson, 1996). They are also referred to as "distributions of distributions" or as confidence intervals on distribution parameters θ. The distribution of the parameters θ_i is noted as $\theta_i \sim F(\gamma_i)$.

The most straightforward method to obtain results in the form of second order variables is called second order or 2-dimensional or embedded Monte Carlo simulation (Cullen & Frey, 1999). It consists simply in two Monte Carlo loops nested one inside the other. The inner one deals with the variability of the input variables, while the outer one deals with uncertainty. For each shot of a (uncertain) parameter value in the outer loop a whole distribution is created in the inner loop based

only on variability. In this way changes in variability-dependent frequency distributions under the influence of parameter uncertainty can be quantified. This is graphically shown in Figure 14.

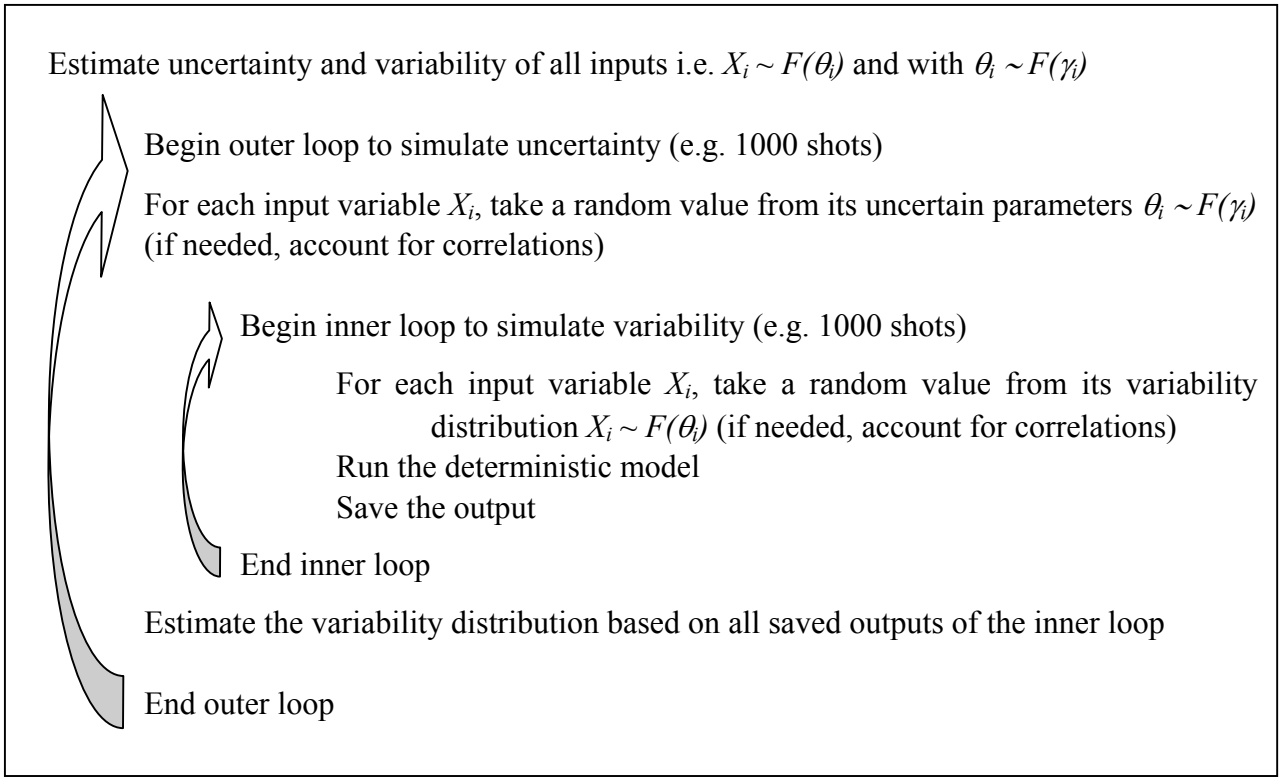


Figure 14: Simulation algorithm of a second order Monte Carlo simulation

3.1.3.2.2. Estimation of Second Order Probability Distributions

Several techniques can be used to characterise a second order random variable or to estimate uncertainty and variability in a data set: expert judgement or data-driven techniques like bootstrapping, classical or Bayesian approaches… These will be discussed in Chapter 3.2. Figure 15 gives an example of a second order random variable. The cumulative distribution function itself is a variability distribution e.g. reflecting the increasing effect of increasing concentrations on a set of species. For each percentile of the variability distribution, a confidence or uncertainty interval can be calculated (i.e. an uncertainty distribution).

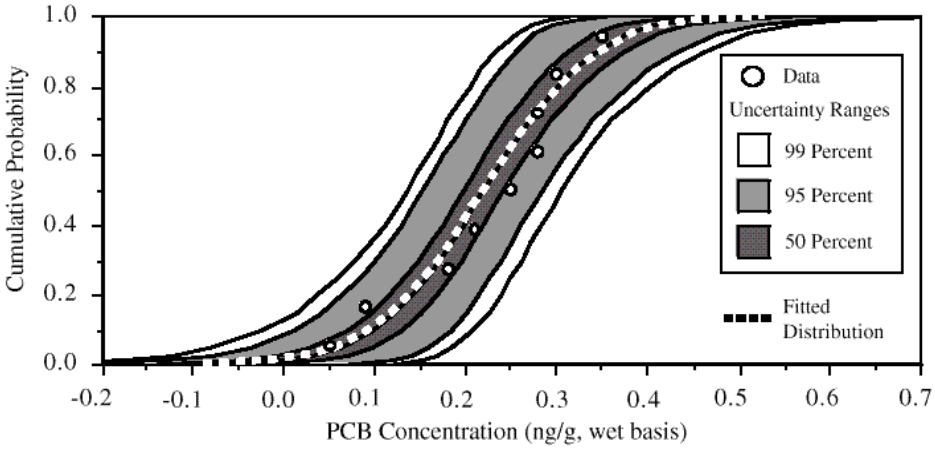


Figure 15: Example of a cumulative distribution function of a second order random variable

3.1.3.2.3. Simulation of correlations

In the example shown in Figure 15, the mean and standard deviation were allowed to vary independently. Thus, a distribution could be defined by a combination of a low mean and a high standard deviation, high mean and low standard deviation, or any other combination in between. The assumption of independence of parameters may not be valid in all cases. It may be unreasonable to assume that for example a high mean soil concentration would occur with a low standard deviation. An alternative assumption would be that the standard deviation of the mean is a constant proportion of the mean (i.e. a constant coefficient of variation). Hence, correlations between parameters should be considered in the design of the probabilistic ecological risk assessment.

A common approach for correlating two parameters of e.g. the normal distribution is to specify a bivariate normal distribution. A bivariate normal distribution allows for the distribution of one parameter to be sampled conditional on the other. This is a special case of a joint distribution in which both *x* and *y* are random variables and normally distributed (as the conditional distribution of *x* or of *y* is always normal) (EPA, 2001). For more information on handling dependencies and correlations, see section 3.1.2.2.5. Resampling the observations (as in bootstrapping, see Chapter 3.2) is a useful, simulation method to assess the relationship between the parameters of a distribution.

3.1.3.3. Case Study

The Exposure Concentration Distribution (ECD) and the Species Sensitivity Distribution (SSD) are characterised by both uncertainty and variability. The resulting risk quotient will also be

characterised by both uncertainty and variability. The main focus of this case study is to show how second order Monte Carlo can propagate the uncertainty and variability to a variable and uncertain risk quotient. No attention was given here to the risk quotient model or EC and SS distribution selection or potential autocorrelation within the EC data set, as these will be discussed in Chapter 3.2 and 4.1. Here, the model is very simple; it only consists of a quotient of two input variables (see Figure 16). The quotient is the risk quotient.
Note that for simple models like this one, it may be possible to calculate the risk quotient analytically as e.g. in case of two lognormal input distributions (see Chapter 4.1).

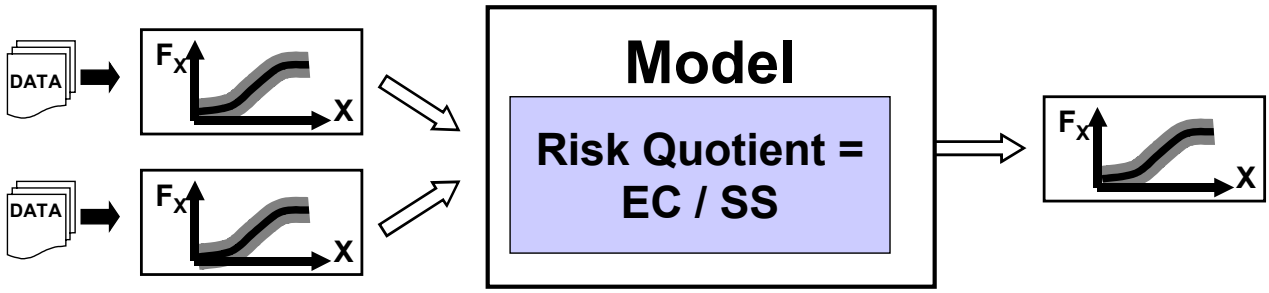


Figure 16:Variability AND uncertainty propagation in risk quotient modelling

The two input variables are variable and uncertain. The first input, the ECD, represents the spatial and temporal variability of Zn-concentrations for the Netherlands in 1998 extracted from RIZA and CIW databases. A regional Zn background concentration was also subtracted from the monitoring data. More details can be found in Van Sprang *et al.* (2002). A lognormal distribution was fitted to the data. Since the data set is very large (number of data points = 2183), the sampling uncertainty (shown as the grey band in Figure 17 left) is very small for the ECD. A nonparametric model is perhaps more suitable because enough data are available, but this is not important considering the goals of this case study.
The second input, the SSD, represents several species sensitivities towards Zn. Long term ecotoxicity data on Zn for the aquatic organisms belonging to the different trophic levels were gathered from the Zn risk assessment report. A Pareto distribution was fitted to the data (selection of Pareto, see Chapter 3.2). Since the data set is rather small (number of data points is 21), the sampling uncertainty is larger (see Figure 17 left).

The uncertainty and variability of these second order random variables were estimated based on their respective data sets using the parametric bootstrap technique (see Chapter 3.2). A second order Monte Carlo simulation was performed. Both inputs (ECD and SSD) were assumed to be independent of each other. As a consequence, no correlations had to be considered in the inner Monte Carlo loop of variability. Potential correlations between the parameters of each distribution type of the ECD and SSD (in the outer Monte Carlo loop of uncertainty) were indirectly modelled because they were based on earlier parameter resample estimates of the parametric bootstrap technique.

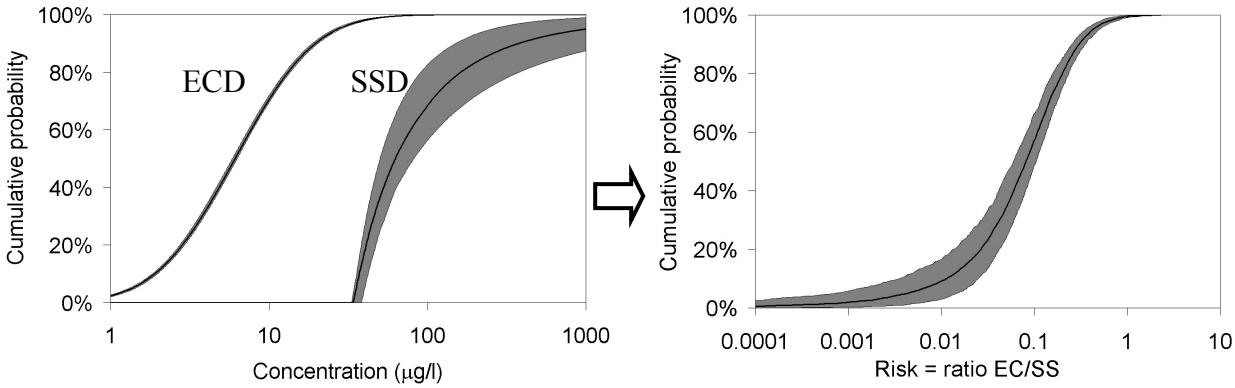


Figure 17: EC and SS input distributions (left) and resulting Risk quotient distribution with uncertainty band simulated by means of second order Monte Carlo simulation (right)

The output, the risk quotient distribution and its uncertainty band, is shown in Figure 17 right. The risk quotient variability distribution can be interpreted as all possible combinations of quotients of EC and SS. The uncertainty band represents the sampling uncertainty of the risk quotient due to the sampling uncertainty of the EC and SS.

3.1.4. Conclusions & Further Research

This chapter demonstrated the possibilities and usefulness of the propagation technique Monte Carlo simulation. The major advantages of simulation techniques are that it can accommodate a wide variety of assumptions regarding model inputs and can be used with a wide variety of models. Simulation methods are also available to evaluate the effect of correlations and dependencies among model inputs, as described. The accuracy of simulation methods can be improved by increasing the number of shots, but this also points to the weakness of the method, its computational burden. However, in our studies, the models were either sufficiently simple or the required number of shots for the derived accuracy was small.

It was stressed and illustrated that separation of uncertainty and variability and the correct application of Monte Carlo analysis simplify the interpretation of a model's output distribution of interest.

Chapter 3.2

-

Uncertainty and Variability Estimation: Comparison of Several Techniques

This chapter was published in:

Verdonck F. A. M., Jaworska J., Thas O. & Vanrolleghem P. A. 2001. Determining environmental standards using bootstrapping, Bayesian and maximum likelihood techniques: a comparative study. *Analytica Chimica Acta*, 446(1-2), 429-438.

Chapter 3.2

Uncertainty and Variability Estimation: Comparison of Different Techniques

In Probabilistic Ecological Risk Assessment (PERA), estimation of uncertainty and variability of existing monitoring data and ecotoxicity tests is needed to determine the Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD) or to determine input distributions of exposure or effects models (see filled arrows in Figure 6 of Part 2).

Research priorities on estimating variability and uncertainty are mainly situated on the effects side of the risk assessment in the determination of a SSD. More specifically, many such estimation methods have been proposed (see Part 2) but only few explicitly consider small sample sizes (i.e. smaller than or equal to 20). In practice, this is very relevant as small ecotoxicity data sets are common for many chemicals; especially tests based on chronic or population level endpoints. On the exposure side of the risk assessment, data sets typically have larger sample sizes. Typical issues for the ECD determination are presence of an upper bound (e.g. the water solubility of an aquatic concentration) or the presence of censored data (e.g. due to a detection limit). No case studies on the estimation of variability and uncertainty of the ECD are given here, but some examples can be found in Chapter 4.1 and 5.1. Note also that the methodologies presented here are valid for estimating uncertainty and variability of any small data set.

3.2.1. Introduction

The basic assumption of the SSD concept is that the sensitivities of a set of species (or inter-species sensitivity/variability) can be described by some statistical distribution. The available ecotoxicological data are seen as a sample from this distribution and are used to estimate the

parameters of the SSD. An SSD can be visualised as a cumulative distribution function (see Figure 1). This is the integral of an associated probability density function. The cumulative distribution function curve follows the distribution of the sensitivity data obtained from ecotoxicological testing; plotting effect concentrations derived from acute or chronic toxicity tests.

The most common current approach is to derive the Predicted No Effect Concentration (PNEC) from the 5th percentile of the SSD (EU-TGD, 1995) as shown in Figure 1 (for the moment, do not mind the uncertainty bands). Historically that value is known as the Hazardous Concentration at p-protection level or HC_p. The 5th percentile of a chronic toxicity distribution has most often been chosen as a concentration that is protective for most species in a community (namely 95%). Researchers also started to determine a confidence or uncertainty interval on the HC_5 (see Figure 1). This was mainly done because the median HC_5 is a conservative estimate of the HC_5 calculated without uncertainty (Aldenberg & Slob, 1993). Until now, unfortunately, only few studies report confidence intervals. A confidence or uncertainty interval can quantify the sampling error in the HC_5 estimate. Sampling error is due to the fact that only a limited sample is collected. In theory, one would need to collect an infinite number of sample points to obtain the correct estimate of the HC_5.

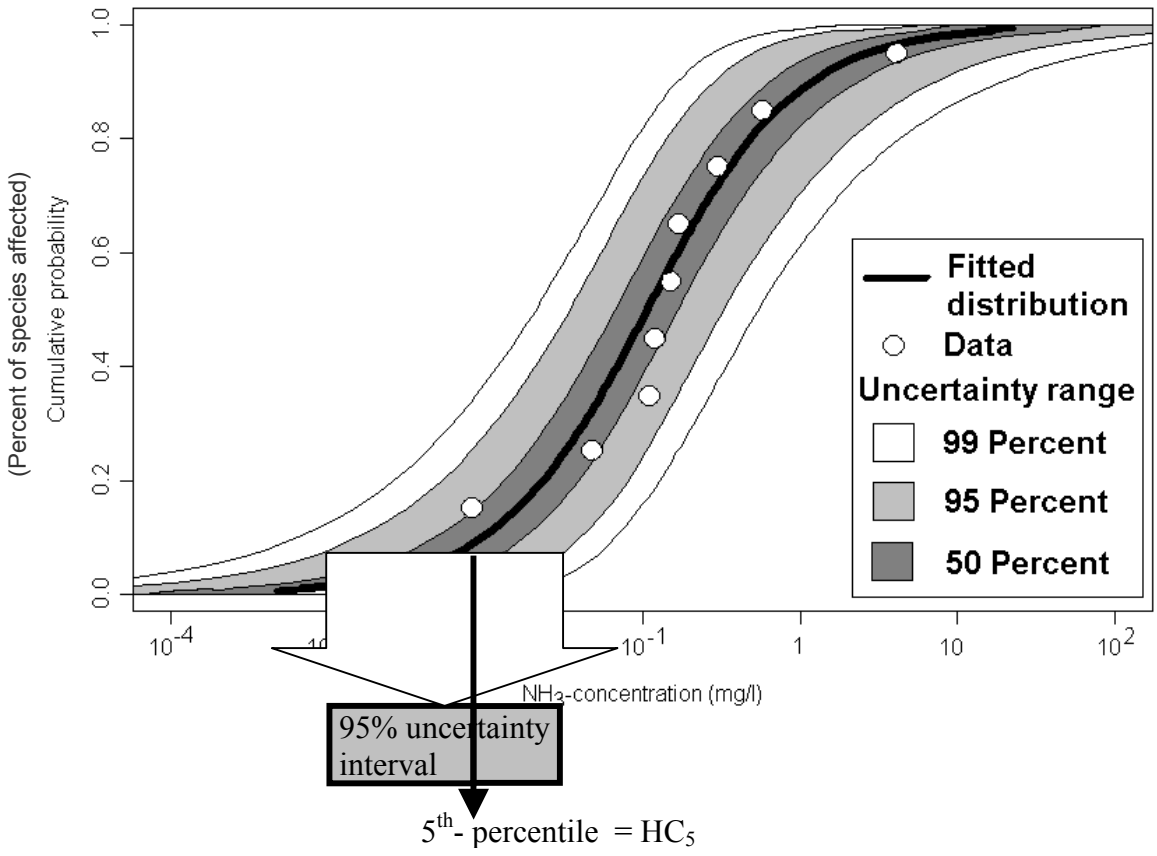


Figure 1: Example of a SSD (Species Sensitivity Distribution) with uncertainty band and HC5 (Hazardous Concentration at 5%)

Several statistical techniques exist to estimate (inter-species) variability and (sampling) uncertainty in a data set. An overview of uncertainty and variability estimation techniques applied in the ecotoxicology field is given in Part 2. This area of quantitative risk analysis is currently an active area of research, but all different techniques for estimating variability and uncertainty and all different parametric distributions for estimating the 5^{th} percentile (or HC_5) from SSDs have not been compared for small data sets (i.e. sample size = 20 or less). In addition, nobody (except Jagoe & Newman (1997) and Shao (2000)) performed a simulation study in which important statistical concepts as coverage and bias were tested to assess the reliability of their proposed method. In statistics, an estimate is unbiased if the mathematical expectation of the estimator is equal to the true value of the parameter. Similarly, a method has accurate coverage if the probability p that a confidence interval does not cover the true parameter is equal to the probability level used to construct the confidence interval. Confidence intervals are expected to enclose a true but unknown parameter according to a specified probability, such as 90% or 95%. This is the expected coverage of the confidence interval, given a specified significance level. The difference between the expected coverage and the actual coverage is one metric for evaluating statistical methods that yield different confidence intervals.

Therefore, the goal of this chapter is to compare and assess the reliability of most of these uncertainty and variability estimation methods at sample size 20 by means of simulation studies and case studies.

3.2.2. Methods for Variability Estimation (One-Dimensional)

There exist several statistical methods to characterise the distribution of a random variable. Focus of this section is on how to estimate the probability distribution and its percentiles. This probability distribution is mostly interpreted as variability but note that also uncertainty (other than sampling uncertainty, e.g. measurement uncertainty) is a possible interpretation. Estimating both uncertainty and variability at the same time is the topic of 3.2.3. A distinction is made between parametric and nonparametric methods for characterising a random variable.

Let $x_1, x_2, \ldots, x_n$ be an independent and identically distributed random sample from x. This variable has distribution function $F(x)$, density function $f(x)$ and depends on some parameters $\theta_1, \theta_2, \ldots, \theta_m$. The sample mean is $\bar{x}$ and the standard deviation is s.

3.2.2.1. Parametric Methods

3.2.2.1.1. Distribution Fitting

Three of the most common approaches to estimate the parameters of a distribution are the method of (matching) moments, the Maximum Likelihood Estimation (MLE) method and the method of least squares (Cullen & Frey, 1999), (EPA, 2001).

In the method of (matching) moments, the parameters of a probability distribution are selected such that the moments of the model match the moments of the data set. The number of moments that are used in this process corresponds to the number of parameters to be estimated. This approach is usually the most straightforward to implement. Therefore, it typically satisfies the criteria of practicality. However, it may not fully satisfy other criteria like consistency, efficiency, bias, and robustness (Cullen & Frey, 1999).

The maximum likelihood estimation method involves selecting values of the distribution parameters θ_1, θ_2, … θ_m that are most likely to yield the observed data set. To do this, a likelihood function is defined. For a continuous random variable for which independent sample points have been obtained, the likelihood function L is:

$$L(\theta_1,\theta_2,...,\theta_m)=\prod_{i=1}^{m} f(x_i|\theta_1,\theta_2,...,\theta_m)$$

where the likelihood function L is evaluated based upon the product of the probability density function evaluated for each sample. The parameters of the probability distribution are selected so as to maximise the value of the likelihood function. Thus, this is an optimisation problem in which the first derivative of the likelihood function with respect to each of the distribution parameters is set to the distribution parameters. For small sample sizes, the maximum likelihood estimates do not always yield minimum variance or unbiased estimates. However, for larger sample sizes, the maximum likelihood method tends to better satisfy the criteria for statistical estimators: consistency, efficiency, bias and robustness. However, it is often much more difficult to implement than the method of moments. Analysts should not feel the method of maximum likelihood is always ideal (Cullen & Frey, 1999).

In some cases, both methods yield the same estimators, e.g. for the normal distribution (Cullen & Frey, 1999). For this reason, the method of moments was selected to estimate the parameters of a normal distribution.

The least squares method is a third distribution-fitting tool. This technique involves finding a probability and data scale that plots the cumulative probabilities of a hypothesized distribution as a straight line. The corresponding linearity provides a measure of the goodness-of-fit of the hypothesized distribution. The species are ranked in order of decreasing sensitivity; the rank of each species is converted to a cumulative probability ($i/(n+1)$), where i is the species rank and n is the total number of species for which data are available, and transformed to a probability using e.g. the normal distribution. The relationship between the normalized rank and the species sensitivity (effect concentration) is determined by least squares regression (ECOFRAM, 1999). Many statisticians feel that maximum likelihood is more rigorous on statistical grounds than the regression approach. However, maximum likelihood would require special computer programs to perform the iterative, numerical optimisations that such an approach requires. Used judiciously, the simple calculations required for lognormal regression will ordinarily prove to be practical. These calculations can be readily implemented in spreadsheets. However, with all of these approaches, biases in the selection of data or lack of availability of data can result in incorrect estimations (ECOFRAM, 1999). Cullen & Frey (1999) point out that probability plotting may not be a primary choice for selecting and fitting distributions because the method violates an important assumption of least squares regression i.e. independence of the observations. This is because the rank-ordered data are no longer independent.

3.2.2.1.2. Distribution Selection

After fitting the distribution to the data using one of the techniques above, the goodness-of-fit should be assessed. For this, several criteria should be considered. More information on selecting and fitting distributions can be found in Cullen & Frey (1999), Vose (1996), EPA (2001) and EPA (1999).

First, graphical methods can provide valuable insights and generally should be used in conjunction with exploratory data analysis. They reveal important characteristics of a data set, including skewness (asymmetry), number of peaks (multi-modality), behaviour in the tails, and data outliers. Examples of graphical methods are frequency distributions (i.e. histograms), dot plots, line plots for discrete distributions, box-and-whisker plots and scatter plots. In a QQ-plot, observed values of a single numeric variable are plotted against the values that would be obtained if the sample were from a normal distribution. If the sample is from a normal distribution, points will cluster around a straight line. Here, the line is plotted through the first and third quartile of the data. The QQ-plot also depends on plotting positions and those are calculated according to Hazen as discussed in section 3.2.2.2.1.

Second, expert judgement refers to inferential opinion of a specialist or group of specialists within an area of their expertise. When there is uncertainty or variability associated with an input variable,

such as a data gap, expert judgement may be appropriate for obtaining distributions. Distributions based on expert judgement can serve as Bayesian priors in a decision-analytic framework. The distributions and Bayesian priors can be modified as new empirical data become available.

Third, goodness-of-fit tests exam how well (or poorly) a sample of data can be described by a hypothesized probability distribution for the population. Goodness-of-fit tests are formal statistical tests of the hypothesis that the data represent an independent sample from an assumed distribution. These tests involve a comparison between the actual data and the theoretical distribution under consideration. However, goodness-of-fit tests have low statistical power and often provide acceptable fits to multiple distributions. Thus, goodness-of-fit tests are better used to reject poorly fitting distributions than for ranking good fits. For small n, goodness-of-fit tests will often fail to reject many of the hypothesized probability distributions.

3.2.2.1.3. Percentile estimation

In parametric methods, a percentile can be calculated based on the inverse cumulative distribution function according to the following equation:

$$\alpha^{th} - percentile = F^{-1}\left(x, \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m\right)$$

with x the population random variable
$\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ the estimated parameters of the distribution
F^{-1} the inverse cumulative distribution function

3.2.2.2. Non-parametric Methods

3.2.2.2.1. Distribution 'Fitting'

Cullen & Frey (1999) summarise several possible methods for constructing a (nonparametric) Empirical Distribution Function (EDF) of an observed data set. These methods are referred to as "plotting positions". The plotting position is an estimate of the cumulative probability of a data point. First, rank ordering the data is needed. Then, the calculation of the cumulative probabilities can be done according to several methods (see Table 1). In the formulas, $F(x_i)$ denotes the probability that the random value X will have values less than that of the sample point x_i. The sample points are ranked in ascending order, the index i refers to the rank of each sample point, and n is the number of data points in the data set.

Table 1: Overview of some plotting positions of the cumulative probability of a data point

Name	Formulae	β	Reference
General	$F(x_i) = \frac{i-\beta}{n-2\beta+1}$	β	(Blom, 1958)
Mean plotting position	$F(x_i) = \frac{i}{n+1}$	0	(Gumbel, 1958) (Davison & Hinkley, 1997)
Modal plotting position	$F(x_i) = \frac{i-1}{n-1}$	1	(Cullen & Frey, 1999)
Hazen's plotting position	$F(x_i) = \frac{i-0.5}{n}$	0.5	(Hazen, 1914)
Standard plotting position for EDF	$F(x_i) = \frac{i}{n}$	-	

Hazen's plotting position, with β = 0.5, appears to be a compromise between the mean plotting position (β = 0) and the modal plotting position (β = 1). A consequence of using the quantity (n-1) in the denominator in the modal plotting is that the smallest and largest values in the sample are given percentiles of 0 and 1, respectively. Some plotting positions have known deficiencies. For example, for small sample sizes the mean plotting position leads to overestimates of the standard deviation, while the Hazen plotting position leads to underestimation of the standard deviation. As the sample size increases, the various plotting positions tend to approach each other and, therefore, differences between them become less important. As to risk assessment, data are often scarce so the importance of an accurate plotting position may not be underestimated. Many investigators use either the mean or Hazen plotting positions (Cullen & Frey, 1999). Cullen & Frey (1999) described that the mean plotting system gives erratic results for small sample sizes. Nevertheless, in the ecotoxicity field, the mean plotting is still often used (e.g. Jagoe & Newman (1997), Van Der Hoeven (2001)). More information on order statistics and plotting positions can be found in section 5.6.4 of Aldenberg *et al.* (2002).

3.2.2.2.2. Percentile estimation

Once the observed data set is plotted, percentiles can be calculated taking the inverse interpolated empirical distribution function. A percentile is then calculated nonparametrically according to the following steps (Cullen & Frey, 1999) (see also Figure 2):
First, the data are rank ordered (COL1 and COL2 → COL3 and COL4).

Then, the Hazen plotting system calculates the cumulative probabilities of a point x_i as follows: $F(x_i) = \frac{i - 0.5}{n}$ (COL 3 → COL 5) where i is the rank number and n is the total number of data points.

Once the data are plotted, percentiles can be calculated taking the inverse (interpolated) empirical distribution function (see Figure 2). One can easily derive that the 20th percentile is 0.6 and the 80th percentile is then 1.75.

COL1	COL2	COL3	COL4	COL5
Shot number	x_i	Ranked number i	x_i (sorted)	$F(x_i)$
3	1	1	0.5	10%
1	0.5	2	0.7	30%
5	2	3	1	50%
4	1.5	4	1.5	70%
2	0.7	5	2	90%

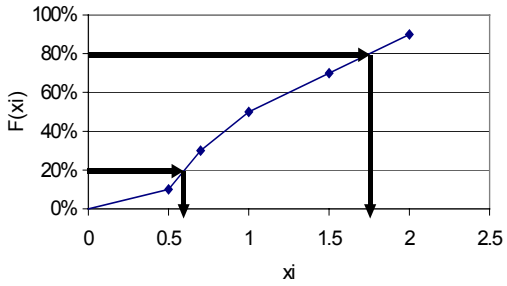


Figure 2: Example of nonparametrically percentile estimation

One of the main shortcomings of nonparametrically determining a percentile of a data set is that the minimum and maximum values obtained are limited by the minimum and maximum values within the data set. Zero is sometimes considered as a minimum. When only small data sets are available, this can lead to biases in the representation of a given model input (e.g. failure to consider possible large, or small values that are not present in the limited data set).

One can even calculate the minimum sample size needed to estimate a 5th-percentile without having to extrapolate (below the smallest data point). The rank order i for the smallest point is 1, $F(x_i)$ is 0.05 (=5 %) and n is unknown. The calculations for different plotting systems are then:

Mean plotting system: $F(x_i) = \frac{i}{n+1} \Leftrightarrow 0.05 = \frac{1}{n+1} \Leftrightarrow n = 19$

Hazen plotting system: $F(x_i) = \frac{i - 0.5}{n} \Leftrightarrow 0.05 = \frac{1 - 0.5}{n} \Leftrightarrow n = 10$

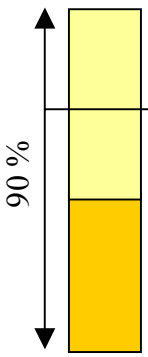
Other plotting system: $F(x_i) = \frac{i}{n} \Leftrightarrow 0.05 = \frac{1}{n} \Leftrightarrow n = 20$

The minimum n=19 for the mean plotting position corresponds to earlier findings in literature (e.g. Van Der Hoeven (2001), Grist *et al.* (2002)).

3.2.3. Methods for Variability and Sampling Uncertainty Estimation (Two-Dimensional)

In this section, an overview is given of methods estimating both uncertainty and variability at the same time in a two-dimensional analysis. The distinction in structure is here made between several methods and not between parametric and nonparametric methods as in the previous section.

The terminology to be used is shown in Figure 3. The vertical bar represents a 90% uncertainty or confidence interval of the 5th-variability percentile. This bar is the 90 degrees left rotation of the horizontal bar in Figure 1. This uncertainty interval has an upper 90% confidence or uncertainty limit (equal to the 95th percentile of the uncertainty distribution) and a lower 90% confidence or uncertainty limit (equal to the 5th percentile of the uncertainty distribution). The line in the uncertainty bar is the estimate of the 5th variability percentile or HC_5. It is estimated by the median of the uncertainty distribution. For ease of use, this will here be named the 5th percentile in a statistical context and HC_5 in a risk assessment context.



→ **upper 90% confidence or uncertainty limit** = 95th- uncertainty percentile

→ true 5th-(variability) percentile = true HC_5

→ median **5th-variability percentile** = median 5th-percentile of the variability distribution = 50th-percentile (median) of uncertainty of the 5th-percentile of variability = median 90 % confidence limit = estimate of **HC_5** (Hazardous Concentration 5%)

→ **lower 90% confidence or uncertainty limit** = 5th- uncertainty percentile

Figure 3: Terminology and visualisation

The distinction between a target population (represented by distribution *F*) and an estimated target population (represented by distribution $\hat{F}$) should be considered carefully (Figure 4). The target population is often considered to be the "population of concern" or the real, unknown distribution. A risk assessor is often interested in quantifying specific attributes of the population (e.g. exposure mean or HC_5). Distributions are generated from representative sample populations to make inferences about the target population. Ideally, a sample should be a subset of the target population and should be selected for measurement to provide accurate and representative information about the variable being studied. However, defining representative samples is a matter of interpretation.

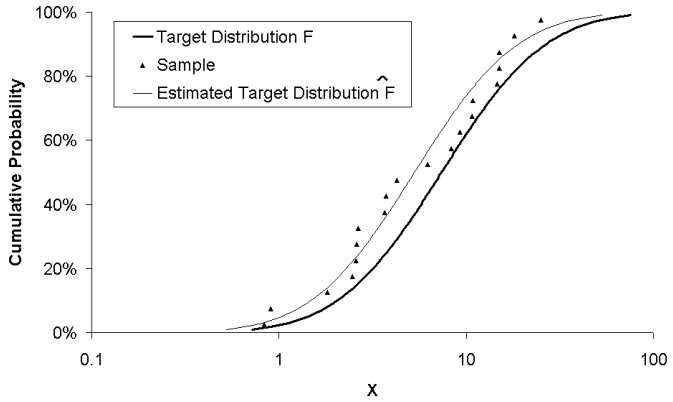


Figure 4: Example of target distribution F and an estimated target distribution $\hat{F}$ based on a sample of 20 data points

3.2.3.1. Bootstrapping

A detailed description of the bootstrapping method can be found in literature (Cullen & Frey, 1999), (Davison & Hinkley, 1997), (Efron & Tibshirani, 1993). Different types of bootstrapping were studied: two non-parametric techniques, each with two different plotting systems for constructing an Empirical Distribution Function (EDF), and one parametric technique (assuming the lognormal distribution).

Given a sample of size *n*, the general approach in bootstrap simulation is to assume $\hat{F}$, to perform *r* replications (e.g. *r* = 5000) of the original data set by randomly drawing, with replacement, *n* values from $\hat{F}$, and then calculate *r* values of the statistic of interest. These can be used to determine the uncertainty distribution of the statistic of interest. Note that the properties of the bootstrap are asymptotic.

One approach is to use the actual data set itself and to randomly select, with replacement, the actual values of the data set. This is sometimes referred to as **resampling**. The data can be represented via an EDF $\hat{F}$. Another approach to achieve the same is to first calculate *n* random samples of a uniform distribution between 0 and 1. The random samples are then found as the inverse cumulative EDF of these random uniform samples. This can be noted as $\hat{F}^{-1}(U(0,1))$. The solid dark line in Figure 5 gives an EDF for a given data set. Basically, resampling is done from that step function. $\hat{F}$ is assumed to be a good estimate of the real, unknown target distribution *F*.

But, $\hat{F}$ can be of any distribution type (nonparametric or parametric (e.g. lognormal, triangular…)) and will determine the type of bootstrapping. From $\hat{F}$ the statistic of interest can be calculated (here the 5th percentile).

An interpolated empirical distribution function $\hat{F}$ (interpolated EDF ≠ EDF) can be fitted to the data. The broken, grey line in Figure 5 shows an **interpolated EDF**. In this approach, samples are taken from the interpolated EDF instead of the EDF. Such a distribution has minimum and maximum values, which can be constrained by the minimum and maximum values in the data set, or have to be determined explicitly. Here, zero was considered as the minimum and a very large number was taken as the maximum. These minimum and maximum values were only used to extrapolate below the smallest and above the largest data point. The minimum and maximum values were not used as two additional data points. As introduced earlier, there are several "plotting positions" to construct an EDF and an interpolated EDF: mean and Hazen plotting.

$\hat{F}$ can also be a parametric distribution. This approach is called **parametric bootstrapping**. The broken, black line in Figure 5 represents a fitted normal distribution for a given data set.

$\hat{F}$ can also be a smoothed distribution. This approach is called smoothed bootstrapping. However, this technique was not considered here because it is a more complex method and it has to deal with additional issues like selection of an optimal bandwidth.

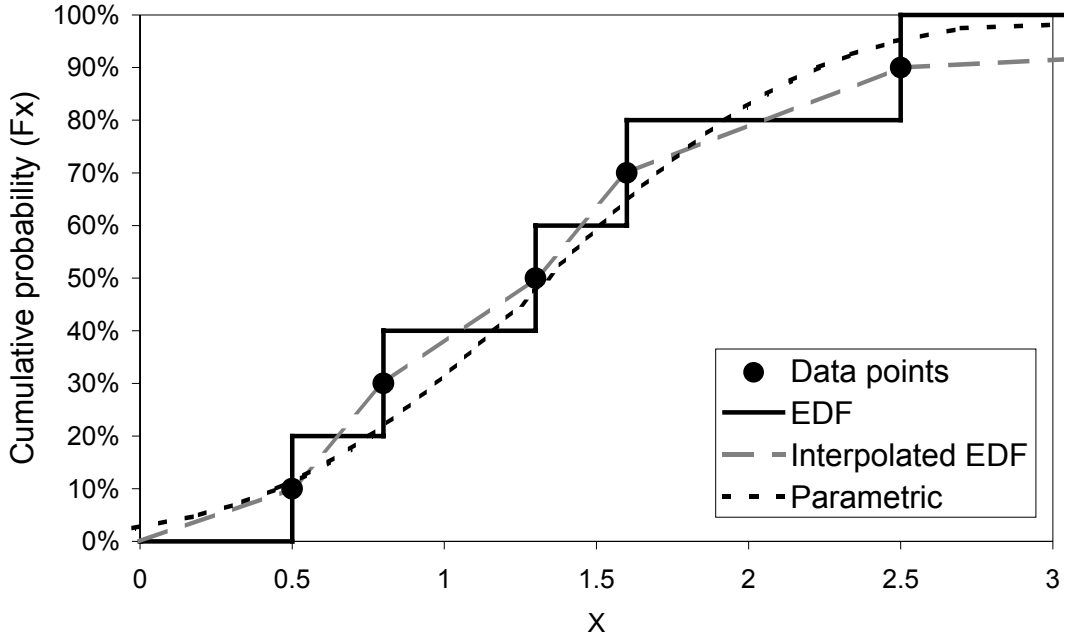


Figure 5: EDF (Empirical Distribution Function), interpolated EDF and a parametric fit (normal in this case) for a given data set

Each approach will lead to a different estimate of the confidence interval of the 5th percentile. Nonparametric or distribution-free approaches do not require assumptions regarding the probability distribution for the underlying population distribution. However, they also tend to yield wider confidence intervals than parametric methods do.

Note that $\hat{F}$ can be a nonparametric or parametric distribution and that the statistic of interest (5th percentile (HC_5) in this case) can be calculated nonparametrically or parametrically. In this study, when a parametric $\hat{F}$ was assumed, the 5th percentile (HC_5) was also calculated parametrically. And analogously, when $\hat{F}$ was considered to be nonparametric, the 5th percentile was also calculated nonparametrically. In literature, some use hybrid versions of the bootstrap. Grist *et al.* (2002), for example, consider the nonparametric EDF as $\hat{F}$, but calculate the 5th percentile parametrically (by means of linear regression). These hybrid bootstrap versions were not considered here because we felt that consistency in the method is more important: either a full parametric or a full nonparametric bootstrap.

The most commonly used method to give confidence intervals for the statistics of interest is the percentile method, which simply consists of taking the inverse of the cumulative distribution function of the bootstrap sample (as described above). But, there are better ways to construct bootstrap confidence intervals. Grist *et al.* (2002) describe these improvements based on Davison & Hinkley (1997) and Efron & Tibshirani (1993). The two most frequently applied nonparametric transformations produce bias-corrected (BC) and bias-corrected and accelerated (BC_a) percentile confidence intervals. Bias correction is actually a centering adjustment that corrects for median bias. It ensures that the resulting bias-corrected confidence interval is median unbiased in the sense that the proportion of bootstrap estimates less than or equal to the sample statistic contained is equal to 50%. The coverage accuracy of the BC confidence interval is improved further by a correction for skewness achieved by employing the quantity referred to as the acceleration constant. Efron (1982) discussed these different methods. The percentile method gives somewhat erratic results, both in terms of the length of the intervals and of their skewness. The BC_a percentiles require more computation but are more accurate. Grist *et al.* (2002) showed indeed an expected improvement of the HC_5 estimate and its confidence interval for some examples on potassium dichromate and lindane with sample size of 97 and 79 respectively. However, BC and BC_a confidence intervals did not improve for examples on ammonia, benzene and endosulfan with sample size of 27, 28 and 76 respectively. This suggests that for small sample sizes, the BC and BC_a confidence intervals fail to improve the basic confidence intervals. Therefore, these improvements were not considered here and the simplest method, the percentile method, will be used.

In order to construct uncertainty bands on the SSD as in Figure 1, one simply needs to calculate and plot the 1st, 2nd, 3rd, … 99th percentile (instead of only the 5th percentile or HC_5) of the SSD and their respective 90% confidence intervals.

3.2.3.2. Methods from Classical Statistics

In classical statistics, reasoning often refers to properties of infinitely repeated samples. For this method, simple formulae and lookup tables from literature for specific parametric distributions are used. A 5th-percentile can be calculated according to the following equation (depending on the lognormal or loglogistic distribution):

$$\alpha^{th} - percentile = \bar{x} - K \cdot s$$

with $\bar{x}$ the mean of the log-transformed data set
s the standard deviation of the log-transformed data set
K a tabulated extrapolation factor depending on the sample size n

There are now not only K-values for calculating the 5th-percentile but also for calculating the lower and upper 90% confidence limit. These are tabulated in Aldenberg & Jaworska (2000) for the lognormal distribution and in Aldenberg & Slob (1993) for the loglogistic distribution.

For example, let $\bar{x} = 2$ and $s = 3$ be the sample mean and sample standard variation of a log-transformed sample of 5 data points. The 5th-percentile and its lower and upper 90% Confidence Interval (CI) are calculated as:

$$5^{th} - percentile = 2 - 1.7793 \cdot 3 = -3.3379 \Leftrightarrow \mathrm{HC}_5 = e^{-3.3379} = 0.0355$$
$$lower\ CI\ of\ 5^{th} - perc. = 2 - 4.2027 \cdot 3 = -10.6081 \Leftrightarrow lower\ CI\ of\ \mathrm{HC}_5 = e^{-10.6081} = 0.0000247$$
$$upper\ CI\ of\ 5^{th} - perc. = 2 - 0.8178 \cdot 3 = -0.4534 \Leftrightarrow upper\ CI\ of\ \mathrm{HC}_5 = e^{-0.4534} = 0.635$$

3.2.3.3. Bayesian Inference Methods

Bayesian statistical methods reverse the role of sample and distribution: the sample is fixed and unique, and the distribution itself is uncertain. This statistical viewpoint corresponds better to the practical situation the individual researcher is facing: there is only one sample and there are doubts what distribution to use, or, if the distribution is chosen, what values the parameters will take. The uncertainty of the distribution is modelled by assuming that the parameters of the distribution are distributed (Aldenberg & Jaworska, 2000).
If one assumes parameter values to be distributed, one has to presuppose a so-called (in this case a non-informative) prior distribution for the parameters, to specify the initial state of knowledge about them, before the data are used. The prior distribution is transformed into the so-called posterior distribution by multiplication with the classical likelihood function, by which the information in the

data is introduced. This is essentially Bayes' theorem. The posterior distribution summarises our increase in knowledge about the parameters due to observing the data. A Bayesian simulation focuses on the evaluation of the joint posterior distribution of the parameters. For further technical details the reader is referred to Box & Tiao (1973).

This Bayesian method for confidence interval estimation was not programmed or used as such. Instead, *K*-values are found from a lookup table from Aldenberg & Jaworska (2000) for the lognormal distribution. Again, the bounds of the uncertainty or confidence interval are calculated as in 3.2.3.2. Note that the correct term for the intervals is actually credibility intervals but these will be named here as Bayesian confidence intervals.

3.2.4. Simulation Study

When comparing alternative approaches for quantifying parameter uncertainty, criteria that are important to consider include the variance of the original data set, and the bias and coverage of the confidence intervals generated by each method. In a simulation study, the bias of the HC_5 estimate and the coverage of its uncertainty interval are found by repetitively taking random samples of the presupposed target distribution *F*. For risk assessment, the most desirable estimation method is one that deals well with high variance, yields confidence intervals that are sufficiently wide (i.e., the confidence interval does not underestimate the probability of enclosing the population parameter), and, more specifically, yields upper confidence limits that are not biased low. The choice of the most appropriate method will depend on the characteristics of the data set and a balance between two objectives: (1) the desire to be ecologically protective and, therefore, have a low probability of underestimating the mean, and (2) a desire to be accurate, in the sense of choosing a method whose expected coverage equals the true coverage. As a general principle for quantitative uncertainty analysis, if alternative methods yield very different answers, it is helpful to explore the reasons for the differences (EPA, 2001).

In this simulation study, a target distribution *F* was assumed: a lognormal distribution ($\exp[N(\mu,\sigma)]$ with $\mu = 2$ and $\sigma = 1$). The arithmetic mean of the target distribution equals $\exp[\mu + 0.5\sigma^2] = 12.2$. Next, 20 positive values (see Table 2) were randomly drawn from *F*. These form the estimated target distribution $\hat{F}$. The arithmetic mean of this sample equals 14, exactly. The target distribution, the sample and its estimated target distribution are shown in Figure 4.

Table 2: The hypothetical data set of 20 sample points, data values drawn from a lognormal distribution of the form exp[N(μ,σ)] with μ = 2 and σ = 1 (true HC_5 = 1.42639)

0.832858	2.573425	3.724999	9.227466	14.99063
0.903766	2.602635	4.258860	10.80821	15.05903
1.821690	2.659332	6.221531	10.85469	18.03431
2.463967	3.689074	8.331888	14.64650	24.97989

Figure 6 gives the 90% uncertainty intervals of the 5th percentile (HC_5) obtained with all methods tested on the hypothetical data set. First, the parametric methods are discussed. The method from classical statistics and the Bayesian approach lead to the same results, as Aldenberg & Jaworska (2000) already concluded. The parametric bootstrap results are similar to the Bayesian analysis results, although with slightly wider confidence intervals. The true HC_5 (1.42639) lies within the 90%-uncertainty interval of all parametric methods. This is to be expected, as the data set is lognormally distributed. The methods are good, given the correct assumption of the distribution.

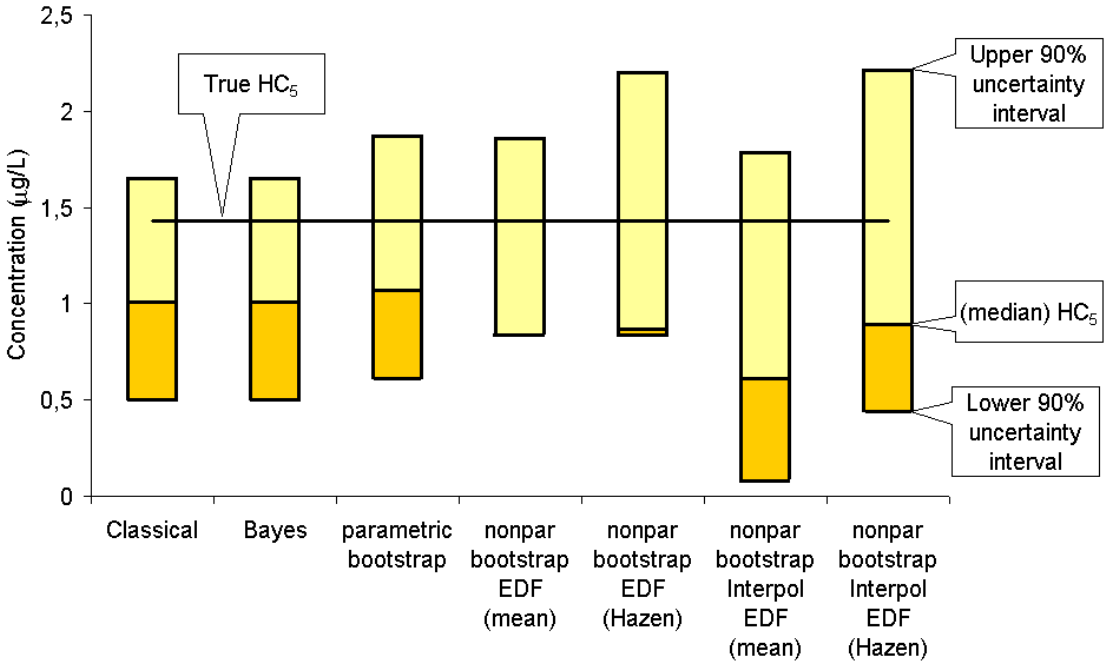


Figure 6: 90% uncertainty or confidence intervals of the HC_5 following various methods for 20 data points of the hypothetical lognormal data set (thick line = true 5th-variability percentile)

Second, the nonparametric methods were evaluated. The lower 90% uncertainty limit and the HC_5 percentile are (almost) equal for the resampling procedure (nonpar bootstrap EDF in Figure 6). In case of small data sets, the 5th-percentile has to be estimated between zero and the first point. If the empirical distribution function (see Figure 5) is used, the 5th-percentile is the first point itself. In other words, the uncertainty interval is bounded by the first (and smallest) data point (namely

0.832858). As a result, the first data point is selected many times. When linear interpolation between zero and the first point is used (as in the interpolated empirical distribution function, see Figure 5), the lower 90% uncertainty limit is not bounded by the first point and as a result linear interpolation accounts for the possibility that the lower 90% uncertainty limit can be smaller than the first data point. The mean and Hazen plotting system (in case of nonparametric bootstrapping with interpolated EDF) show significant differences. A factor of 4 was observed between the minimum and the maximum of the estimated lower 90% uncertainty limit.

Note that Figure 6 is only one possible realisation of confidence intervals. Therefore, as a validation exercise, 20 new data points from the same hypothetical lognormal target distribution were sampled. The uncertainty interval of the HC_5 was again estimated. The coverage of the true HC_5 over the uncertainty interval was checked for every method. If this process is repeated 1000 times, the uncertainty interval should cover the HC_5 percentile 900 times, i.e. the method with coverage closest to 90% should be considered as the most suitable method. The results of this simulation study are shown in Table 3.

Table 3: Coverage, or percentage (%) of the samples that the actual HC_5 value is included in the 90% confidence interval, calculated for different methods and distributions

Method	"Distribution"		Coverage (%)
1) Method from classical statistics	Lognormal		90.6
2) Bayesian statistics	Lognormal		90.6
3) Bootstrapping:			
- Parametric	Lognormal		88.9
- Nonparametric (resampling)	EDF	Mean	58.6
	EDF	Hazen	63.6
- Nonparametric	Interpolated EDF	Mean	93.7
	Interpolated EDF	Hazen	94.5

Differences between methods are mostly determined by the choice of the probability distribution. The three parametric methods assuming lognormal distribution give similar results. The results show that these parametric methods also give the best results. This is to be expected as the hypothetical data set is lognormally distributed. The nonparametric bootstrapping, based on an interpolated EDF, overestimates the uncertainty interval, which is expected as nonparametric techniques tend to have larger uncertainty estimates. All in all, the coverage is sufficiently accurate.

The nonparametric resampling procedure, on the other hand, clearly underestimates the uncertainty. In conclusion, the assumption that the EDF $\hat{F}$ is assumed to be a good estimate of the real, unknown target distribution F does not seem to hold. It can even be shown theoretically that the basic bootstrap interval will not in general achieve this coverage (Efron & Tibshirani, 1993).

3.2.5. Case Studies

Three case studies on the use of several techniques for uncertainty and variability estimation are discussed here. The focus in the first case study (on LAS and Cu) is on estimating HC_5 and its uncertainty interval using several parametric and nonparametric techniques. The second case study (on Cd and a hypothetical chemical) focused on the influence of upper and lower outliers on the HC_5 and its uncertainty interval using parametric and nonparametric techniques. The third case study (on Zn) in collaboration with EURAS bvba (Ghent, Belgium) focused on the comparison between several parametric distribution distributions, more in particular the difference between threshold and non-threshold distributions.

3.2.5.1. Species Sensitivity Distribution for LAS and Cu

These data sets come from laboratory and field measurements reported in literature. They consist of a toxicity database of Cu (Copper - 20 data points) and LAS (Linear Alkylbenzene Sulfonate - 17 data points), which can be found in Versteeg *et al.* (1999).

Distribution Selection

Both data sets are lognormal distributed according to the Kolmogorov Smirnov statistic for lognormality. To explore lognormality, normal QQ-charts or probability plots of the log-transformed data were plotted (see Figure 7).

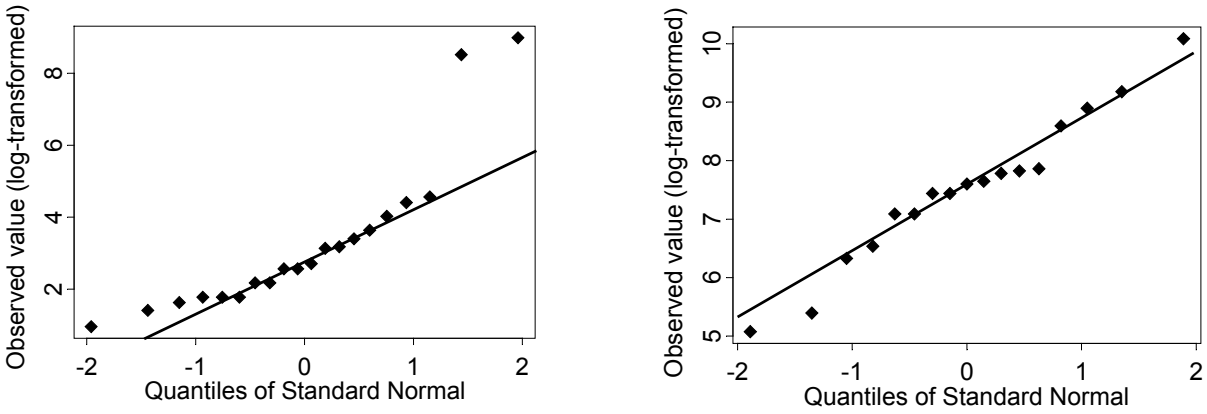


Figure 7: QQ-plots for log Cu (left) and log LAS (right)

The QQ-plots indicate that the data are lognormal distributed around the mean, but tend to deviate at the tails, especially the upper tail for the Cu data set (Figure 7 left). A better fitting distribution

could be selected for the Cu data set especially because the lognormal distribution can lead to protection levels (PNEC) below normal background levels, particularly for metals. For example, Hopkin (1993) showed that use of SSDs could lead to supposedly toxic PNECs for essential metals below those that are a physiological requirement for beneficial soil arthropods. Other distribution types will be investigated in the third case study. Attention in this case study was focused on the difference between parametric and nonparametric methods.

Uncertainty estimation

All previously discussed methods were assessed for their performance in calculating the HC_5 and its uncertainty/confidence estimates.

An illustration of the results for LAS is shown in Figure 8. One can conclude that there is a distinct difference in shape between the parametric and nonparametric distributions estimated from the data. The parametric methods tend to produce smoother and smaller uncertainty or confidence bands compared to the non-parametric methods. Nonparametric techniques follow more the data. This clearly demonstrates that the choice of distribution is an important, general problem. In order to select a lower bound, it is necessary to specify both the desired percentiles of variability and uncertainty. For example, one point estimate would be the 5^{th}-percentile of uncertainty for the 10^{th}-percentile of variability (point A in Figure 8).

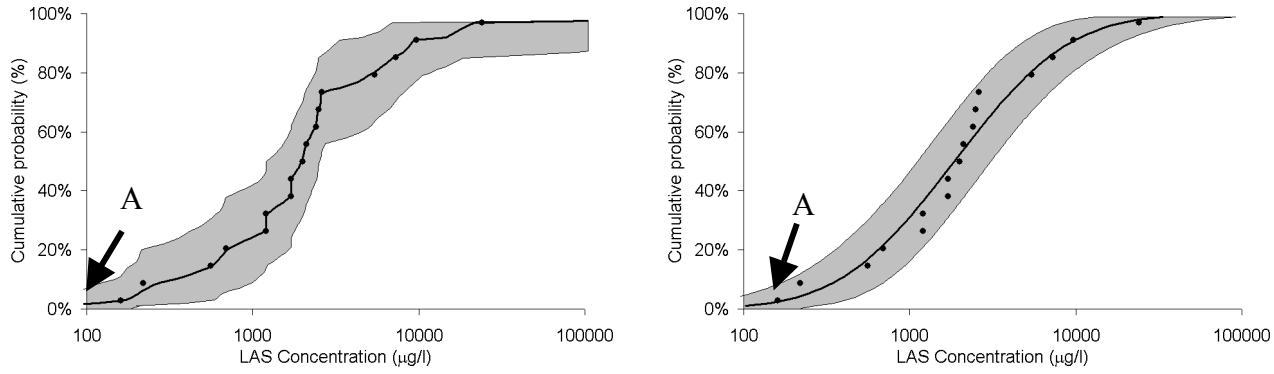


Figure 8: SSD and its 90%-uncertainty band for the LAS data set based on non-parametric bootstrapping (Interpolated EDF and Hazen plotting method) (left) and parametric bootstrapping (lognormal distribution) (right)

For LAS, a possible HC_5 could be identified (median of all methods is around 200 µg/l) that could be situated within the 90%-uncertainty intervals of all methods (see Figure 9). However for Cu, no possible HC_5 could be identified that would lie within the 90%-uncertainty intervals of all methods, since these do not overlap (see Figure 10). For LAS, a factor of 2.4 and for Cu, a factor of almost 5 was found between the minimum and the maximum of the HC_5 estimated by the various methods.

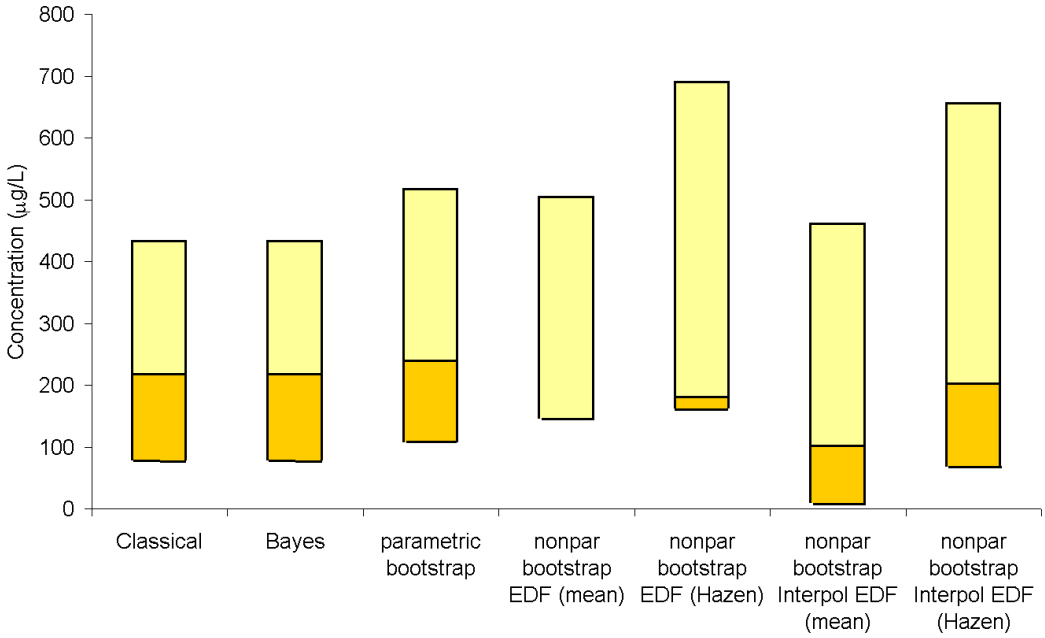


Figure 9: 90% uncertainty or confidence intervals of the HC_5 following various methods for 17 data points of the LAS data set (concentration in µg/l)

From Figure 10, it can be seen that the results are very sensitive to the choice of the assumed distribution (parametric or non-parametric). Furthermore, as already outlined above, the influence of potential outliers may not be underestimated. A detailed outlier study should be performed.

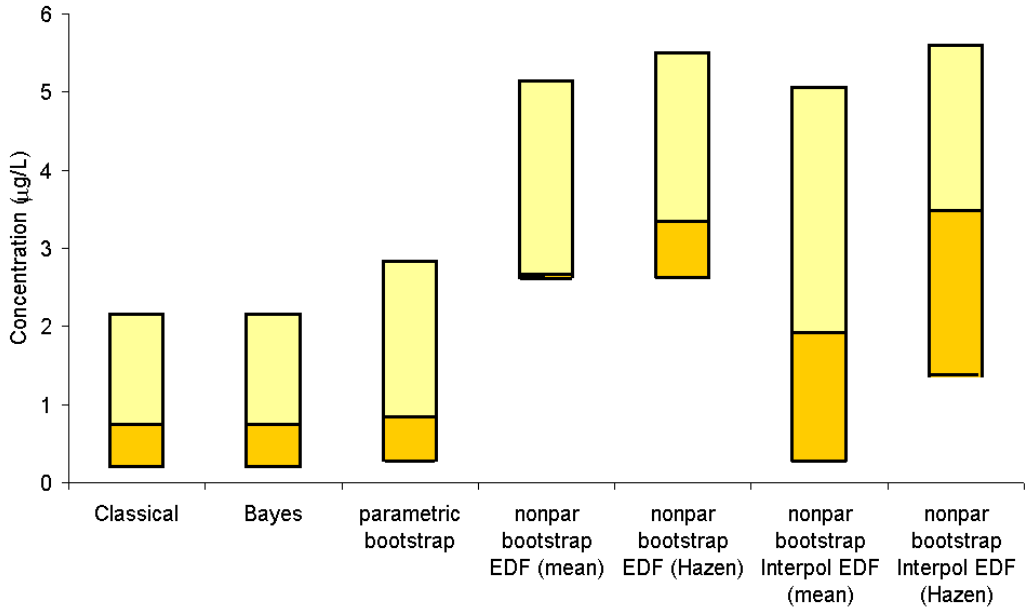


Figure 10: 90% uncertainty or confidence intervals of the HC_5 following various methods for 20 data points of the Cu data set (concentration in µg/l)

3.2.5.2. Species Sensitivity Distribution for Cd and a Hypothetical Chemical X

The Cd data set (Cadmium - 19 data points) comes from laboratory measurements, which were compiled from literature databases and Janssen (2001). The hypothetical chemical X (18 data points) was also compiled based on realistic ecotoxicity tests found in literature. This second case study focuses on the influence of outliers or extreme values on the HC_5 and its uncertainty interval using parametric and nonparametric methods.

Distribution Selection

To explore lognormality, normal QQ-plots of the log-transformed data were again made (see Figure 11).

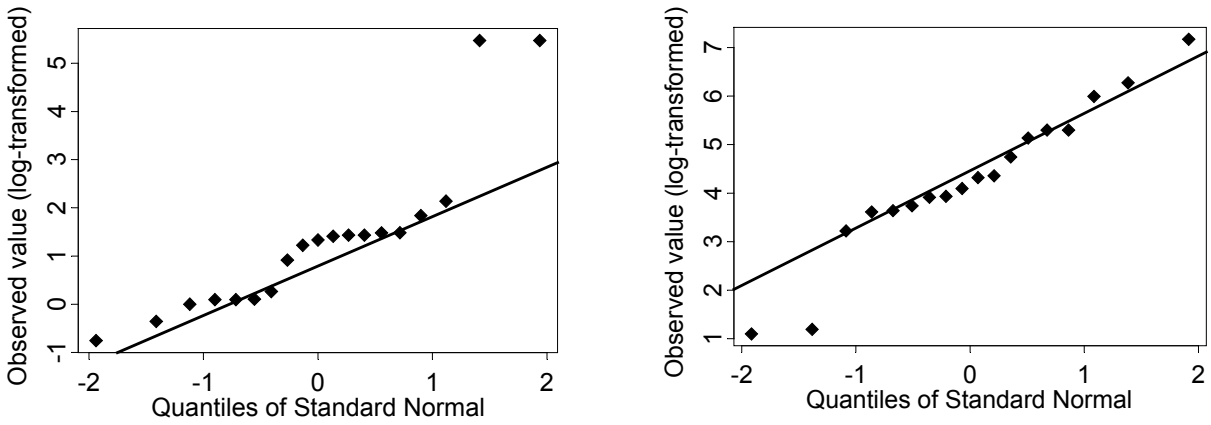


Figure 11: QQ-plots for log Cd (left) and log X (right)

Only the X data set is lognormal distributed according to the Kolmogorov Smirnov statistic for lognormality. The QQ-plots also indicate that the data are lognormal distributed around the mean, but tend to deviate at the tails, especially the upper tail for the Cd data set (Figure 11 left). The bad fit at the tail could be caused by outliers. Several objective criteria exist to detect an outlier. These will not be used here. Outliers were detected based on visual inspection of cumulative distribution plots and QQ-plots. If a certain point deviates too much from the others, it is regarded as a potential outlier. This criterion is obviously very subjective, but this case study should be regarded as a sensitivity analysis of extremely low or high values rather than an outlier study. The Cd data set has two potential upper outliers whereas the X data does not have any outliers based on its QQ-plot although it has two extreme low data points.

Uncertainty estimation with and without outliers or extreme values

The influence of outliers was investigated by removing or adding a potential outlier. The parametric and nonparametric bootstrap (using the interpolated EDF and Hazen plotting) were assessed for their performance in calculating the HC_5 and its uncertainty/confidence estimates.

The results of the influence of the two upper outliers in the Cd data set are shown in Figure 12 for nonparametric and parametric bootstrapping. The numeric values for the HC_5 and the PNEC are presented in Table 4.

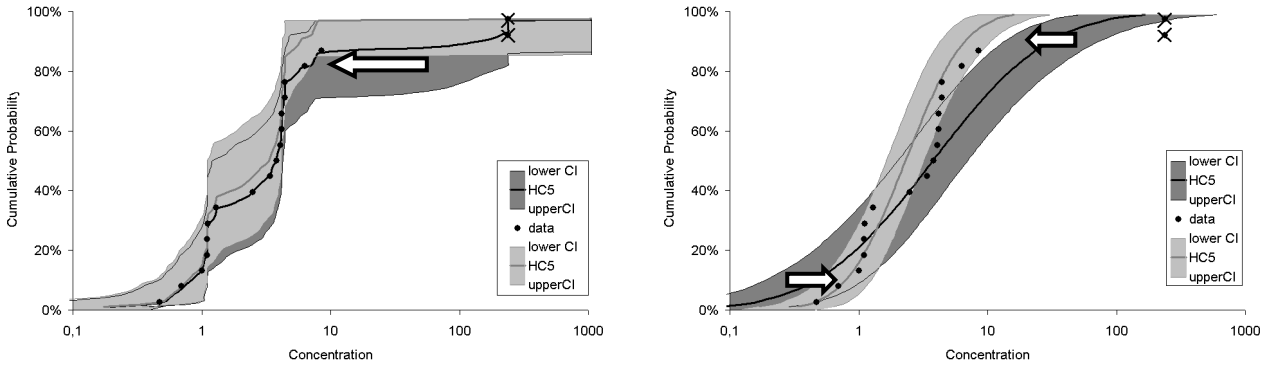


Figure 12: SSD-curves + uncertainty band of the Cd-toxicity data set according to nonparametric and parametric bootstrapping with and without upper outliers or extreme values

For calculating HC_5, the nonparametric methods are almost insensitive to two upper outliers in contrast with the parametric methods (for the Cd data set; a factor 2.5 was observed for HC_5). The data with outliers are not lognormally distributed according to a Kolmogorov Smirnov (KS) test, while the data set without the outliers is consistent with a lognormal distribution according to the KS test, i.e. removing the outliers improves the parametric fit. Removing an upper outlier only affects the uncertainty band at the upper tail for the nonparametric methods, while the uncertainty bands calculated using parametric methods are affected at both the upper and lower tail.

The results of the influence of two lower extreme values in the X data set are shown in Figure 13 for nonparametric and for parametric bootstrapping. For calculating HC_5, the nonparametric methods appear to be very sensitive to lower outliers (for the X data set, a factor 7.2 was observed for HC_5). This can be expected as the lowest data points are used for HC_5 calculation. In the examples, the parametric methods result in similar sensitivities to upper and lower outliers or extreme values.

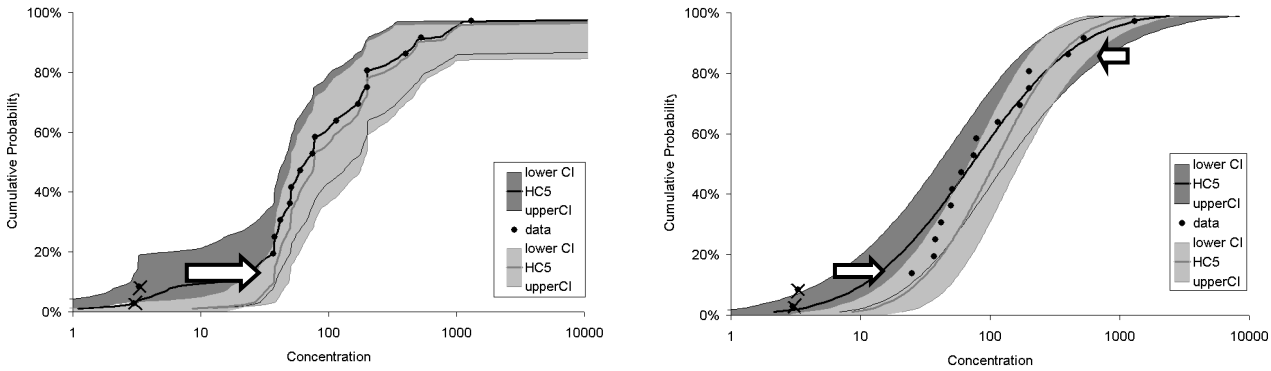


Figure 13: SSD-curves + uncertainty band of the X-toxicity data set according to nonparametric and parametric bootstrapping with and without lower outliers

Table 4: Influence of upper (for Cd) and lower outliers or extreme values (for X) on HC_5

			With outliers	Without outliers	Factor
Cd	Lognormal distribution?		No	Yes	
	Parametric	HC_5	0.25	0.59	2.4
		Lower CI*	0.17	0.47	2.8
	Nonparametric	HC_5	0.60	0.58	0.97
		Lower CI*	0.44	0.40	0.91
X	Lognormal distribution?		Yes	Yes	
	Parametric	HC_5	5.97	17.97	3.0
		Lower CI*	4.12	13.46	3.3
	Nonparametric	HC_5	4.07	29.33	7.2
		Lower CI*	2.65	19.46	7.3

* 25^{th}-perc of HC_5 (= lower limit of a 50% confidence interval)

3.2.5.3. Species Sensitivity Distribution for Zn

In the framework of the European new and existing chemicals policy, an overall risk assessment for zinc (Zn) is being prepared. Traditionally, risk assessments in the European framework are performed according to the methodology laid down in the Technical Guidance Document (EU-TGD, 1996). The potential risks are typically estimated in a deterministic way using point estimates. Currently, overly conservative assumptions are used in an attempt to account for the uncertainty. Recently, the use of probabilistic approaches for characterizing effects in SSDs has been suggested (Van Straalen & van Leeuwen, 2002).

The main objective of this case study is to address the ongoing discussion in the European Union Zn Risk Assessment Report (EU-Zn RAR) and use its Zn data in a probabilistic way in order to provide a quantifiable and improved effects assessment. For this, several parametric (and nonparametric) distributions will be assessed to model the Zn ecotoxicity data. This study was part of collaboration with EURAS. More details of the study can be found in Van Sprang *et al.* (2002).

Distribution Selection

Selection of the most appropriate distribution was based on several criteria: goodness-of-fit statistics, expert knowledge and graphical exploration (e.g. QQ-plots or normal probability plots). Goodness-of-fit statistics (Anderson-Darling test statistic) were generated and compared among the different distributions. The smaller the discrepancy between the hypothesized and the observed distribution, the better the fit. Lognormality was explored using normal QQ-plots of the log-transformed data (see Figure 14). The investigated Zn data set was not lognormal distributed according to the Kolmogorov-Smirnov and Anderson-Darling test statistics. The QQ-plot also indicates that the data deviates from lognormal at the lower tail (see Figure 14). As a consequence, the HC_5 derived from a lognormal distribution is expected to be an underestimation. Similar conclusions were drawn for the other investigated parametric non-threshold distributions (i.e. distributions going through zero such as lognormal, extreme value, Pearson, Weibull, loglogistic). Results not shown here but can be found in Van Sprang *et al.* (2002). Among all non-threshold distributions, the best fit was obtained with the lognormal distribution (both based on graphical inspection and statistical tests). The lognormal fit is shown in Figure 15.

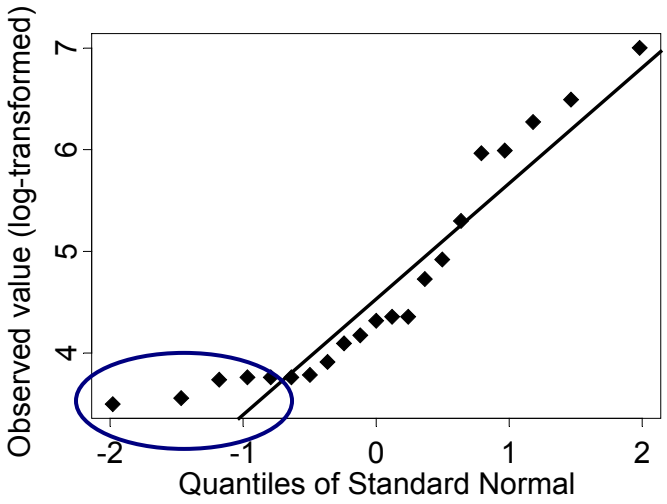


Figure 14: QQ-plot for the log-transformed Zn data

Parametric threshold distributions such as the Pareto, Beta and Triangular distribution do not force the lower tail of the SSD to go through zero. They were fit to the log-transformed Zn data (only Pareto fit is shown here in Figure 15 but others can be found in Van Sprang *et al.* (2002)). Threshold distributions are justified if there is an underlying interpretation of the threshold. Because Zn is an essential metal, there is a certain concentration range of that metal required for normal metabolic functioning of the organism. Except for the triangular distribution, threshold distributions fitted the chronic Zn NOEC data more properly, especially in the lower tail. While the triangular distribution tends to underestimate toxicity in the lower tail, better fits were achieved with the Pareto, Beta and sigmoidal distributions. The best fit at the lower tail was achieved by the Pareto distribution on the log-transformed data (Figure 15).

Note that only for this case study, all distribution types (Pareto, Beta, extreme value, normal, …) except the lognormal distribution were fitted to log-transformed data. This means that these distribution types are not invoked as such but for ease of use, their names are not changed in this case study.

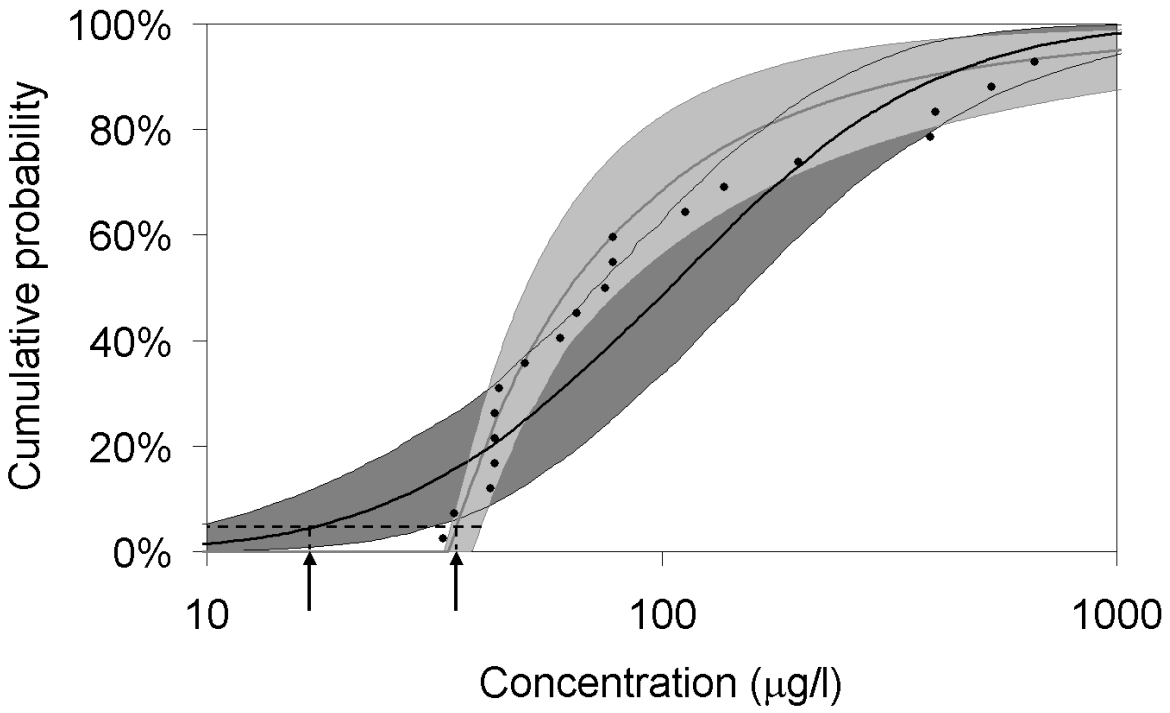


Figure 15: Lognormal (dark grey) and Pareto (light grey) species sensitivity distribution with uncertainty band on the Zn ecotoxicity data set (arrows indicate 5th percentile or HC_5)

Uncertainty estimation

In this study, bootstrap simulations are used to estimate the uncertainty due to sampling error. For each resampling run, 21 random resample points (i.e. the sample size of the chronic NOEC database for Zn) were taken from the estimated distribution (parametric or nonparametric). The number of resampling runs was set at 2000. Doubling the number of resampling runs to 4000 did not change the uncertainty estimates (data not shown).

Discussion

The estimated HC_5 values derived from non-threshold distributions ranged between 14 and 24 µg/l and were consistently lower than the better fitting threshold distributions. The fit of these latter distributions resulted in HC_5 estimates ranging between 27 and 35 µg/l. Threshold distributions produced smaller 90% confidence bounds compared to the non-threshold distributions, especially on the lower confidence limits, which are bound by the threshold value.

According to the Zn risk assessment report, an additional assessment factor of 2 on the HC_5 was applied. This is argued by the presence of 2 high quality chronic NOEC values for the cladoceran *Ceriodaphnia dubia* (i.e. 14 µg/l and 17 µg/l), which fall below the estimated HC_5 of 17.2 µg/l. However, the validity of the application of this additional assessment factor of 2 in the EU Zn risk assessment report for the protection of very sensitive species can be questioned when environmental risks are characterised using probabilistic techniques. In the effects assessment of a PERA, the entire SSD is considered instead of reducing the available information to a point estimate (as HC_5) prior to risk characterisation. The very sensitive organisms (i.e. those with NOEC < HC_5) will also be covered in the risk characterisation (see Chapter 4.1). Thus, the argument to introduce such additional assessment factor in order to be sufficiently protective for the most sensitive species in a risk assessment can be overruled.

3.2.6. Conclusions & Further Research

The reliability of several uncertainty and variability estimation methods at sample size 20 was compared and assessed by means of simulation and case studies.

The considered methods display varying robustness and accuracy in determining lower confidence limits of the HC_5. The most suitable methods to estimate lower end percentiles such as the 5^{th}-percentile were found to be the method from classical statistics, Bayesian analysis and nonparametric bootstrapping (using interpolated empirical distribution function and the Hazen plotting system).

At this stage, there is no direct reason to prefer parametric or nonparametric methods. However, the results are very sensitive to the choice of the method (a factor of 5 difference was observed when results from different methods were compared). Differences between methods are for a large part determined by the choice of the probability distribution: parametric or nonparametric, threshold or non-threshold distributions. Consequently, the proper use of distribution selection methods should not be underestimated. Statistical tests, graphical exploration and expert knowledge can help in identifying the appropriate distribution.

Some nonparametric methods should not be used for estimating low percentiles given a small sample size. All resampling techniques showed they were rather arbitrary and inaccurate because they are bounded by the smallest data point.

For estimating 5^{th}-percentiles of small sample sizes, the Hazen plotting and the mean plotting system are used in literature but one should be aware that both systems give different results (a factor of 2 was observed here) at low sample sizes (see also Aldenberg & Jaworska (2000) and Chapter 3.3).

Further research on the influence of sample size may reveal more information. This is the topic of Chapter 3.3.

Chapter 3.3

-

Uncertainty & Variability Estimation: Sample Size Issues

Chapter 3.3

Uncertainty & Variability Estimation: Sample Size Issues

After introduction of Species Sensitivity Distributions (SSDs) to assess effects on the ecosystems and to derive environmental quality criteria, the method has been extensively debated. The most common current approach is to derive Predicted No Effect Concentration (PNEC) from the (median) 5th percentile of the SSD (EU-TGD, 1995). One of the remaining issues in SSD determination is the choice of sample size. This issue is less important for the Exposure Concentration Distribution (ECD) determination because usually, either enough monitoring data are available or exposure models are used. Therefore, this chapter, that reviews different criteria for determination of an appropriate sample size, will focus on SSDs.

3.3.1. Introduction

One of the other remaining issues in SSD determination is the choice of sample size. The choice of an appropriate sample size is an essential component of any experimental design. Two important considerations need to be made to determine the appropriate sample size. First, the accuracy and scientific reliability of the method to estimate the 5th percentile should be assessed. Some methods cannot be applied at small sample sizes (say < 20). Second, the desired level of precision should be defined and assessed. Several papers (see Table 1) have already been published on the determination of a minimum or optimal sample size for SSDs. However, each researcher uses his/her own considerations to estimate the 5th percentile.

Table 1: Overview of proposals on minimum or optimal sample size for SSDs (Species Sensitivity Distributions), together with the criterion considered to obtain that size (consideration 1: the level of scientific reliability, consideration 2: the desired level of precision in the estimation)

Reference	Consideration	Minimum/optimal sample size
(Stephan *et al.*, 1985)	Unknown	8
(van Leeuwen, 1990)	2	5
(Baker, 1994)	Unknown	4-8
(Cowan *et al.*, 1995)	1	20
(Solomon, 1996)	1	9
(Roman *et al.*, 1999)	Both	No proposal
(Vega *et al.*, 1999)	1	10
(Newman *et al.*, 2000)	2	15-55
(Van Der Hoeven, 2001)	2	No proposal
(Wheeler *et al.*, 2002)	1	10

Consequently, different sample sizes for SSDs are proposed depending on the criterion and method used (in an often incomplete analysis) (see Table 1). Therefore, there is a need for a proper, standardised and scientifically sound procedure to determine a minimum sample size.

This chapter has two main objectives. First, some studies and methods in literature on sample size determination will be reviewed in detail. Second, this chapter aims to clarify and illustrate these two important considerations for sample size determination. This will be done via a combination of a literature study and own simulation results.

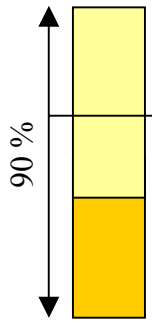
3.3.2. Terminology, Simulation and Case Studies

Several methods for estimating variability and uncertainty can be used (more information can be found in Chapter 3.2).

Let $x_1, x_2, \ldots, x_n$ be an independent and identically distributed random sample from x. This variable has distribution function $F(x)$, density function $f(x)$ and depends on some parameters $\theta_1, \theta_2, \ldots, \theta_m$. The sample mean is $\bar{x}$ and the sample standard deviation is s. The 5^{th} percentile of x or the HC_5 of SS (Species Sensitivity) is the estimator of interest.

The terminology to be used remains the same as in Chapter 3.2 and is shown in Figure 1. The vertical bar represents a 90% uncertainty or confidence interval of the 5^{th} percentile or HC_5 taken from the SSD. This uncertainty interval has an upper 90% confidence or uncertainty limit (equal to

the 95th percentile of the uncertainty distribution of the HC_5) and a lower 90% confidence or uncertainty limit (equal to the 5th percentile of the uncertainty distribution of the HC_5). The line in the uncertainty bar is the estimate of the 5th (variability) percentile or HC_5. It is estimated as the median of the uncertainty distribution. For ease of use, this will here be named the 5th percentile in a statistical context and HC_5 in a risk assessment context.



→ **upper 90% confidence or uncertainty limit** = 95th- uncertainty percentile

→ **true 5th-variability percentile**

→ median **5th**-variability **percentile** = median 5th-percentile of the variability distribution = 50th-percentile (median) of uncertainty of the 5th-percentile of variability = median 90 % confidence limit = estimate of **HC_5** (Hazardous Concentration 5%)

→ **lower 90% confidence or uncertainty limit** = 5th- uncertainty percentile

Figure 1: Terminology and visualisation

The distinction between a target population (represented by distribution *F*) and an estimated target population (represented by distribution $\hat{F}$) should be considered carefully as described in Chapter 3.2. In the following sections, some hypothetical target distributions will be assumed for some simulation studies. The first hypothetical target distribution *F* is a lognormal distribution with parameters mean $\mu = 2$ and standard deviation $\sigma = 1$ ($X \sim \exp[N(\mu,\sigma)]$). The second hypothetical target distribution *F* is also a lognormal distribution but with parameters $\mu = 5$ and $\sigma = 2.82$ ($X \sim \exp[N(\mu,\sigma)]$). Both distributions have the same 5th-percentile (namely 1.43). Other lognormal distributions, like the standard lognormal distribution ($\exp[N(\mu,\sigma)]$ with $\mu = 0$ and $\sigma = 1$), could also be studied but will give similar results as every lognormal distribution can be transformed into another one. The third hypothetical target distribution *F* is a transformed lognormal distribution ($X \sim \exp[N(5,2.82)]+100$). The idea behind this third hypothetical distribution is to create a "lognormal-like" data set with bad fit at the lower tail. To explore lognormality, normal QQ-plots of the log-transformed data were plotted. The normal QQ-plot for this last data set is shown in Figure 10a.

These target distributions can be used for simulation studies to assess the bias and coverage of the confidence intervals. In statistics, an estimate is unbiased if the expected value of a statistic as the mean or 5th percentile of the estimated target distribution $\hat{F}$ is equal to the true value of the parameter:

$$bias = E\left[\hat{HC}_5\right] - HC_{5,true}$$

Similarly, a method has accurate coverage if the probability p that a confidence interval does not cover the true parameter is equal to the probability level used to construct the confidence interval. Confidence intervals are expected to enclose a true but unknown parameter according to that specified probability, such as 90% or 95%. This is the expected coverage of the confidence interval, given a specified significance level. The difference between the expected coverage and the actual coverage is one metric for evaluating statistical methods that yield different confidence intervals. In a simulation study, these coverage and bias can be found by repetitively taking random samples of the target distribution *F*. For risk assessment, the most desirable estimation method is one that deals well with high variance, yields confidence intervals that are sufficiently wide (i.e. the confidence interval does not underestimate the probability of enclosing the population parameter), and, more specifically, yields upper confidence limits that are not biased low. The choice of the most appropriate method will depend on the characteristics of the data set and a balance between two objectives: (1) the desire to be ecologically protective and, therefore, have a low probability of underestimating the mean, and (2) a desire to be accurate, in the sense of choosing a method whose expected coverage equals the true coverage.

The other data sets discussed in this chapter consist of real toxicity databases (NOECs) of respectively Cu (Copper) and LAS (Linear Alkylbenzene Sulfonate), which can be found in Versteeg *et al.* (1999). For the Cu data set, the lognormal distribution fits well around the mean, but it fits badly at the tails (see QQ-plot in Figure 2 left). For the LAS data set on the other hand, the lognormal distribution fits very well (see QQ-plot in Figure 2 right).

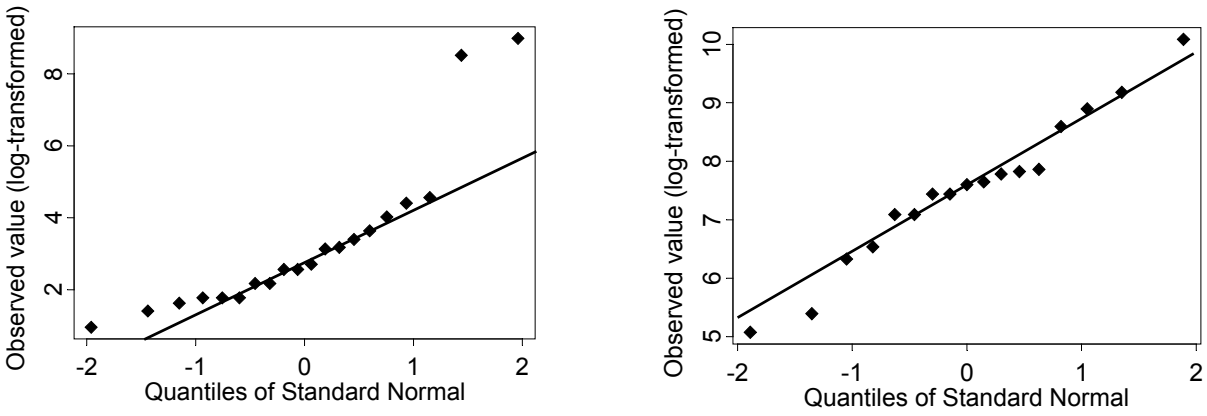


Figure 2: QQ-plots for Cu (left) and LAS (right)

3.3.3. Comments on Existing Sample Size Determinations for the SSD

Each researcher uses his/her own method and criteria to determine a minimum or optimal sample size. This makes it extremely difficult to compare several sample size studies. Unfortunately, some of these sample size derivation methods are, in our opinion, wrong or could be significantly improved. In this section, two of such methods are reviewed in detail and criticised. Both deal with resampling and its underlying theory.

3.3.3.1. Comments on Subsampling for Sample Size Plots

In sample size determination problems, plots are often made showing the sample size on the abscis and some parameter estimate (HC_5 in this case) and its uncertainty interval on the ordinate (e.g. see Figure 3b). Given a data set of n data points, the construction of these plots can be done in many ways. For every sample size smaller than or equal to n, the HC_5 and its uncertainty interval needs to be calculated. Next, it is discussed whether to subsample with or without replacement from the original data set in order to construct these sample size plots.

The "without replacement"- sampling sequence strategy takes random subsamples from the original data set without replacement. This corresponds with the practical point of view in risk assessment. A particular species sensitivity can only be observed once (e.g. a species can only die once). Unfortunately, only a limited number of sampling sequences are possible. For sample size m given a data set of size n (with $m < n$), the number of combinations is $\frac{n!}{m!(n-m)!}$. Many sampling sequences are needed because by investigating only a few selected sampling sequences, only a few possible realisations of many possible sequences are evaluated and consequently they are highly dependent on coincidence. Therefore, the sampling needs to be repeated many times in order to study all possible combinations. For the “without replacement”- sampling sequence strategy, the HC_5 and its uncertainty interval should only be calculated for sample sizes less than or equal to $n/2$ because the number of combinations is too small when the sample size is larger than $n/2$ (Bros & Cowell, 1987). Roman *et al.* (1999), Bros & Cowell (1987), Van Der Hoeven (2001), Vega *et al.* (1999) and Wheeler *et al.* (2002) used the "without replacement"- sampling sequence strategy.

In the other case, the "with replacement"- sampling sequence strategy, replacement is allowed when subsampling. This corresponds better with the theoretical point of view that always starts with a new sample. This strategy is more general compared to the previous one. This strategy is used by Newman *et al.* (2000), Bros & Cowell (1987), Manly (1992)…

The effect of different sampling sequence strategies was illustrated for the LAS data set. The results are shown in Figure 3. The uncertainty intervals are smaller in case no replacement is used. This is

due to the fact that only a limited number of sampling sequences is possible for this strategy. In the example of Figure 3a, results should not be interpreted from sample size 9 (=18/2) onwards as described above. Because of the limitations of the “without replacement”- sampling sequence, only the "with replacement"- sampling sequence should be and was used for further simulations.

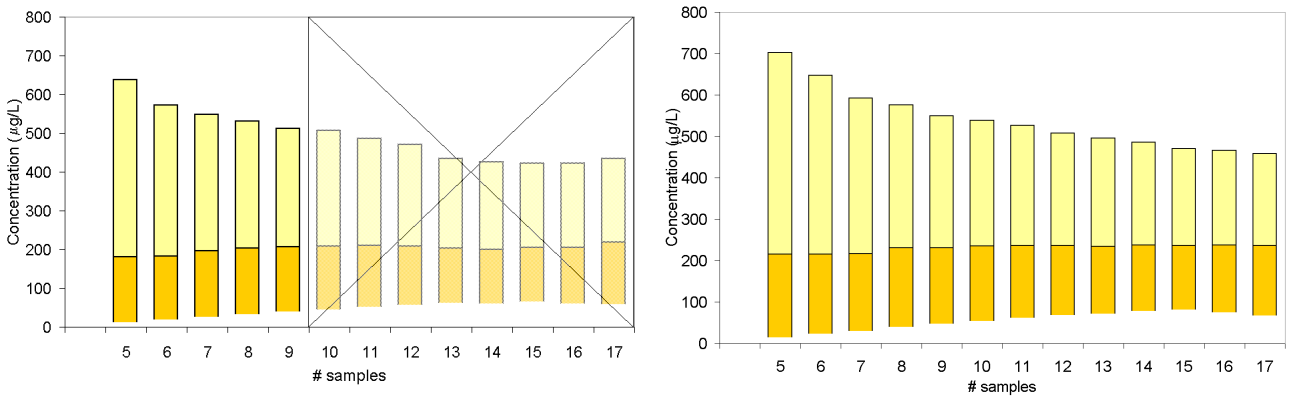


Figure 3: Sample size figures for LAS, (a) "without replacement"- sampling sequence, (b) "with replacement"- sampling sequence

3.3.3.2. Comments on Bootstrap Resampling Strategy Based on Newman et al. (2000)

3.3.3.2.1. Introduction

In this commentary the recent approach of Newman *et al.* (2000) is analysed to find optimal sample size and only the statistical considerations pertaining to accuracy of the prediction of the p^{th}-percentile are focused on.

Among other topics, Newman *et al.* (2000) describe a method for determining a sufficient number of species in SSDs using a modified version of bootstrapping. The novel modification concerns the use of a resample size bigger than the actual sample. Their results are: "*Approximate optimal sample sizes for HC_5 estimation ranged from 15 to 55 with a median of 30 species-sensitivity values. Similar sample sizes were needed for HC_{10} and HC_{20} estimation: estimates ranged from 10 to 75. ...These sample sizes are much higher than those recommended as acceptable for regulatory purposes*". Since indeed these numbers are rather high and different from existing recommendations of 5-8 depending on the source, Newman’s approach was investigated both theoretically and numerically.

3.3.3.2.2. Theoretical Considerations of Bootstrapping

The general approach in bootstrap simulation is to consider the empirical distribution function or to assume a distribution which describes the quantity of interest, to perform x replications of the data set of n by randomly drawing, with replacement, $m = n$ or smaller values, and then calculate x values of the statistic of interest (Efron & Tibshirani, 1993). However Manly (1992) writes: "... *One of the key aspects of bootstrapping is that samples are taken with replacement rather than without replacement. Hence, a pilot sample of size P* (here named *n*) *can be bootstrap-sampled to produce samples of any size, including sizes that are > P* (here named *m) (Bickel & Freedman, 1981)....*". This is somewhat misleading as Bickel & Freedman (1981) only make a statement on bootstrapping statistics when *m* and *n* are varied separately under the condition that both *m* and *n* have to tend to infinity i.e. they study the asymptotic properties only.

3.3.3.2.3. Newman et al. (2000) Approach

Newman *et al.* (2000) modified the bootstrap by taking more resample points m than the actual, original data set size *n*. In Newman's (2000) study, the resample size varied between 5 and 100 in increments of 5. For each resample size, the HC_5, HC_{10} and HC_{20} and their 95% confidence interval were calculated and plotted (Figure 4 for HC_{20}). Logically, the confidence interval around the HC_p estimate decreased as resample size increased. The resample size where no further visual improvement was found in narrowing of a confidence interval (note the subjectivity) was taken as the optimal *m*. Newman *et al.* (2000) set the optimal sample size *n* equal to the optimal resample size *m* found from such bootstrap study.

3.3.3.2.4. Simulation Examples

The Newman *et al.* (2000) procedure was investigated in two ways: first, for the convergence of the confidence intervals, as the resample size increases; second, for the consistency of the method by varying the initial sample size *n*.

First, the procedure was repeated for 60 data points for a Zn toxicity data set. The data only serve as an example, data quality and other SSD issues are not discussed here. The resample size varied from 5 to 1000 in order to investigate the convergence of the confidence intervals (Figure 4). The optimal resample size of 65 was selected as Newman (2000) did by a visual assessment.

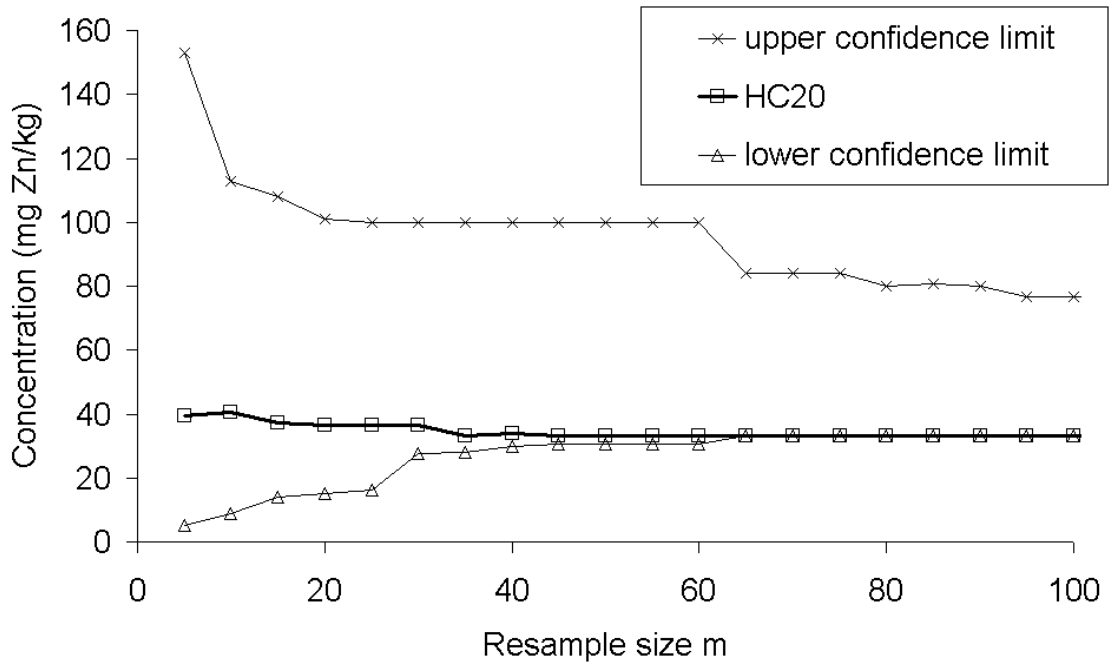


Figure 4: Curves for estimating sample size (HC_{20} as illustrated with the Zn data (60 points). The symbols indicate the HCp values ranked at the 50% (□), 5% (Δ), and 95% (X) of the 10,000 values generated by bootstrapping

Simulations show that as the resample size increases (becomes 1000), the confidence intervals converge to zero (Figure 5). In other words, the ideal resample size *m* would be infinity because the uncertainty would become zero and since these are simulations, it is no issue to resample as much as possible. This is in conflict with confidence interval theory that shows that there is always uncertainty at finite sample size.

Second, suppose the original data set contained 30 sample points instead of 60 (30 sample points were ad random removed), and the entire procedure was repeated, what would then be the optimal (re)sample size? Using the same criteria as above 40 is a good sample size (Figure 5). So, starting with 60 data points leads to an optimal sample size of 65, whereas starting from half of these data leads to an optimal sample size of 40. This indicates that the method for optimal sample size determination is not consistent and starting from different initial conditions leads to different conclusions. This is not due to lack of a statistical criterion for a cutoff. Using a possible criterion of $HC_{20}^{95}/HC_{20}^{50} < 2$ the recommended numbers would be 39 for 30 points and 360 for 60 points.

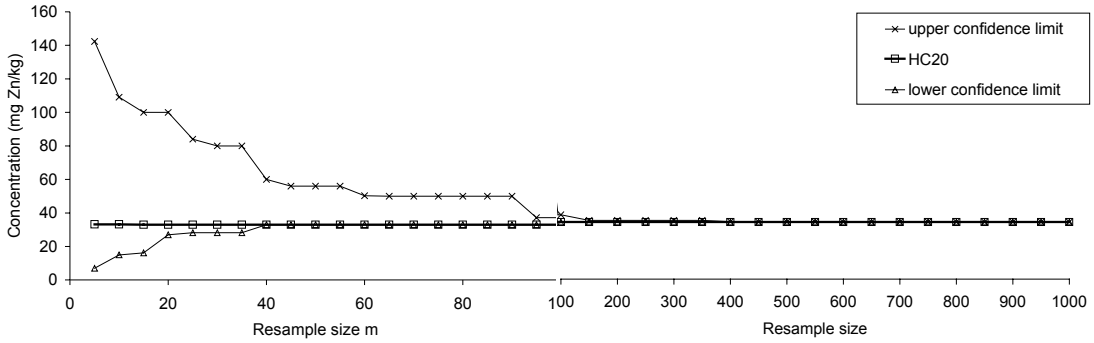


Figure 5: Curves for estimating sample size for (HC_{20}) illustrated with the limited Zn data (sample size is 30)

3.3.3.2.5. Conclusions

No theoretical background was found to support the approach of Newman *et al.* (2000). Further, simulation examples show that when the resample size exceeds the sample size logical and statistical inconsistencies arise. Indeed, one cannot get more information from a data set than the data itself contains. After this analysis we feel that Newman's *et al.* (2000) recommendations on sample size should not be considered.

This was later also confirmed by Grist *et al.* (2002) in their paper. The theory of the *m* out of *n* bootstrap is currently not well understood (Bickel *et al.*, 1997). Although the *m* out of *n* bootstrap percentile confidence intervals of Newman *et al.* (2000) generally converged with increasing *m*, there are no theoretical grounds for assuming that their deviation away from convergence provides information on minimum sample size. In particular, analysis of the situation in which resamples are generated with $m > n$ (in other words, where the bootstrap resamples contain more elements than the original sample) currently remains largely unexplored (e.g. Sakov & Bickel (2000)).

3.3.4. Sample Size Considerations

Two important considerations need to be made to determine the appropriate sample size. First, the accuracy and scientific reliability of the method to estimate the 5^{th} percentile need to be assessed (relevance of parameters, representativeness of toxicity data both in terms of number and kind of species and taxonomic groups (which were not discussed here)). Some methods cannot be applied at small sample sizes (say < 20). Second, the desired level of precision needs to be defined and assessed. A minimum sample size can then be determined based on this desired level of precision.

3.3.4.1. Level of Reliability

First, an appropriate method needs to be selected to estimate the 5th percentile or HC_5. Every estimation method has its own properties and is based on its own assumptions (see Chapter 3.2). However, some of these properties or assumptions may not be valid at small sample sizes (say < 20). Therefore, it should always first be checked if a HC_5 estimation method is reliable and accurate enough at small sample sizes. This can be done by investigating the underlying theory or by performing some simulation studies. This last one will be illustrated in this section.

3.3.4.1.1. Reliability of the Sample: Representativeness

All statistical techniques used for estimating uncertainty and variability in SSDs depend on the assumption that the data set is a randomly drawn sample from the real, target SSD. The EDF, constructed on the basis of the sample, will converge to the real, target SSD when the sample size increases. The experimental setup must provide the guarantee that the sample is a random selected one. Unfortunately, a lot of practical problems still need to be solved to guarantee this. From a statistical point of view, it is very difficult to quantify and investigate "representativeness". From an ecological point of view, it is also important to assess the representativeness of several taxonomic groups in a SSD. This is not discussed in this dissertation, as this issue needs a different (i.e. not statistically related) approach. Examples are given in e.g. Stephan *et al.* (1985).

3.3.4.1.2. Reliability of the Estimation Method

Simulation studies are a way to assess the reliability and accuracy of the statistical methods for small sample sizes. The bias and the variance of the HC_5 estimate and the coverage of the estimated uncertainty interval are then assessed for several sample sizes. Definitions of these concepts can be found above in section 3.3.2. Here, a simulation study of 1000 replications was performed on the first hypothetical target distribution. The bias and coverage will be exact for all sample sizes when the number of replications becomes infinity.

The results for the classical approach (according to Aldenberg & Jaworska (2000)) can be found in Figure 6. Based on the bias and coverage, a sample size of two is already sufficient provided the data were lognormally distributed and provided an average bias of 1.4 µg/l and a HC_5 variance of 14.7 µg/l is acceptable to the policy-maker. How acceptable a particular HC_5 bias or variance should be, is the main issue in the next consideration. The HC_5 bias and variance decrease as the

sample size increases. This corresponds with statistical estimation theory. The results also show that the coverage of the 90%-confidence interval varies around 90%, which was expected as these methods are based on this coverage concept (Aldenberg & Slob, 1993). This corresponds to earlier findings in literature (Aldenberg & Slob, 1993), (Kooijman, 1987): parametric approaches can result in very small sample sizes but these techniques are valid only if the assumption of the underlying distribution is correct. Hence, if the selected distribution assumption does not hold, parametric methods should not be used to obtain confidence intervals.

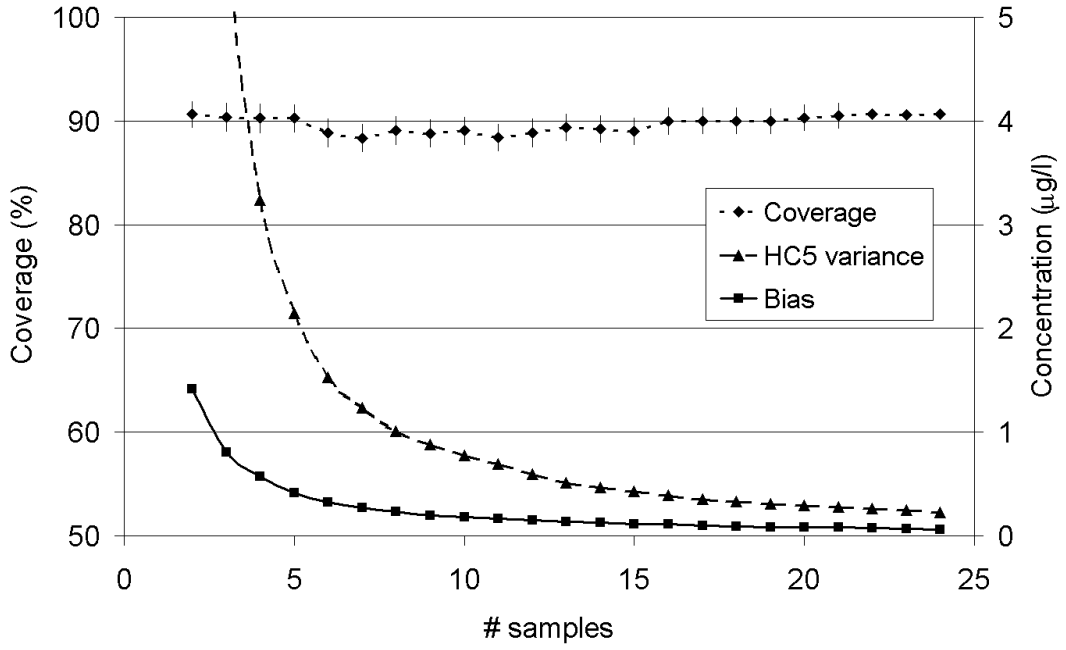


Figure 6: Bias, variance and coverage of the uncertainty intervals of HC_5 for the bootstrapping method (Classical approach, Hazen plotting, 1000 repetitions)

If an underlying distribution cannot be specified, a nonparametric technique should be used. The results of the simulation study for the nonparametric method using Hazen plotting and interpolated EDF are shown in Figure 7. The HC_5 bias and variance are less dependent on the sample size compared to the parametric methods. The bias even decreases when the sample size decreases below sample size 10. This is apparent because it is inconsistent with what is expected from a theoretical point of view i.e. increase of bias at increasing sample size. The coverage varies between 95 and 100%. Hence, the uncertainty intervals are always overestimated, but this is acceptable from a conservative point of view in a risk assessment.

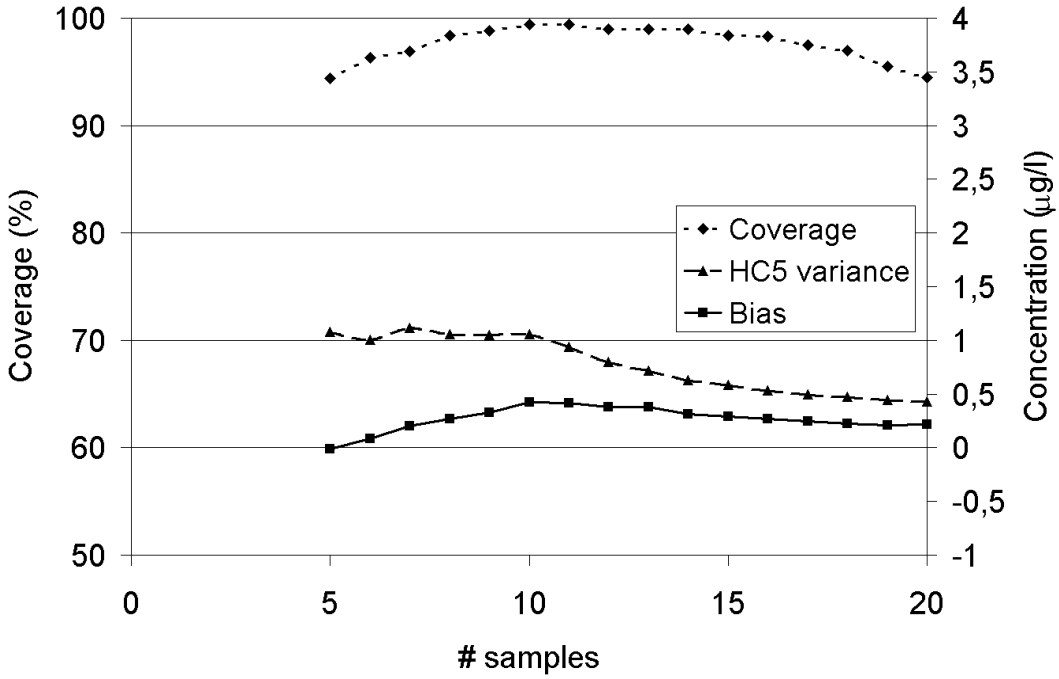


Figure 7: Bias, variance and coverage of the uncertainty intervals of HC_5 for the bootstrapping method (interpolated EDF, Hazen plotting, 1000 repetitions)

Such simulation studies are not often found in the ecotoxicological literature. Instead, some researchers investigate the "stability" or convergence of the HC_5 in function of the sample size. A consistent estimator (e.g. for calculating the 5^{th} percentile) converges to the "true" value as the sample size increases. The result given by the method should not depend on the number of species tested (Roman *et al.*, 1999). Although this is a rather qualitative and subjective assessment, this criterion is also explored, illustrated and compared here for several methods estimating uncertainty and variability. Special attention was given to the difference between parametric and nonparametric methods. Examples of parametric results are shown in Figure 12, Figure 3, Figure 10b, Figure 8 left and Figure 9 left for several hypothetical and real data sets. The results of the nonparametric techniques are shown in Figure 8 right and Figure 9 right for the LAS and Cu data set respectively.

The 5^{th} percentile (or HC_5) should not depend on the sample size. Instead, it should be constant. In case of LAS, parametric and nonparametric methods do not show instable HC_5 estimations (Figure 8). The stability of HC_5 in parametric methods is even independent of the sample size, indicating that even very small sample sizes (of 5 data points) are accurate enough, provided the assumed distribution is right. However, the uncertainty intervals for the nonparametric method do not show the expected "trumpet"-shape, indicating an estimation problem below sample size 10.

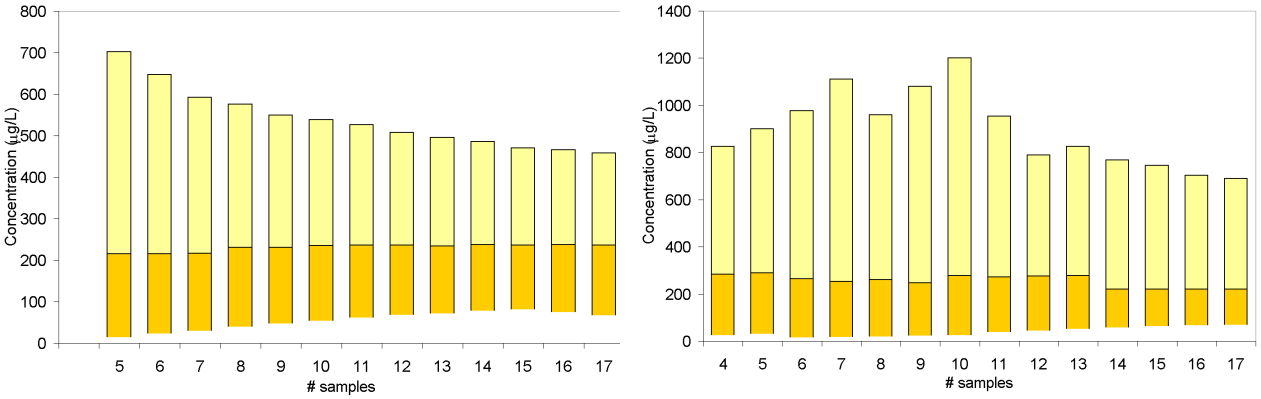


Figure 8: Influence of sample size on the 5th-percentile using (left) parametric method (Aldenberg & Jaworska, 2000) and (right) nonparametric bootstrapping (Interpolated EDF - Hazen plotting) for the LAS data set

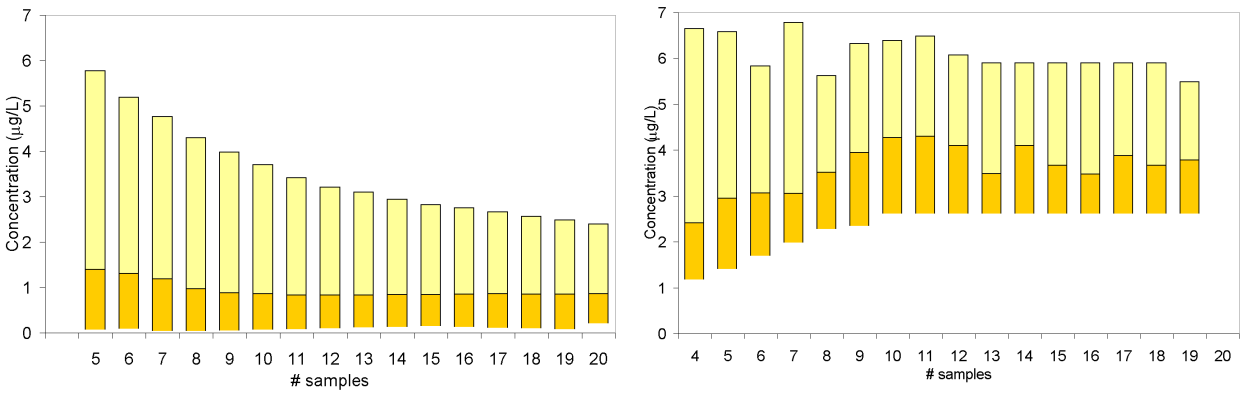


Figure 9: Influence of sample size on the 5th-percentile using (left) parametric method (Aldenberg & Jaworska, 2000) and (right) nonparametric bootstrapping (Interpolated EDF - Hazen plotting) for the Cu data set

In case of Cu, the HC_5-estimates seem to converge above sample size 8 for the parametric methods (Figure 9). Note that the lognormal assumption may be less fulfilled considering its QQ-plot in Figure 2 and the large difference between parametric and nonparametric HC_5s.

For nonparametric methods below sample size 10, the HC_5 gradually drops with decreasing sample size and the confidence intervals are not steadily increasing, as one would expect from a theoretical point of view (less data points, more uncertainty). This number 10 is not a coincidence and can be explained by analytical calculations. Recall the formulae from Chapter 3.2 to calculate a percentile nonparametrically. Based on these, one can calculate the minimum sample size needed to estimate a 5th-percentile (HC_5) without having to extrapolate below the smallest data point. The rank order i for the smallest point is 1, $F(x_i)$ is 0.05 (=5 %) and n is unknown. The calculations are:

mean plotting system: $F_x(x_i) = \frac{i}{n+1} \Leftrightarrow 0.05 = \frac{1}{n+1} \Leftrightarrow n = 19$

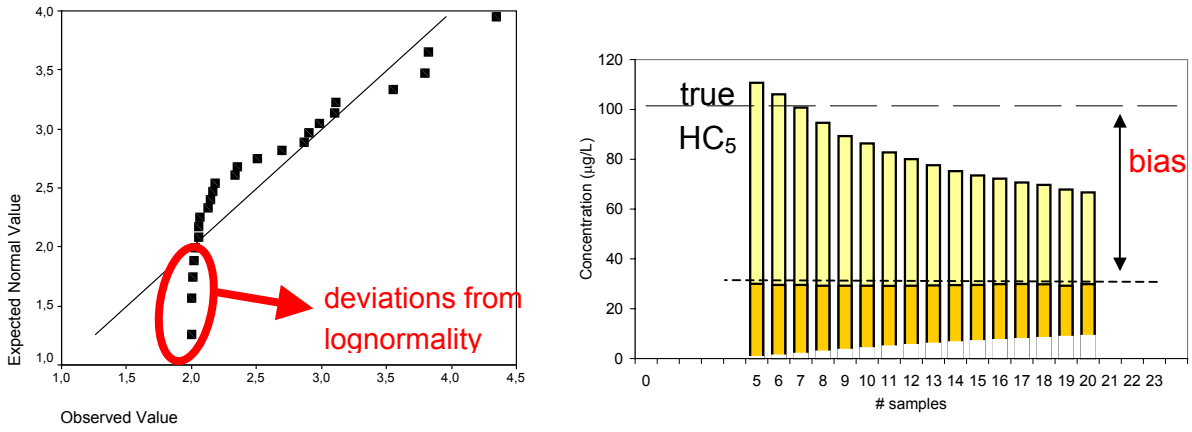
Hazen plotting system: $F_x(x_i) = \frac{i-0.5}{n} \Leftrightarrow 0.05 = \frac{1-0.5}{n} \Leftrightarrow n = 10$

The minimum n=19 for the mean plotting position corresponds to earlier findings in literature (e.g. Van Der Hoeven (2001)). The minimum n=10 for the Hazen plotting position corresponds to the simulation results found here. Similarly, such a minimum sample size can also be calculated for the 20th-percentile (HC_{20}):

Mean plotting system: $F_x(x_i) = \frac{i}{n+1} \Leftrightarrow 0.20 = \frac{1}{n+1} \Leftrightarrow n = 4$

Hazen plotting system: $F_x(x_i) = \frac{i-0.5}{n} \Leftrightarrow 0.20 = \frac{1-0.5}{n} \Leftrightarrow n = 2{,}5 \Rightarrow n = 3$

To further investigate the relevance of the distribution type, twenty-five random sample points were drawn from the third hypothetical target distribution. The QQ-plot for checking lognormality is shown in Figure 10a. In this example, it is clear that the lower data points are deviating from the lognormality assumption and that these data points are not outliers. It is now investigated how such deviations from the lognormal assumption influence the sample size problem.



(a) (b)

Figure 10: Hypothetical "lognormal-like" data set (a) QQ-plot, (b) sample size plot for the parametric method (Aldenberg & Jaworska, 2000)

The result in Figure 10b shows that the usage of the wrong statistical distribution in the first place results in error or bias. The bias is the difference between the estimated HC_5 and the true HC_5. The

thin, black line in Figure 10b represents the true HC_5. The conceptual problem is that this value is only known when a hypothetical distribution is used in practice. The bias, being the difference between the thin line and the middle, black bar, will in practice not be known. It does not affect the sample size determination problem according consideration 2 (level of precision, see below), nor does it affect most other criteria of consideration 1 (e.g. no apparent divergence was observed). Of course, if it is clear from the QQ-plot and the data that they are not lognormal distributed, the lognormal distribution should not be used to determine the minimum sample size.

Based on the results above, it can be summarised that sample size 2 is the absolute minimum for all parametric methods and sample size 10 is the absolute minimum for the nonparametric bootstrap based on Hazen plotting and the interpolated EDF. In Chapter 3.2, it is already shown that the minimum sample size for nonparametric bootstrap based on EDF (resampling) is much larger than 25. Such findings should be considered when the HC_5 estimation method is selected.

3.3.4.2. Level of Precision

Once an appropriate method is selected, the level of precision can be specified and the 5^{th} percentile or HC_5 can be estimated for several sample sizes. The minimum sample size can then be determined. The level of precision can be specified in many ways.

Most papers on sample size determination use this criterion (e.g. Roman *et al.* (1999), Newman *et al.* (2000), Van Der Hoeven (2001) and Manly (1992)). The uncertainty intervals as a function of sample size typically have a "trumpet"-shape: uncertainty intervals become smaller as the number of data points increases because more information becomes available. The optimal sample size is then selected as the sample size where a small increase in sample size only results in a small increase in precision (i.e. a slight decrease in confidence intervals). For example in Figure 12a, 18 could be selected as a minimum sample size. However, these assessments are very qualitative and subjective. As a consequence, different scientists propose their own minimum sample size based on their own level of precision. It is, in our opinion, up to the policy-maker to decide and quantify how large uncertainty intervals may be. For example for Figure 12b, if the policy-maker is already confident with a 90% confidence interval width of 7 µg/l, then a sample size of 19 would be sufficient for this criterion. Another approach is to select the optimal sample size by balancing the cost of obtaining additional data points versus precision statements (as e.g. in Bros & Cowell (1987)).

The level of precision can be specified on an absolute or a relative scale. This is explained by means of the two first hypothetical target distributions. Twenty data points were drawn from each distribution. Both SSDs have different means and variances but the same 5^{th}-percentile (i.e. HC_5)

and are shown in Figure 11. A parametric method was used to construct the uncertainty band and intervals. Chemical (a) only has a narrow range of species sensitivities (i.e. a small variance = 1). All species have more or less the same sensitivity. Chemical (b) on the other hand has a wide range of species sensitivities (i.e. larger variance = 2.05). As can be seen from Figure 11, the larger the variability (i.e. variance of the SSD), the larger the uncertainty or confidence intervals for the HC_5 (for the same sample size).

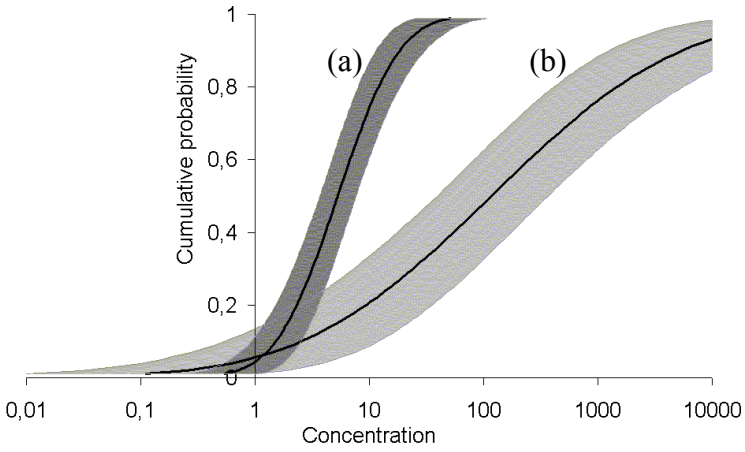


Figure 11: Two SSDs with different means and variance but the same 5th-percentile (uncertainty bands are also shown)

The results for different sample sizes are presented in Figure 12. There is clearly a distinct difference in the width of the confidence intervals. It is clear that more data points are needed for chemical (b) to achieve the same precision (i.e. uncertainty interval width) as chemical (a).

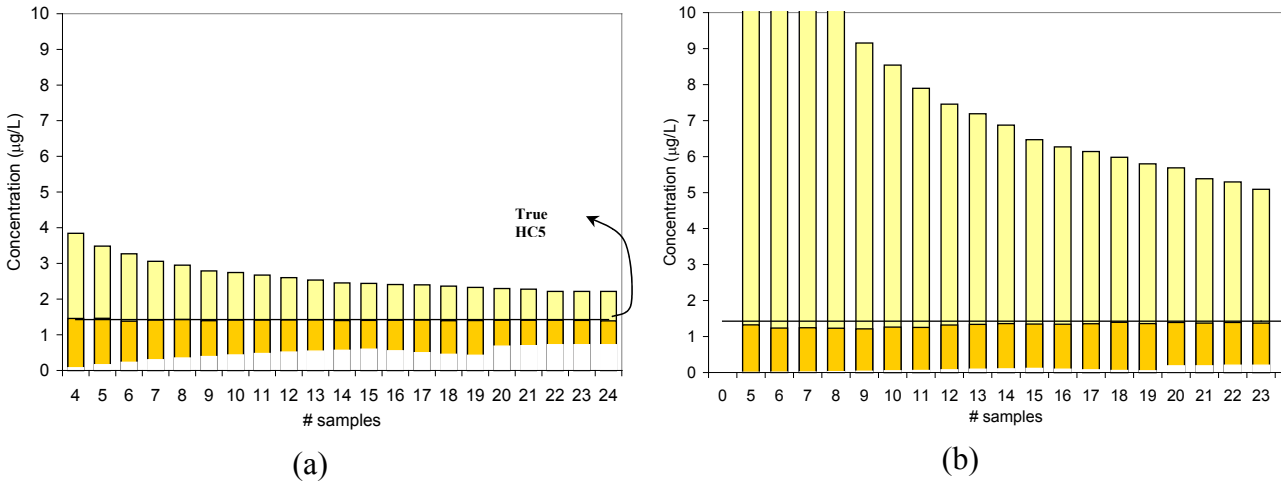


Figure 12: Influence of the level of variability on the sample size - uncertainty interval relation for the 5th-percentile using a parametric method Aldenberg & Jaworska (2000) (for two hypothetical lognormal data sets (a) and (b))

The SSD can be standardised to put the level of precision on a relative scale as illustrated by Roman *et al.* (1999). Standardising is obtained by subtracting the data points by the mean and dividing the result by the standard deviation. On this relative scale, both chemicals (a) and (b) will have the same width of uncertainty intervals because the uncertainty intervals for scale location distributions are independent of the data. They only depend on the sample size.

However, it is suggested here that it is up to the policy-maker to decide whether to put the level of precision on a standardised or a non-standardised SSD. In case the SSD is standardised, the policy-maker feels there is no need to have an equally precise HC_5 estimate of a chemical with a wide range of species sensitivities (as chemical (b) in the example) compared to an HC_5 estimate of a chemical with a small range of species sensitivities (as chemical (a) in the example). Therefore, it is more important to obtain a precise HC_5 estimate because of the narrow range of species sensitivities. In case the SSD is not standardised, the policy-maker feels HC_5 estimates should be precise independent of the wide or small range of species sensitivities.

3.3.4.3. Determination of Minimum Sample Size

The minimum sample size was determined here for the Cu and LAS data set as an illustration. The data sets were considered to be representative of all species sensitivities. For the LAS data set, a parametric method (assuming lognormal distribution) was selected. Given this method selection, it was derived from the reliability assessment above that sample sizes from 2 onwards can be considered. For the Cu data set, a nonparametric method (with interpolated EDF and Hazen plotting) was selected since its QQ-plot indicates potential deviations from lognormality and since essential metals can be characterised by a threshold (see Chapter 3.2). Given this method selection, it was derived from the reliability assessment above that sample sizes from 10 onwards can be considered.

For consideration 2, the policy-maker should specify the level of precision (on absolute or relative scale). In this example, the policy-maker decided to specify the level of precision as the ratio of HC_5 and its lower 90% confidence limit to be smaller than 5 on the absolute scale, i.e. the policy-maker feels HC_5 estimates should be precise independent of the wide or small range of species sensitivities. Ecological considerations also needs to be specified but are not considered here. Based on this level of precision, the minimum sample size for the LAS and Cu data set is respectively 10 and 14.

3.3.5. Conclusions & Further Research

First, a review was made of existing minimum or optimal sample size derivation procedures. No theoretical background was found to support the approach of Newman *et al.* (2000). Further, simulation examples show that when the resample size exceeds the sample size logical and statistical inconsistencies arise. Indeed, one cannot get more information from a data set than the data itself contains. After this analysis it is felt that Newman's *et al.* (2000) recommendations on sample size should not be adopted.

Next, this chapter has illustrated two important considerations needed for sample size determination. The general methodology is first to determine the scientific reliability and accuracy of the HC_5 estimation method, representativeness of the data set and second to specify the desired level of precision needed, by the policy-maker or risk manager.

Based on simulation studies to assess the reliability and accuracy of the statistical methods for small sample sizes, it can be summarised that sample size 2 is the absolute minimum for all parametric methods and sample size 10 is the absolute minimum for the nonparametric bootstrap based on Hazen plotting and the interpolated EDF. In Chapter 3.2, it is already shown that the minimum sample size for nonparametric bootstrap based on EDF (resampling) is much larger than 25. The findings should be considered when the HC_5 estimation method is selected. Once an appropriate method is selected, the level of precision can be specified on an absolute or relative scale by the decision-maker and the minimum sample size can then be determined.

Chapter 3.4

-

Hierarchical Uncertainty and Variability Estimation

Chapter 3.4

Hierarchical Uncertainty and Variability Estimation

The Exposure/Environmental Concentration (EC) and the Species Sensitivity (SS) are characterised by uncertainty and variability. The uncertainty and variability of the EC and SS can be separated in different hierarchical levels. For example, the hierarchical variability of the EC of a chemical in European rivers can be separated in a spatial variability level (different monitoring locations throughout Europe) and a temporal variability level (different time measurements per monitoring location). Govaerts *et al.* (2001) and Lecoutre (2001) discussed several hierarchical approaches for the EC estimation. One of their main issues was to deal with censoring. But their case study dealt with large data sets (range 1000-25000). For the SS, unfortunately, only small data sets are available (range 10-100). The main focus of this chapter will therefore be on SS.

3.4.1. Introduction

Several organisms have different sensitivities towards a certain chemical. Species sensitivities are based on ecotoxicity data to assess effects on ecosystems in ecological risk assessment and to derive environmental quality standards. After Species Sensitivity Distributions (SSDs) were introduced by Van Straalen & Denneman (1989), the method has been extensively debated. The most common current approach is to derive Predicted No Effect Concentration (PNEC) from the 5^{th} percentile of the SSD (EU-TGD, 1995). Historically, that value is known as the Hazardous Concentration at p-protection level, HC_p (in this case p = 5).

In the derivation process of this HC_5, a lot of information (sources of variability and uncertainty) is currently discarded because summary statistics (e.g. averages) are taken in each step of the HC_5 derivation process. Currently, in the conventional method, only inter-species variability is modelled in the SSD characterisation. Sometimes, sampling uncertainty of selecting species from a

community is modelled as in e.g. Aldenberg & Slob (1993) or Aldenberg & Jaworska (2000) and Chapter 3.2. In this way, other information like the variability of intra-species sensitivity (i.e. differences in sensitivity between several individuals of the same species towards a toxicant) and inter-laboratory variability (i.e. differences due to laboratory test populations and/or test conditions) is discarded.

Consider for example in Figure 1 the two cumulative probability distributions. These could either represent inter-laboratory variability, intra-species or inter-species sensitivity. Both curves have the same 20th percentile but different means and variances. The location and shape of the distribution (which contain information on the sensitivities and variations) are not accounted for if only the 20th percentile is taken for further analysis.

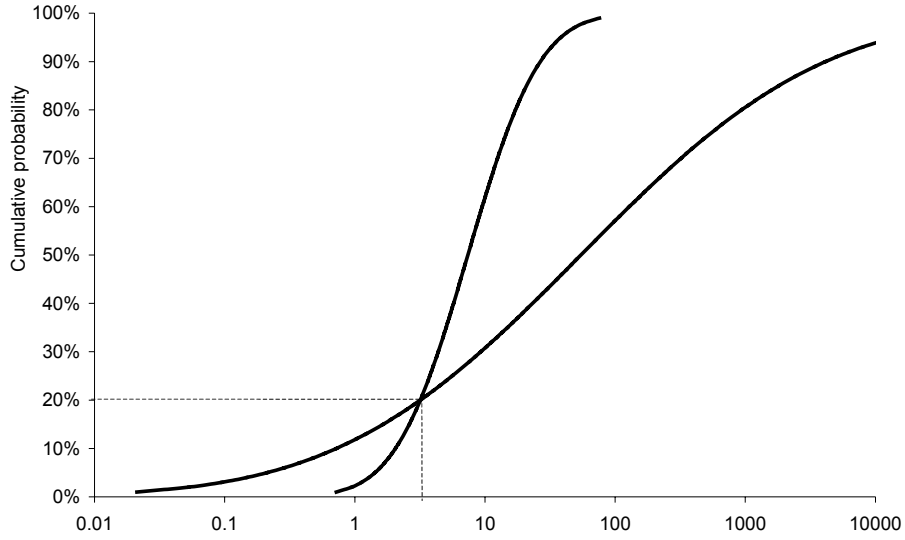


Figure 1: Two cumulative distribution functions with different mean and variation but the same 20th percentile

There are many statistical methods that can account for this hierarchical variability and uncertainty on the contrary to what Wheeler *et al.* (2002) claims. It is however still unclear in the chemical risk assessment field how to account for this hierarchical structure because of its case-specific nature and whether it is worthwhile to account for all these extra hierarchical levels of information.

The goal of this chapter is twofold: (1) understand how to account for hierarchical variability and uncertainty in a SSD and (2) evaluate which methods should be used. To address the first question, the conventional method and two (alternative) methodologies will be evaluated. To address the second question, all methods will be compared with each other. Simulation studies as well as case studies of SSDs for some organic and inorganic chemicals will be given.

Note that next to these extra forms of hierarchical uncertainty and variability, there are also other sources of uncertainty (e.g. lab to field extrapolation, mixture toxicity…). These are not dealt with here nor are other (dis-)advantages of the use of SSDs. Rather, this chapter should be viewed as a step forward in revealing, quantifying and propagating more sources of uncertainty and variability in a scientifically defendable manner. The more sources are quantified; the better the decision-maker can assess the reliability of a risk assessment outcome.

3.4.2. Description of Hierarchical Variability and Uncertainty

The left panel of Figure 2 shows a community of n species. Each species i was investigated by m_i laboratories and each laboratory j performed a toxicity test for l_{ij} individuals. An overview of all hierarchical variability and uncertainty levels in the HC_5 derivation process is given in Figure 2 and Table 1 (partly based on Forbes *et al.* (2001), Moore *et al.* (2000), Janssen *et al.* (2000), Hart (2001)).

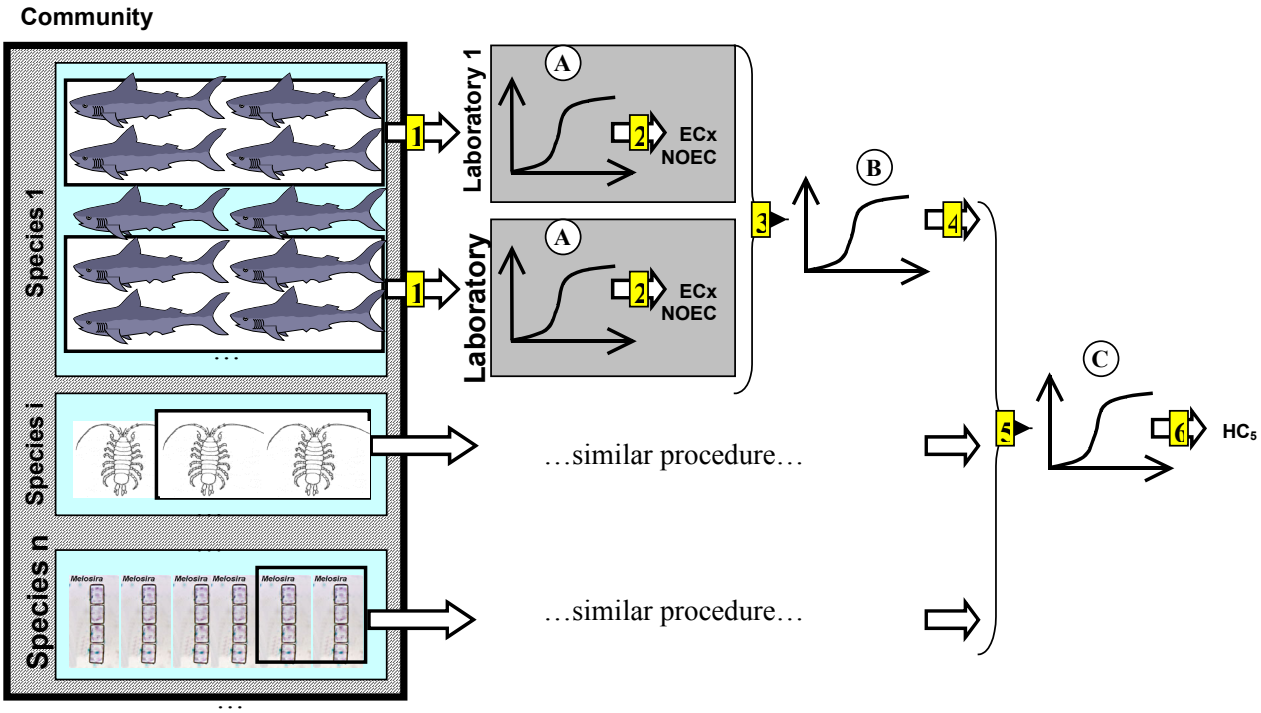


Figure 2: Procedure for deriving a SSD where information is lost due to (1) sampling uncertainty per individual, (2) taking the x^{th}-percentile for EC (e.g. 50^{th}-percentile for EC_{50}) or calculating a significant difference for NOEC, (3) sampling error per lab, (4) averaging, (5) sampling error per species, (6) taking the 5^{th}-percentile for HC_5. The distributions represent (A) intra-species variability (EC-curve), (B) inter-laboratory variability (C) SSD (variation due to inter-species variability)

Table 1: Overview of several types of (hierarchical) uncertainty and variability

N	Type	Description	How discarded?
1	Sampling uncertainty*	Due to selecting individuals from a particular species	Not considered
2	Intra-species variability	The selected individuals of the same population typically have different sensitivities towards a chemical. The results are shown in an EC-curve (effect-concentration or dose-response curves as e.g. A in Figure 2). Smit *et al.* (2001) concluded on the basis of the evaluation of more than 300 EC-curves, that the intra-species variability is considerable.	Either the x^{th}-percentile of the EC-curve is taken, resulting in an Effective Concentration EC_x (e.g. EC_{50} for the median) or a No Observed Effect Concentration (NOEC) is calculated based on a significant difference with the lowest observed test concentration.
3	Sampling uncertainty*	Due to selection of several references (laboratories) for a particular species in a literature search of toxicity tests	Not considered
4	Inter-laboratory variability (and uncertainty)	Here, inter-laboratory variations are mainly contributed by variability although by several sources**: • **variability** in different testing methods in different labs (test-type, test-component (e.g. $ZnSO_4$, ZnO, $ZnCl_2$ …), experimental time, test-water (artificial, river, lake, … resulting in different physico-chemical properties as pH, hardness, DOC, suspended solids…)); • **variability** in endpoint criteria (growth, maturation, reproduction, survival…); • **variability** due to acclimatisation (a species can adapt to higher or lower concentrations); • **variability** due to different life-stages of the species (e.g. larvae, adults…); • **uncertainty** due to measurement error… Inter- (and intra-)laboratory variations can be quite high despite the use of standardised toxicity test protocols and species (Moore *et al.*, 2000).	The mean (Janssen *et al.*, 2000), (Stephan *et al.*, 1985) (Wheeler *et al.*, 2002) or the lowest value (Forbes *et al.*, 2001) (Wheeler *et al.*, 2002) of all NOECs or EC_xs observed for several laboratories of each species is taken. Some sources of variability are accounted for by standardising the toxicity endpoints to equal physico-chemical test conditions.

Table 1 (continued): Overview of several types of (hierarchical) uncertainty and variability

N	Type	Description	How discarded?
5	Sampling uncertainty*	Due to species selection in a community (i.e. in theory, a random, representative sample of species in the environment is taken, in practice this often boils down to the species toxicity tests taken from literature).	Sometimes simulated as in e.g. Aldenberg & Slob (1993). Usually, the resulting confidence interval is not considered for further analysis.
6	Inter-species variability	The averages of all available NOECs or EC_xs of each species are combined into a SSD representing the different sensitivities the different species have towards a chemical.	The SSD is often determined, but usually the 5^{th} percentile (HC_5) of the SSD is taken.

* Sampling uncertainty is due to the random process of sample selection. One sample (e.g. a set of species sensitivities) is only one possible realisation of a random process. If this analysis would be repeated independently from the previous one, other samples might be taken and different toxicity results would be obtained.

** Inter-laboratory variations were considered to be mainly due to either variability because in our opinion, collecting more results from more laboratories will hardly result in smaller inter-laboratory variations. And this reflects better the definition of variability than the definition of uncertainty.

All levels of variability (inter-laboratory variability, intra- and inter-species variability) could be aggregated into one single SSD reflecting all inherent natural heterogeneities. All sources of sampling uncertainty could be included in an aggregated uncertainty band around the SSD.

3.4.3. Methods

Next to the conventional, two alternative methods are proposed and discussed. An overview of the properties of each method is given in Table 2.

The conventional method is based on summary statistics provided by several laboratories, which in their turn are based on ecotoxicity summary statistics of several individuals. Consequently, only the inter-species variability is determined (step 6 in Table 1 and Figure 2). Sometimes, sampling uncertainty of selecting species from a community is modelled (step 5 in Table 1 and Figure 2). The other sources of hierarchical variability are not considered in the effects analysis. The first (alternative) method considers the hierarchical structure of the data in a hierarchical method. The second method ignores the hierarchical structure and considers all raw data points on the same level. Both are based on all (raw) data and are therefore designed to include the hierarchical

variability and uncertainty. Recently, the non-hierarchical method (but without weighting) was also studied by Wheeler *et al.* (2002).

Table 2: Overview of the properties of all methods discussed

Method	Type of data	Data weighted (see 3.4.3.1)?	Types of variability modelled?*	Types of sampling uncertainty modelled?*	Practical use?
Conventional	Summary data	Not relevant	6	5	Easy
Hierarchical	All (raw) hierarchical data	Yes	2,4,6	1,3,5	More effort needed
Non-hierarchical	All (raw) data	Yes	2,4,6	1,3,5 but no interpretation	Easy

* Number refers to N in Table 1 and Figure 2

Weighting the ecotoxicity data can be done for several reasons and correspondingly according to several methods. Weighting according to the reasons outlined in 3.4.3.1 was only relevant for the two alternative methods.

For the "practical use"-property, putting summary or all raw data into one distribution is easier and can be easily done in spreadsheet software compared to applying a less easy method like a hierarchical method.

3.4.3.1. Assigning Weights to the Data

Weighting the ecotoxicity data can be done for several reasons and correspondingly according to several methods. First, some species may play a more dominant or suppressive role in an ecosystem and should for this reason be assigned a larger or smaller weight. Second, one of the important assumptions in the SSD methodology is that the selected species should be representative for an ecological community (which is often mentioned in literature e.g. Forbes *et al.* (2001), Fuchsman *et al.* (1999)). Weights could resolve the overrepresentativeness of e.g. a taxonomic group in the SSD. These two reasons are not considered any further because they are ecologically inspired. Third, a lot of literature data can be available for one particular species (e.g. because the species is well studied or easy to cultivate), leading to this species being overrepresented in the SSD. This reason is not

relevant for the conventional approach because averages are taken for each species and consequently, each species is equally represented in the SSD. But when considering all raw data, some species will be overrepresented if a lot of laboratory data for that species are available.

Weighting the data is a possible solution but unfortunately not often applied (as e.g. in Wheeler *et al.* (2002)). Weighting would not complicate the interpretation of the SSD since each species would be equally represented. One possible option for assigning weights is presented here: Weights are assigned to each data point such that all data for one particular species only counts as one species despite for example its abundant lab data. In the general case, the weights are calculated as

$$w_{ijk} = \frac{1}{l_{ij} \cdot m_i \cdot n},$$

where

w_{ijk} is the weight for individual k for laboratory j and for species i where $\sum w_{ijk} = 1$

l_{ij} is the number of individuals for laboratory j and for species i

m_i is the number of laboratories for species i

n is the number of species

In this way, every particular laboratory and species (which was over- or underrepresented in the raw data set) will now be equally treated compared to the other laboratories or species.

In case intra-species variability is not considered (k = 0), a simplified formula $w_{ij} = \frac{1}{m_i \cdot n}$ is used.

This formula is more applicable in practice, as data-availability on intra-species sensitivity is often problematic. Also in the case studies presented below, intra-species variability is not considered and therefore, it was also excluded from the simulation study.

3.4.3.2. The Conventional Method for SSD Building

Sampling error or uncertainty can be modelled in many ways. Here, the bootstrap technique was selected. Given a sample of size n, the general approach in bootstrap simulation is to assume $\hat{F}$, to perform r replications (e.g. r = 5000) of the original data set by randomly drawing, with replacement, n values from $\hat{F}$, and then calculate r values of the statistic of interest. These can be used to determine the uncertainty distribution of the statistic of interest. Note that the properties of the bootstrap are asymptotic. One approach is to use the actual data set itself and to randomly select, with replacement, the actual values of the data set. This is sometimes referred to as **resampling**. The data can be represented via an Empirical Distribution Function (EDF) $\hat{F}$. Another approach to

achieve the same is to first calculate random samples of a uniform distribution between 0 and 1. The random samples are then found as the inverse cumulative distribution function of these random uniform samples. This can be noted as $\hat{F}^{-1}(U(0,1))$. Basically, resampling is done from that step function. $\hat{F}$ is assumed to be a good estimate of the real, unknown distribution *F*. But, $\hat{F}$ can be any distribution type (nonparametric or parametric (e.g. lognormal, triangular,…) distribution) and determines the type of bootstrapping. From $\hat{F}$ the statistic of interest can be calculated (here the 5th percentile). The reader is referred to Chapter 3.2 for more information on the bootstrap technique.

3.4.3.3. Alternative Method 1: Hierarchical Method ("Weighted Hierarchical Bootstrap")

Hierarchical methods can be separated into several classes. Parametric methods assume a certain underlying distribution or model (e.g. a random effects model (Warren-Hicks *et al.*, 2000)) whereas a nonparametric method relies on the data themselves. The parametric methods are most often used in literature. There are frequentist and Bayesian viewpoints of how to deal with hierarchical data sets (Warren-Hicks *et al.*, 2002). Bayesians like to add prior knowledge in the analysis and have a different interpretation of uncertainty. There are numerical, simulation and analytical techniques. These different classes of methods have their advantages and disadvantages. These different methods will not be discussed here. The main focus is to discuss the difference between a hierarchical and a non-hierarchical method.

Here, a hierarchical bootstrap technique was selected. The bootstrap is easy to understand and implement (see Chapter 3.2). In a bootstrap, the data (or a specified distribution $\hat{F}$) are resampled as if the analysis was completely repeated. In the hierarchical bootstrap, this principle remains the same (see Figure 3). The entire data generation process of Figure 2 is repeated several times.

First a species is randomly selected. Second, for that species, a laboratory is randomly selected. And third, for the selected species and laboratory, an individual is randomly selected. In particular, the species are resampled *n* times. For each selected species, the laboratories are resampled m_i times. For each selected species and laboratory, this resampling of individuals is repeated l_{ij} times. Basically, these are three embedded loops. In each resample, there are l_{ij} x m_i x *n* number of data points on which a nonparametric and parametric distribution is fitted and the HC_5 is calculated. Repeating this resampling procedure a large number of times (e.g. 500 times) results in an integrated variability distribution with an uncertainty band aggregating all sampling uncertainties. The weights of each randomly selected species, laboratory or individual are accounted for as described above. A similar procedure was applied by Lecoutre (2001).

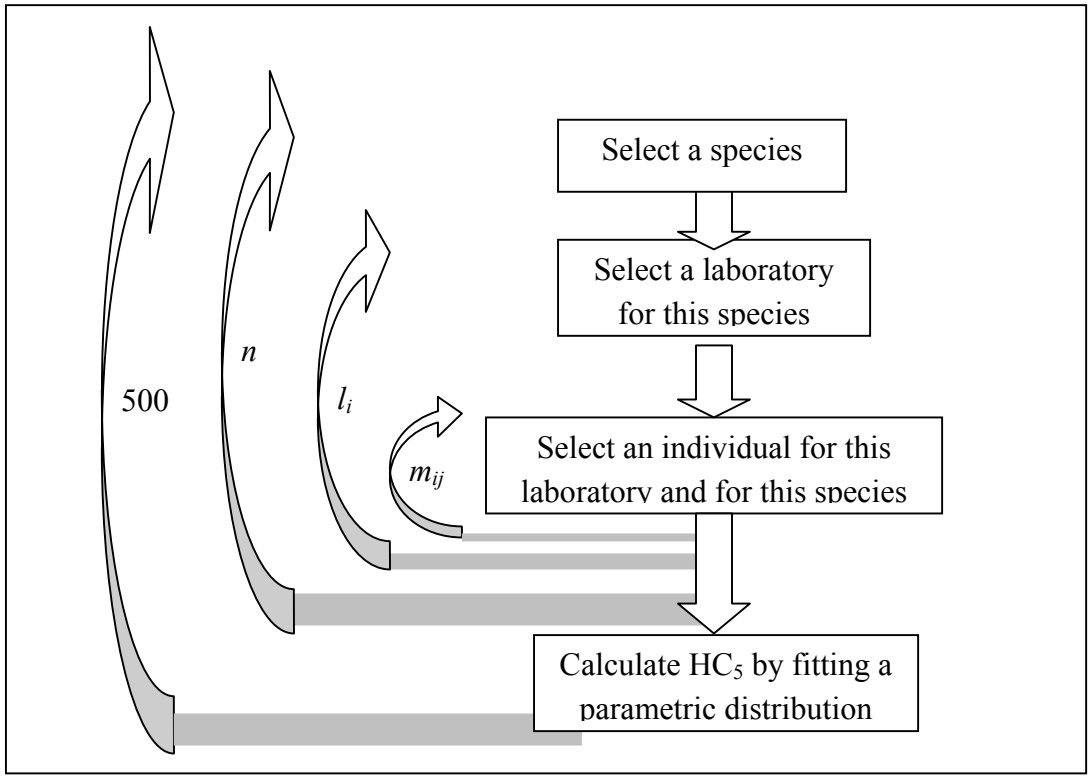


Figure 3: Principle of hierarchical bootstrap

There is however a substantial practical drawback to this method. Often, the data sets are too small to apply nonparametric HC_5 calculation. For example, the number of laboratories per species is mostly very small (range of 1-5). Therefore, only parametric HC_5 calculations were performed.

3.4.3.4. Alternative Method 2: Non-Hierarchical Method (Data Also Weighted)

The bootstrap was again selected for the estimation of uncertainty and variability, but it had to be extended in order to incorporate weighting. In the next paragraph, it is explained how the estimated target distribution $\hat{F}$ was modified for the nonparametric and parametric bootstrap.

In the nonparametric bootstrap, the weighted interpolated Empirical Distribution Function (EDF) is constructed by plotting the sorted data $(x_1 \le x_2 \le ... \le x_n)$ against the cumulative probability calculated as $F(x_i) = \frac{i - 0.5}{n}$ (here i represents the rank of the data point). Figure 4 illustrates how the EDF changes when the data points are weighted. The values of the data points (the X-values) remain the same but their cumulative probability changes according to their assigned weights

because the cumulative probabilities are now calculated as $F(x_i) = \sum_{i=1}^{i} w_i$ (cumulated weights after ranking). In the bootstrap technique, samples are now taken from the weighted interpolated EDF instead of the interpolated EDF.

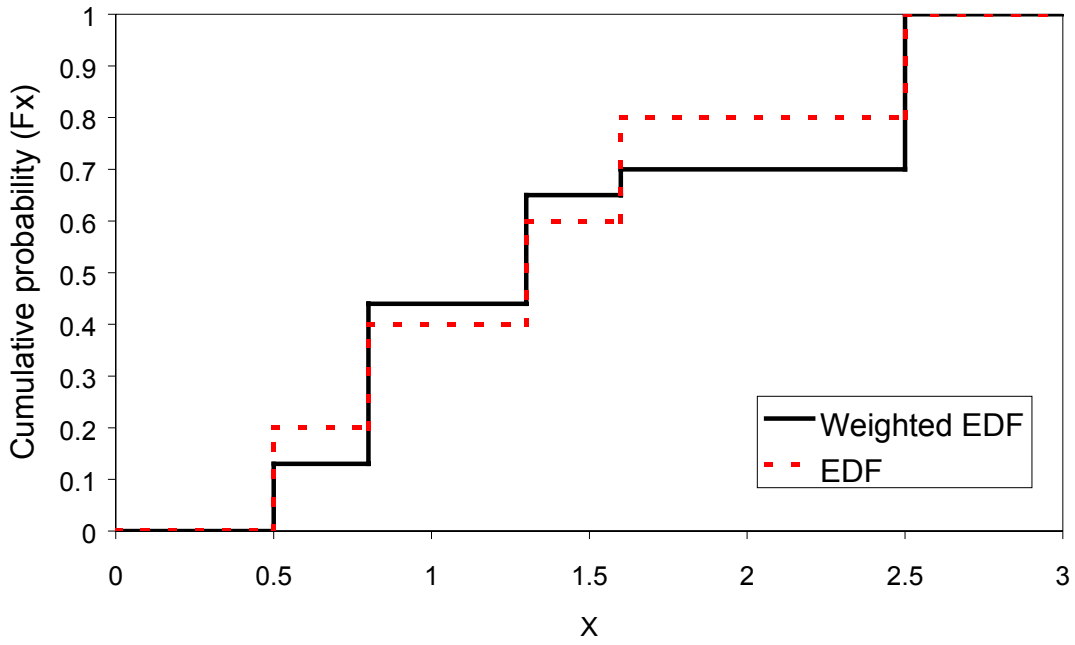


Figure 4: Hypothetical example of an empirical distribution function (EDF) and a weighted empirical distribution function (weighted EDF)

For the weighted parametric bootstrap, the weighted mean and variance need to be calculated. They are calculated using the following formulas (NIST, 2001):

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad s_w{}^2 = \frac{\sum_{i=1}^{n} w_i \left(x_i - \bar{x}_w\right)^2}{\frac{(n-1)\cdot \sum_{i=1}^{n} w_i}{n}}$$

where x_i the i^{th} observation

w_i the weight of the i^{th} observation

n the number of observations

$\bar{x}_w$ the weighted mean of the observations

$s_w{}^2$ the weighted variance of the observations

3.4.4. Simulation Study

3.4.4.1. Description

The distinction between a target population (represented by distribution *F*) and an estimated target population (represented by distribution $\hat{F}$) should be considered carefully as discussed in Chapter 3.2. In statistics, an estimate is unbiased if the expected value of a statistic as the mean or 5th percentile of the estimated target distribution $\hat{F}$ is equal to the true value of the parameter. The bias is therefore a measure for the accuracy of the HC_5 estimation of each method. The variation of the HC_5 and is a measure for the precision of the HC_5 estimation. Similarly, a method has accurate coverage if the probability *p* that a confidence interval does not cover the true parameter is equal to the probability level used to construct the confidence interval. In a simulation study, these coverage and bias can be found by repetitively taking random samples of the target distribution *F*. Furthermore, a good method has a small bias and a small variation of the HC_5. In this simulation study, it was assumed that a community consisted of 30 species. Each species was tested in a number of laboratories. An overview of the species and their inter-laboratory variation (lognormal distributions were assumed) can be found in Table 3.

Table 3: Assumptions for the simulation study (Mean and standard variation of the lognormal distribution representing inter-laboratory variability (true HC5 = 4.77))

Species	Mean*	Standard deviation*	Number of laboratories	Species	Mean*	Standard deviation*	Number of laboratories
1	2.08	0.1	2	16	2.5	0.7	10
2	2.94	0.74	4	17	2.66	0.23	4
3	2.2	0.69	4	18	4.34	0.42	1
4	2.66	0.73	10	19	2.35	0.3	4
5	4.01	0.42	5	20	5.17	0.4	5
6	2.65	0.6	1	21	2.2	0.47	1
7	5.67	0.5	1	22	2.76	1.53	1
8	2.9	0.27	3	23	4.27	0.3	3
9	2.46	1.53	3	24	2.58	1.23	1
10	4.27	0.5	1	25	2.32	0.34	2
11	2.88	1.13	4	26	3.9	0.35	3
12	2.82	0.24	3	27	4.17	0.4	1
13	3.8	0.2	1	28	2.78	1.13	3
14	2	0.2	3	29	2.22	0.14	3
15	3	0.7	4	30	3	0.2	1

* Mean and standard deviation of the corresponding normal distribution

Based on this hypothetical, but realistic data set, the true HC_5 can be determined (also in Table 3). In the simulation study, samples are taken from that true, target distribution *F*. In this study, 5 and 10 species were randomly selected. Next, a number of samples were taken from each species inter-laboratory variability distribution (as in practice many laboratories test the same species). Based on this set of data, the HC_5 was estimated according to all three methods discussed above. Repeating this process many times allows to estimate the bias, the HC_5 variation and the coverage for every method. Here, the process was repeated only 100 times (due to computational limitations) but this number appeared to be sufficient. The results of the simulation study are summarised in Table 4.

Table 4: Comparison of the 'conventional' method based on summary statistics versus the alternative methods (hierarchical or non-hierarchical) based on raw data (number of trials = 100)

		Summary data		Raw/all data			
		Conventional Method		Hierarchical Method		Non-hierarchical Method	
	Sample size	Parametric	Non-parametric	Parametric	Non-parametric	Para-metric	Non-para-metric
Mean bias	5	0.67	- 0.47	0.39	Unreliable at such small sample sizes	- 0.17	- 0.02
	10	- 0.12	2.36	- 0.54		- 0.81	- 0.18
Std. Dev. HC_5	5	3.23	2.12	2.41		2.58	2.98
	10	1.91	2.15	1.36		1.44	2.00
90%- coverage	5	89%	87%	91%		80%	85%
	10	92%	100%	92%		80%	82%

3.4.4.2. Comparison of the Methods

It is important to clearly differentiate the statistical and environmental interpretation of each of the methods. In the alternative methods, the SSD does not only stand for inter-species sensitivity but also for inter-laboratory variability (and if data obtained with different test results obtained for the same species within the same laboratory were also included, also intra-species sensitivity). In the hierarchical method, the uncertainty band represents an aggregated/integrated sampling error of selecting individuals of a species, selecting laboratory test results from literature and selecting species from a community (1, 3 & 5 in Figure 2). This sampling uncertainty is larger since more uncertainties are considered and quantified. For the non-hierarchical method, treating all forms of variability on the same level results in a smaller uncertainty band around the SSD, but this band has no real interpretation, mainly because the hierarchy of the data set was ignored.

The smaller the bias, the more accurate the HC_5 estimation is. A negative bias is preferred over a positive one for conservative reasons (it is better to overestimate the effect). Based on these two criteria, the (nonparametric) non-hierarchical method seems to be the best method (especially for small sample sizes). However, only small differences were found between all considered methods.
The smaller the variation of the HC_5, the more precise the HC_5 estimation is. Based on this criterion, the alternative methods perform better than the conventional method.
The closer the coverage is to 90%, the better the confidence interval is estimated. The coverage is close to 90% for the hierarchical method. The non-hierarchical method underestimates the sampling uncertainty but this was expected, as the non-hierarchical method does not simulate the hierarchical structure of the SS. It basically has no interpretation. It is peculiar, however, that the conventional method can have a good coverage. There appears to be some kind of compensating mechanism that makes its coverage accurate. On one hand, smaller data sets (as is the case with the summary data in the conventional method) result in larger confidence intervals. On the other hand, ignoring some types of uncertainty results in smaller confidence intervals (only sampling uncertainty at the top level is considered). Both opposing driving forces seem to compensate each other.

3.4.4.3. Parametric versus Nonparametric

Since the comparison between parametric and non-parametric methods was not part of this study, the advantages and disadvantages will only be described in short. There are no real differences nor preferences found for parametric or non-parametric methods based on bias, the variation of the bias or the sample size. The coverage is larger for the non-parametric methods, as is commonly experienced in literature (see Chapter 3.2). Nonparametric methods cannot be applied for very small sample sizes (smaller than 10, see Chapter 3.3).

3.4.5. Case Study

3.4.5.1. Description of the Data Sets

For the case study in this paper, NOECs of 4 chemicals for different species analysed in different laboratories have been investigated, namely Cu, Zn, LAS (Linear Alkylbenzene Sulphonate) and atrazine. As previously mentioned, only inter-laboratory variability and inter-species variability, but no intra-species variability will be assessed. The data sets consisted of compiled databases from literature (based on Versteeg *et al.* (1999) and Janssen (2001)). The data sets are only used for

illustrative purposes. No discussion is made on data quality as this hardly influences the evaluation of the methodology. More information on the number of species and laboratories can be found in Table 5.

Table 5: Overview of the data sets of the case study (# is the number of)

	# species with abundant data (7-11 labs)	# species with few data (2-6 labs)	# species analysed in one lab	Total # species
LAS	2	1	15	18
Zn*	3	3	12	18
Cu	1	8	8	17
Atrazine	0	4	13	17

* Two very sensitive species

The LAS data set is characterised by the fact that a lot of lab data (individuals from different laboratories) are available for two species namely *Daphnia* and *Pimephales.* Most other organisms only have one toxicity value. The Zn data set is characterised by two very sensitive species: *Ephydatia fluviatilis* and *Epeorus latifolium*. However, both are represented by only one lab data point. As a result, they will be assigned a large weight in the alternative methods. The Cu data set is characterised by the fact that there are many species and the individuals are equally distributed over all species. The atrazine data set is characterised by 4 species for which more than one lab data point is available: *Selenastrum* (4), *Lemna* (3), *Chlorella* (6) and *Chlamydomonas* (4). For all other species only one lab data point is available.

3.4.5.2. Results and Discussion

The results of the practical case study are summarised in Table 6 and shown for LAS in Figure 5. In the first row of this figure, results of nonparametric bootstrapping (with interpolated EDF) are presented. In the second row, the parametric bootstrapping results (with assumed lognormal distribution) are shown. The first column presents the conventional method, the second column alternative method 1 and the third column alternative method 2.

Table 6: HC5's and its lower 90% confidence limit (in μg/l) for all methods discussed and for all chemicals (3000 - 10000 shots)

		Summary data (17-18)		Raw/all data (30-55)			
		Conventional Method		Hierarchical Method		Non-hierarchical Method	
		Parametric	Non-parametric	Parametric	Non-parametric	Parametric	Non-parametric
Cu	Lognormal?°	No				Yes	
	HC_5	5.4	8.2	4.7		4.5	4.9
	Lower 90% CI	3.1	3.4	3.0		3.0	1.9
Zn	Lognormal?°	No				No	
	HC_5	6.0	4.1	5.8		5.4	2.7
	Lower 90% CI	2.4	1.3	2.3		3.2	1.0
LAS	Lognormal?°	Yes				Yes	
	HC_5	239	203	243		232	153
	Lower 90% CI	106	67	122		134	46
Atra-zine	Lognormal?°	No				Yes	
	HC_5	0.77	0.99	0.93		0.84	0.20
	Lower 90% CI	0.20	0.10	0.23		0.34	0.15

Unreliable at such small sample sizes

° According to Kolmogorov-Smirnov of the 5% level of significance

The case study shows that in case of chemicals like LAS and Zn with a number of species with abundant lab data and the rest almost no lab data, the conventional method is close to the hierarchical method and therefore can be considered sufficiently accurate (see Table 4 and Table 6), i.e. there is no need for a hierarchical method. In case there are no species with abundant lab data, i.e. the case of Cu and atrazine, the non-hierarchical method is closer to the hierarchical method in comparison with the conventional method (see Table 4 and Table 6).

Weighting induces a change in the shape of the SSD. For the non-parametric methods, the uncertainty band becomes smaller at some places and broadens at other places depending on the weights of the lab data points (see Figure 5). When all lab data are equally distributed over all species, weighting becomes less important (as in the Cu data set), but when some species are represented by a lot of individuals, weighting has a large influence on the results and therefore becomes important (as in the LAS data set). The individual sensitivity distribution curves of the two abundant species in the LAS data set, namely *Daphnia* (range 210-4900 mg/l) and *Pimephales* (range 350-2500 mg/l), cannot be recognised in the shape of the SSD, since the individual range of each of these species is largely overlapping with the sensitivities of the other species.

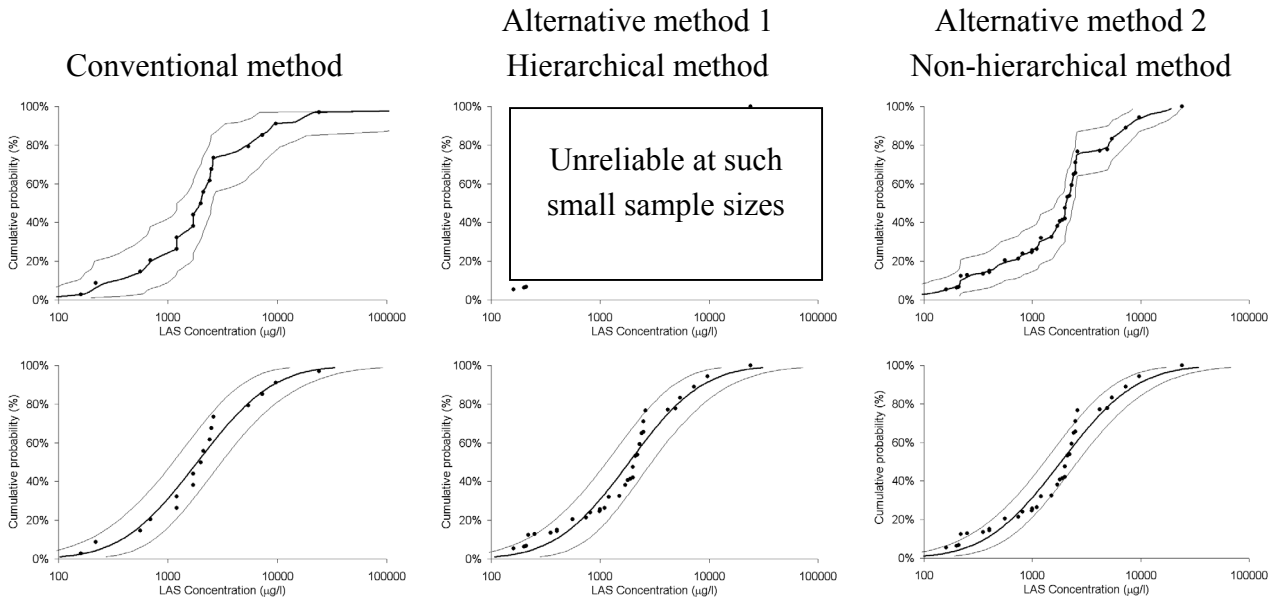


Figure 5: LAS-SSDs with 90%-uncertainty band for a nonparametric (first row) and parametric bootstrap (second row), and for the conventional method (1st column), alternative method 1 (2nd column) and alternative method 2 (3rd column)

3.4.6. Conclusions

Several (statistical) methods were proposed to account for hierarchical variability (i.e. inter-laboratory variability, inter- and intra-species sensitivity) and hierarchical uncertainty (mainly sampling uncertainty). There are three conclusions for the simulated data sets. First, of all the studied methods, the hierarchical method was found to be the most accurate and precise method and the only method with a scientifically reliable interpretation. Second, all methods were found to produce similar results. This indicates that the conventional method does not perform as bad as one would expect based on the fact that it is ignoring underlying information. Third, the non-hierarchical method seems to be most conservative for the simulations performed here, but should not be used for confidence interval estimation. Further testing and research on (non-)hierarchical methods is however needed to generalise these conclusions.

The resulting SSD can then be used entirely (instead of taking the 5th percentile) for the probabilistic risk characterisation (see Chapter 4.1).

Part 4

-

Probabilistic Risk Characterisation

Chapter 4.1

-

Probabilistic Risk Quotient Method

Large parts of this chapter were published in:

Verdonck F. A. M., Jaworska J., Janssen C. R. & Vanrolleghem P. A. 2002. Methodologies to determine risk of chemicals in rivers under data uncertainty. *Proceedings IWA 3rd World Water Congress*. Melbourne, Australia. 7-12 April 2002.

Verdonck F. A. M., Jaworska J., Janssen C. R. & Vanrolleghem P. A. 2002. Probabilistic ecological risk assessment for chemical substances. *Proceedings iEMSs 2002, Integrated Assessment and Decision Support* 1, 144-149. Lugano, Switzerland. 24-27 June 2002.

Chapter 4.1

Probabilistic Risk Quotient Method

Once the Exposure Concentration Distribution (ECD) and the Species Sensitivity Distribution (SSD) are determined and characterised by variability and uncertainty (see Part 3), all elements are available for the actual risk characterisation (see Figure 6 in Part 2). Conventionally, the deterministic risk characterisation is a yes/no statement (see Part 2).

Consider Figure 1 where three hypothetical ECDs are shown with the same 95^{th} percentile (considered as PEC (Predicted Exposure Concentration) in the conventional risk assessment) and three hypothetical SSDs with the same 5^{th} percentile (considered as PNEC (Predicted No Effect Concentration) in the conventional risk assessment). These different combinations of ECDs and SSDs lead to the same deterministic risk quotient PEC/PNEC of 10. These different environmental situations are considered under risk of adverse effects although some combinations will be worse than others. Clearly, other risk characterisations are needed that can differentiate such environmental situations.

In this chapter, risk will be characterised as a probability accompanied with an uncertainty or confidence interval. In this way, both the variability and uncertainty of the EC (Exposure Concentration) and SS (Species Sensitivity) will be accounted for. The chapter both contains methodological developments and practical case studies.

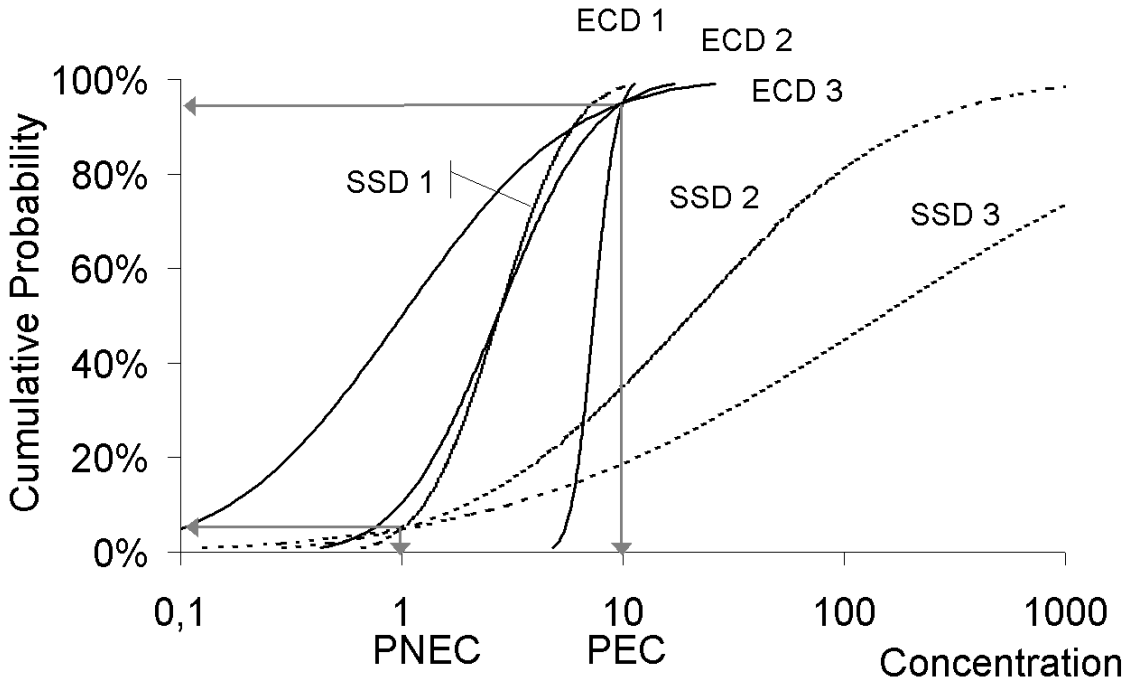


Figure 1: 3 ECDs (Exposure Concentration Distributions) with same PEC as 95th percentile (Predicted Exposure Concentration) and 3 SSDs (Species Sensitivity Distributions) with same PNEC as 5th percentile (Predicted No Effect Concentration)

4.1.1. Introduction

The calculation of a probabilistic risk can be done in many ways (see Part 2). Both the overlap between the EC and SS probability density functions or between the respective cumulative distribution functions have been suggested as a measure of risk (cf. Solomon *et al.* (2000)). However, such graphical measures of risk are mathematically not correct. How to specifically calculate this overlap can be implemented in various ways, as well.

Aldenberg *et al.* (2002) compared different methods mathematically and concluded that the discrete summation for the expected risk of Cardwell *et al.* (1999), Van Straalen's ecological risk (Van Straalen, 1990), the numerical integration of risk distribution curves in the WERF methodology (Solomon & Takacs, 2002) (Warren-Hicks *et al.*, 2002), as well as the area under the curve (AUC) of Joint Probability Curves (JPCs) are all numerically equal to, and may be interpreted as, *the risk of some log EC to exceed some log SS*, as originally implemented by the probability of failure in reliability engineering. The graphical interpretation of this risk is the Area Under the Curve (AUC) of the product of the ECD cumulative distribution with the SSD probability density function, or

alternatively, the AUC of the product of the ECD probability density function with the SSD cumulative distribution function. The reader is referred to Figures 5.10 and 5.11 in Aldenberg *et al.* (2002).

The two most common methods used are the AUC of Joint Probability Curves (JPCs) of Solomon & Takacs (2002) and the risk lookup tables proposed by Aldenberg *et al.* (2002). More information on these methods can be found in Part 2.
JPCs come in two forms: either as a graph of ECD exceedance against fraction of species affected (cumulative probabilities of SS), or as a graph of fraction of species affected against cumulative probabilities of EC. The first is called an Exceedance Profile Plot (EPP) (Giesy *et al.*, 1999), and involves plotting one minus the cumulative probability of the ECD against the cumulative probability of the SSD for any given concentration. The second JPC curve results from plotting the cumulative probability of the SSD on the ordinate against the cumulative probability of the ECD on the abscissa for any given concentration. The latter JPC plots are called Cumulative Profile Plots (CPP) (Aldenberg *et al.*, 2002). CPP JPCs are somewhat easier to draw and interpret, since they only involve cumulative distribution functions. However, each represents the same risk curve, only visualised in a different way. Both EPP and CPP are shown in this chapter. An example of an EPP and CPP is given in Figure 8. The AUC of either JPC can also be considered as a measure of risk. Mathematically, it can be shown that JPC AUCs are equal to the AUCs in the overlap plots of the ECD with the SSD. Hence, the AUC of a JPC expresses the same risk of a random EC to exceed a random SS (Aldenberg *et al.*, 2002).
Aldenberg *et al.*, (2002) tabulates probabilistic risk for two independent normal distributions (table 5.3 on p. 73). By scaling the ECD to the SSD, a two-parameter dependent risk is obtained, by only varying the mean and the standard deviation of the ECD (log ECD) relative to the SSD (log SSD) (see also Part 2).

Not all of the above methods or the other risk characterisation methods described in the literature study in Part 2 are capable of handling all types of parametric and nonparametric EC or SS distributions. The probabilistic risk lookup table of Aldenberg *et al.*, (2002) for example is, unfortunately, only for (log)normal distributions and can only be extended for parametric distributions. Not all methods are easy to use and interpret at the same time. The JPC and AUC methodology, developed in ECOFRAM (1999) for example, is unfortunately, although relatively easy to construct and calculate, sometimes difficult to understand and interpret by decision-makers and risk managers. This was experienced at a SETAC Pellston workshop on the application of uncertainty analysis to ecological risks of pesticides (from 24 February till 1 March 2002, Pensacola, Florida, USA). Almost all literature sources investigated in Part 2, do not calculate an uncertainty or confidence interval on their risk estimate although many acknowledge the need to distinguish between uncertainty and variability. As a consequence, little attention is given to the visualisation of the risk and its uncertainty interval. A good risk visualisation would especially be useful when geo-referenced probabilistic risks need to be displayed on a map because displaying

more than one value for each location in a Geographical Information System (GIS) is often cluttered.

The goals of this chapter are three-fold. First, the goal is to mathematically describe a general, easy-to-use risk characterisation tool that can handle all types of parametric and nonparametric distributions based on the conventional deterministic risk quotient (for easy understanding by risk managers and decision-makers as they are familiar with the risk quotient concept). The second goal is to calculate an uncertainty or confidence interval for this risk. And the third goal is to investigate how a probabilistic risk and its uncertainty interval can be visualised and communicated. Two case studies will be described to illustrate and further discuss the applicability of the methods developed. The results are discussed in the subsequent sections.

4.1.2. Probabilistic Risk Quotient Method

The estimation of the probabilistic risk and its uncertainty or confidence interval according to the risk quotient method is done in two steps. First, the probabilistic risk is determined. In this way, only variability is considered. Second, its uncertainty interval is estimated.

4.1.2.1. When Only Variability is Considered: Probabilistic Risk

The probability of some randomly selected EC exceeding some randomly selected SS has been demonstrated to be a common measure of risk (Aldenberg *et al.*, 2002). This can be written as:

Equation 1: $\text{Risk} = P(EC > SS)$ where P( ) denotes "the probability of"

As described above, several probabilistic risk calculation methods are available. Here, it will be shown that the probabilistic risk fits well into the paradigm of the deterministic quotient method broadly used in chemical management (EU-TGD, 1995). The risk quotient (RQ) is an index of risk calculated by dividing an exposure estimate (EC) by a toxicity value (SS). Its properties have been well described in literature (Burmaster & Bloomfield, 1996) (Rai *et al.*, 1996) (Campbell *et al.*, 2000) and in Part 2.2. The ecological quotient estimates are used to define risks to selected species representing an ecosystem. A critical value of the risk quotient may form the basis for regulatory action, including possible collection of more information or performing a more refined analysis (Warren-Hicks & Moore, 1995).

In a probabilistic framework, however, EC and SS are regarded as random variables having probability distributions rather than point estimates. As a result, the quotient will also have a

probability distribution. The probability of EC exceeding SS (this probability can be considered as a measure of risk of adverse effects) is equal to the probability that the quotient EC/SS becomes larger than one or that $\log_{10}$(EC/SS) becomes larger than zero since:

Equation 2:

$$\begin{aligned} \text{Risk} &= \mathrm{P}(\mathrm{EC} > \mathrm{SS}) = \mathrm{P}\left(\frac{\mathrm{EC}}{\mathrm{SS}} > 1\right) \\ &= \mathrm{P}\left(\log_{10}\left(\frac{\mathrm{EC}}{\mathrm{SS}}\right) > 0\right) = \mathrm{P}(\log_{10}(\mathrm{EC}) - \log_{10}(\mathrm{SS}) > 0) \end{aligned}$$

The random variables EC and SS in Equation 2 can be described by a parametric or nonparametric distribution. If EC or SS is described by a nonparametric distribution, the risk has to be calculated numerically or by means of simulation, e.g. a Monte Carlo analysis. In case EC and SS are described by parametric distributions (e.g. lognormal for EC and Pareto for SS), the risk can be calculated numerically or sometimes also analytically. Simulations will be demonstrated in the first case study.

Logarithmic transformations were made in Equation 2 because the risk can be easily calculated analytically when lognormal distributions are assumed for the ECD and the SSD. The result of a quotient of two lognormal distributions (EC and SS) is again a lognormal distribution. But it is much easier to work with the difference of two normal distributions ($\log_{10}$(EC) and $\log_{10}$(SS)) because its parameters can easily be calculated. The difference of two independent normal distributions is also a normal distribution with parameters (based on Burmaster & Bloomfield (1996), see also Figure 2):

Equation 3: $\mu_{\log(\mathrm{EC/SS})} = \mu_{\log(\mathrm{EC})\text{-}\log(\mathrm{SS})} = \mu_{\log(\mathrm{EC})} - \mu_{\log(\mathrm{SS})}$

Equation 4: $\sigma_{\log(\mathrm{EC/SS})} = \sigma_{\log(\mathrm{EC})\text{-}\log(\mathrm{SS})} = \sqrt{\sigma^2_{\log(\mathrm{EC})} + \sigma^2_{\log(\mathrm{SS})}}$

with μ and σ respectively the mean and standard deviation of the $\log_{10}$-transformed data

Note that $\mu_{\log(\mathrm{EC/SS})}$ is not the risk. It is the mean of the log(RQ) distribution (see Figure 2). Rather, the risk of some randomly selected EC exceeding some randomly selected SS (see Equation 2) is given by the probability of $\log_{10}$(EC/SS) exceeding 0. This is equal to one minus the cumulative probability of the above log(RQ) distribution for $\log_{10}$(EC/SS) = 0 or EC/SS = 1 (see Equation 2 and also shown graphically in Figure 2). This calculated risk is equal to the AUC of a JPC.

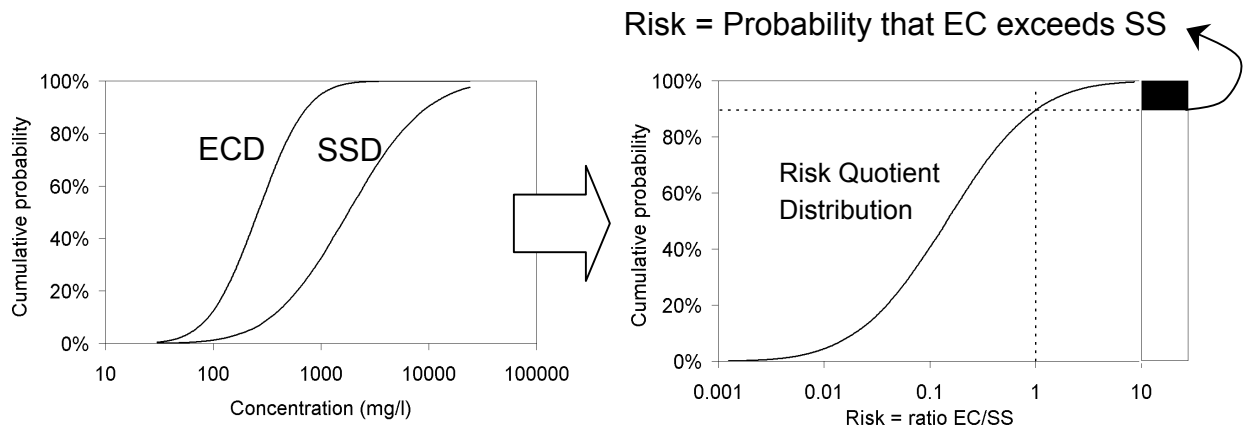


Figure 2: Calculation of the Probabilistic Risk Quotient distribution and the risk (right panel) based on the Exposure Concentration Distribution (ECD) and the Species Sensitivity Distribution (SSD) (left panel on log scale), The risk is found as one minus the cumulative probability for the ratio = 1 (here: 9%)

The formula for the probabilistic risk in the case of two normal distributions (because log-transformed) is described by:

Equation 5: $$P(\log EC - \log SS > 0) = 1 - \Phi_{\mu_{\log EC}-\mu_{\log SS},\sqrt{\sigma^2_{\log EC}+\sigma^2_{\log SS}}}(0) = \Phi_{0,1}\left(\frac{\mu_{\log EC}-\mu_{\log SS}}{\sqrt{\sigma^2_{\log EC}+\sigma^2_{\log SS}}}\right)$$

where $\Phi_{m,s}(x)$ is the cumulative normal distribution of x with mean m and standard deviation s

This is a consequence of Equation 3 and Equation 4 given earlier. In the second panel of Figure 2, the cumulative log risk quotient distribution $\Phi(\log RQ)$ is shown. The exceedance (reverse cumulative) log risk quotient distribution is $1-\Phi(\log RQ)$.

Two comments should be made at this point. First, an important condition for using these formulae is that the EC and SS are independent variables. This is generally considered to be the case. Second, in order to assess the quotient of EC and SS, both sets of values have to be compatible (Aldenberg *et al.*, 2002). One should not compare 96h toxicity tests endpoints with hourly fluctuating concentrations at a discharge point (see Chapter 5.2). The resulting probabilistic risk cannot be interpreted. Instead, either one hour toxicity tests should be used or e.g. weekly fluctuating concentrations should be used. The time interval of EC measurements or simulation results should be equal to (or larger than) the time interval of SS toxicity testing.

4.1.2.2. When Variability and Uncertainty is Considered: Probabilistic Risk and its Uncertainty Interval

In the previous section, only the variability of the ECD and SSD was considered. This resulted in a probabilistic risk quotient (variability) distribution. The ECD and SSD are also uncertain because of sampling error. Adding a Monte Carlo sampling loop to the risk calculation can capture this uncertainty. In each run, an ECD and SSD will be selected from their respective uncertainty bands and the risk quotient distribution will be calculated. After many runs, the risk quotient distribution will also have an uncertainty band and consequently, the probabilistic risk will be accompanied with an uncertainty interval (see Figure 3 for lognormal ECD and SSD). More details can be found in Chapter 3.1.

Note that other sources of uncertainty than sampling uncertainty, such as SS lab to field extrapolation uncertainties, the representativeness of the species in an SSD, model uncertainty and others, are not dealt with here but the generality of the methodology should make it relatively easy to include these as well in the future.

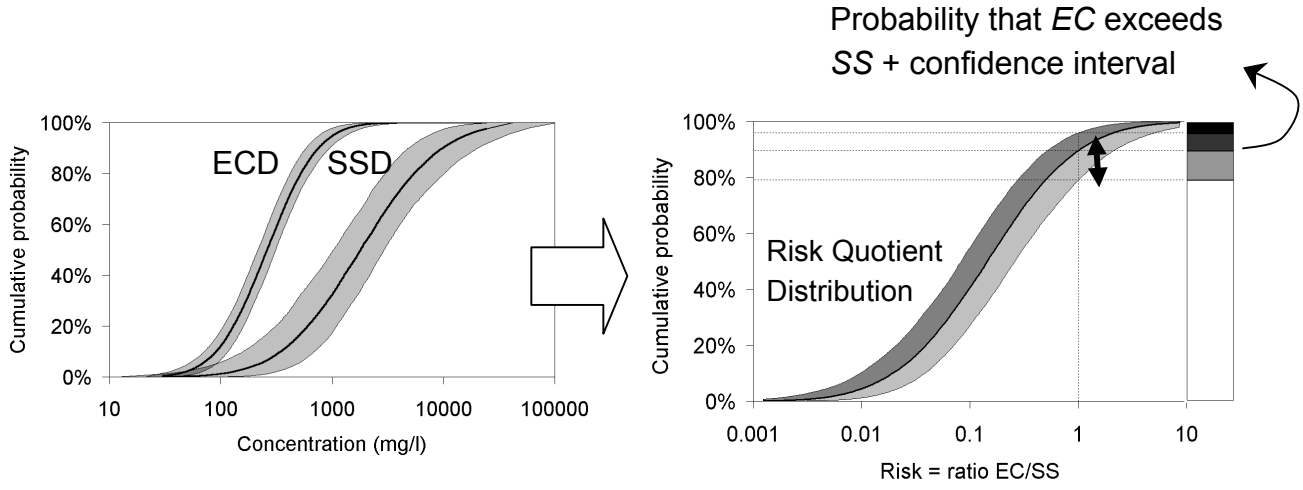


Figure 3: Calculation of the probabilistic risk and its uncertainty interval based on a lognormal ECD and a lognormal SSD

When data are collected sequentially, there is often a tendency for those taken close together (in time or space) to be more alike than those taken farther apart. Hourly ECs may show great variation over a long period of time, while ECs 1 hour apart are very similar. This tendency to be alike is called serial dependence or autocorrelation. The distance between the observations that are examined for correlation is called the lag. One method to detect serial dependence is the autocorrelation function. This is a plot of correlation coefficients for different lags (Berthouex & Brown, 1994). If the ECs are autocorrelated, than the assumption of independent observations in confidence interval theory is violated and the resulting uncertainty estimates may be underestimated

(in case of positive autocorrelation). Further research should be undertaken to account for the autocorrelation in the uncertainty and variability estimation (e.g. block resampling (Davison & Hinkley, 1997)).

A proposal will be made on how to visualise the probabilistic risk and its uncertainty or confidence interval. The probabilistic risk visualised as a column chart in the right panel in Figure 3 can also be visualised as a pie chart as in Figure 4. The entire pie represents 100%. The grey shades indicate how large the probabilistic risk is with a pre-defined certainty. The larger the white slice, the lower the risk is. The more black, the larger the risk is. The larger the grey slices are, the more uncertainty there is on the risk. The example shows that the median risk is 23% (50% certainty) and there is 95% certainty that the risk is smaller than 45%. The 90% uncertainty interval is then 9-45%. This Figure 4 is one possible (new) way of visualising probabilistic risk and its uncertainty interval. The possibility of plotting many of these pie charts on geographical maps is potentially useful and will be further explored in Chapter 5.1.

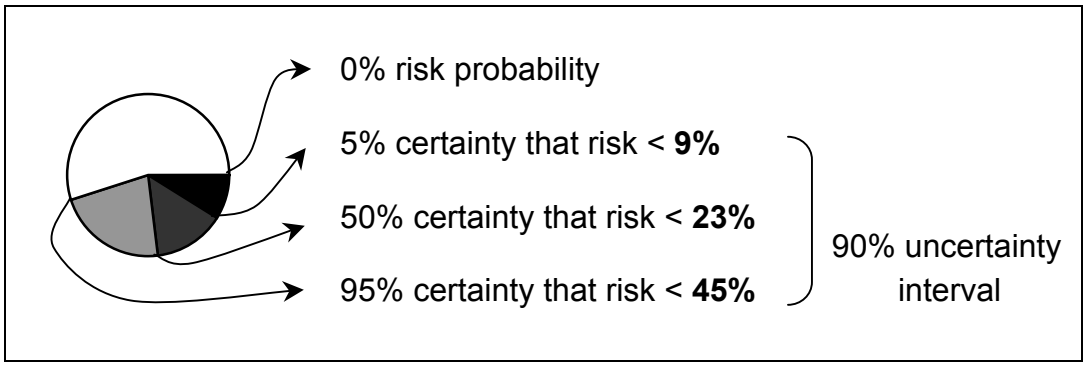


Figure 4: Visualisation of the risk of 23% and its 90%-uncertainty interval

4.1.3. Case Studies

Two case studies were accomplished. The first deals with the probabilistic risk of a heavy metal, namely Zn, in The Netherlands and was part of a collaboration with EURAS. The main focus was on the risk quotient method and its potential. The second case study deals with the probabilistic risk of the pesticide atrazine in two monitoring locations in Flanders, Belgium. The main focus here was on the uncertainty intervals and their visualisation.

4.1.3.1. Probabilistic risk of Zn in The Netherlands

In the framework of the European new and existing chemicals policy, an overall risk assessment for zinc (Zn) is being prepared. Traditionally, risk assessments in the European framework are performed according to the methodology laid down in the Technical Guidance Document (EU-TGD, 1995). The potential risks are typically estimated in a deterministic way using point estimates for both exposure (EC) and effects (SS). Currently, overly conservative assumptions are used in an attempt to account for the uncertainty (see also conventional deterministic risk assessment in Part 2.2).

The goal of this case study for Zn in The Netherlands was two-fold: first, to illustrate how a probabilistic risk characterisation is more realistic compared to the conventional deterministic risk assessment and second, the risk quotient method is illustrated.

Long term ecotoxicity data on Zn for the aquatic organisms belonging to different trophic levels were collected from the EU Zn risk assessment report and can be found in Van Sprang *et al.* (2002). The ecotoxicity data were corrected for background Zn concentrations in the ecotoxicity tests. Two distributions were fitted to the data: the lognormal distribution (a non-threshold distribution, i.e. a distribution going through zero) and the Pareto distribution (a threshold distribution, i.e. a distribution not going through zero). Note that the Pareto distribution was fitted on the log-transformed data (but for ease of use, the Pareto distribution name will further be used). These SSDs with uncertainty bands can be found on the right side of Figure 5. Clearly, there is a large difference between the two distributions at the lower tail. In Chapter 3.2, it was found that the Pareto distribution was the best fitting distribution. The question which distribution type is the best is not addressed here again. Instead, focus is made on the influence of different (good and bad fitting) distribution types on the resulting probabilistic risk. More information on fitting distributions to the Zn data can be found in Chapter 3.2 and Van Sprang *et al.* (2002).

The Dutch (exposure) monitoring data originate from RIZA and CIW databases. Total Zn concentrations were compiled for the year 1998 and were considered as representative for the whole Dutch region. A lognormal distribution was fitted to the data (see left side of Figure 5). Further details can be found in Van Sprang *et al.* (2002). The serial spatial autocorrelation function could not be determined because there was not sufficient geographical information available. Some monitoring stations indicated a (positive) temporal autocorrelation. To conclude, one can expect that the estimated uncertainty band on the ECD (see left side of Figure 5) is most probably underestimated.

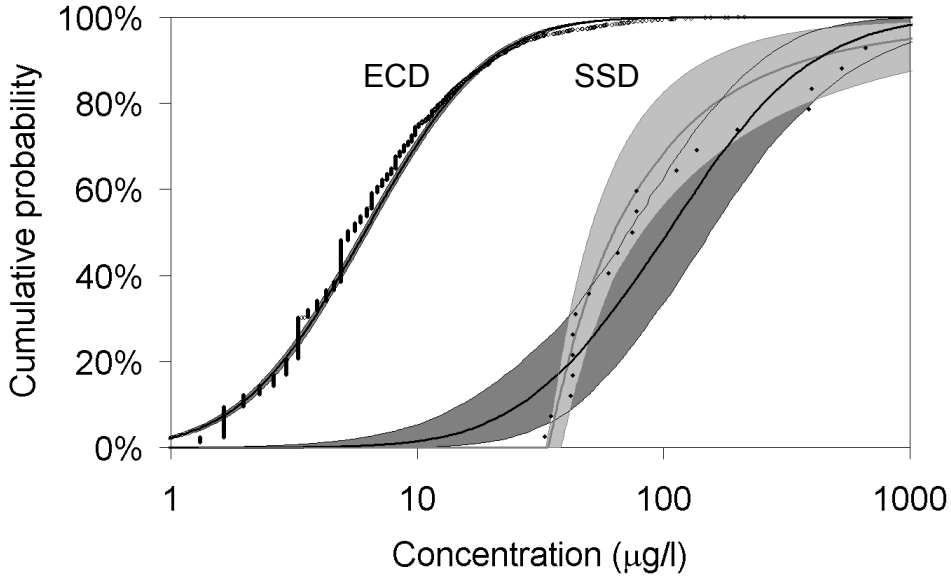


Figure 5: ECD (Exposure Concentration Distribution) of Zn in the Netherlands (left) and lognormal (dark grey) and Pareto distributed (light grey) SSD (Species Sensitivity Distribution) (right)

The conventional deterministic approach revealed a potential risk associated with Zn for the Dutch surface waters. A risk quotient PEC/PNEC (ratio of 90th percentile of ECD and 5th percentile of SSD) of 1.1 and 0.6 is obtained respectively for the lognormal and Pareto SSD. A probabilistic risk characterisation, on the other hand, revealed small probabilistic risks, *defined as the probability that a randomly selected EC exceeds a randomly selected SS*, of 2.2% and 0.5% for the lognormal and Pareto SSD respectively. Their risk quotient distributions are shown in Figure 6. The probabilistic risk characterisation considers the quantitative information of the full range of the ECD and SSD instead of only considering the upper tail of the ECD and the lower tail of the SSD. Therefore, the probabilistic risk characterisation is more realistic and refined.

In addition, the uncertainty on this probabilistic risk was estimated. There is 95% certainty that the probabilistic risk is smaller than 5.3% and 0.9% for respectively the lognormal and Pareto distributed SS. Or alternatively stated, 0.7 - 5.3% and 0.1 - 0.9% are 90% confidence intervals of the probabilistic risks for respectively the lognormal and Pareto distributed SS. The uncertainty interval for the risk based on Pareto distributed SS is clearly smaller. This is also reflected in the uncertainty band (see Figure 6) and is due to the small uncertainty in the lower tail of the Pareto distribution. Note again that both uncertainty intervals may be underestimated because the uncertainty on the ECD can be underestimated due to serial dependence of the EC data.

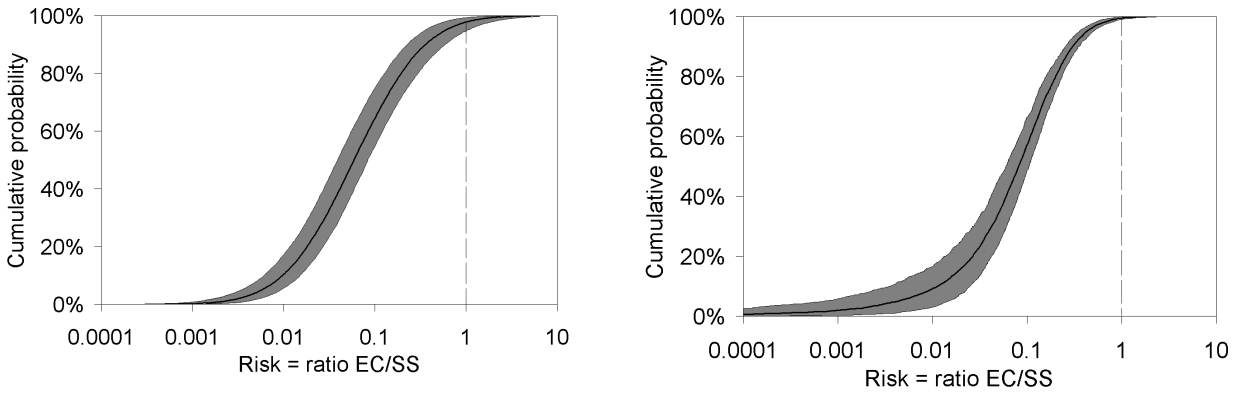


Figure 6: Zn risk quotient distributions based on lognormal ECD (Exposure Concentration Distribution) and lognormal SSD (Species Sensitivity Distribution) (left) or Pareto distributed SSD (right)

4.1.3.2. Probabilistic risk of atrazine at two monitoring locations in Belgium

In this case study, probabilistic risks and their 90%-uncertainty intervals were predicted for the pesticide atrazine in two monitoring locations in the river catchments of Flanders in Belgium. Since atrazine is such a widely used herbicide and persistent compound, it is considered a great potential for groundwater and surface water contamination. Therefore, it is frequently detected (Lipishan & Lee, 1996).
The focus of this case study is two-fold. First, attention is paid to the uncertainty intervals of the resulting probabilistic risks. Second, several possible visualisations of probabilistic risk and its uncertainty interval are tested and discussed.

The data set for the atrazine SSD consists of chronic ecotoxicity values (NOECs: No Observed Effect Concentrations) and can be found in Versteeg *et al.* (1999). A lognormal distribution was assumed and fitted to the data. The fit was satisfactory. More information on fitting distributions to data can be found in Chapter 3.2.

The exposure concentrations (EC) were obtained from the Flemish environmental agency (VMM, 2001). Atrazine was (mostly monthly) measured from 1991 till 2000. Only the reliable data from the years 1997 till 2000 were considered.

Two monitoring stations in the "Westsluisbeek" in Alveringem (VMM nr. 914012) and in the "kanaal van Gent naar Oostende" in Aalter (VMM nr.777000) were selected for a probabilistic risk characterisation. Their ECD and SSD (lognormal fit and uncertainty band) is shown in Figure 7.

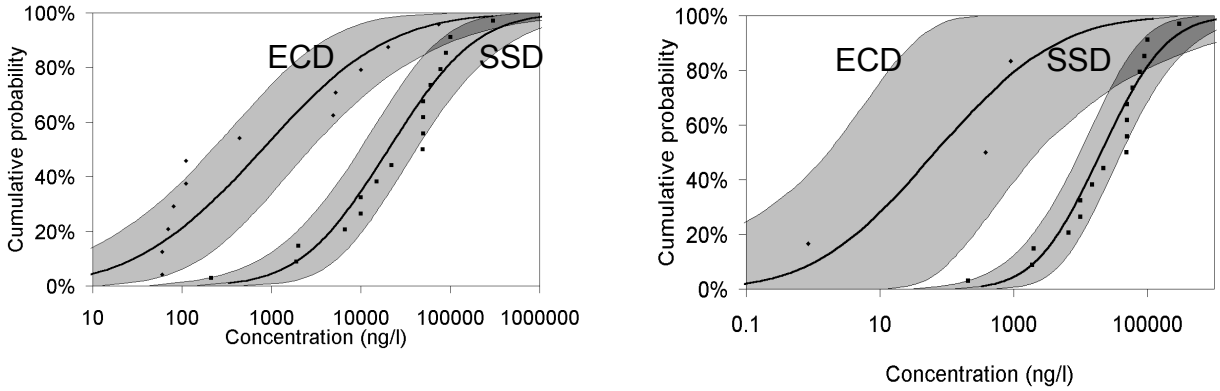


Figure 7: Atrazine ECD (Exposure Concentration Distribution) and SSD (Species Sensitivity Distribution) for the monitoring points in Alveringem (left) and Aalter (right)

Note that the lognormal distribution may not be the most appropriate distribution for the ECD in Alveringem. Nevertheless, the lognormal distribution was selected because it is the most suitable distribution for most other monitoring points and because the distribution choice will not affect the goals of this case study. Moreover, the effect of selecting another distribution is discussed in the previous case study. Autocorrelations (temporal dependence) were not observed for these two monitoring locations although serial dependence could be expected from a theoretical point of view.

All the results are shown from Figure 8 to Figure 10. Figure 8 shows the AUC of the EPP-JPC and CPP-JPC for the two monitoring stations. The AUC of the JPC is an estimate of the (probabilistic) risk. The AUC of the uncertainty bands of the JPC is an estimate of the uncertainty interval of the risk. Figure 9 shows the probabilistic risk quotient distribution for the two monitoring stations. One minus the cumulative probability at risk quotient equal to one is then an estimate of the (probabilistic) risk. One minus the cumulative probabilities of the uncertainty band at risk quotient equal to one are then estimates of the uncertainty interval of the risk. The risk and its uncertainty interval are shown at the bottom of Figure 10. The uncertainty distribution of the risk is shown at the top of Figure 10. It can be derived from this figure that the uncertainty distribution of the probabilistic risk can have different shapes. For the Alveringem location, it has the shape of a normal distribution whereas for the Aalter location, it has the shape of a lognormal distribution. The numerical values of the estimated probabilistic risk and its uncertainty interval are shown in Table 1.

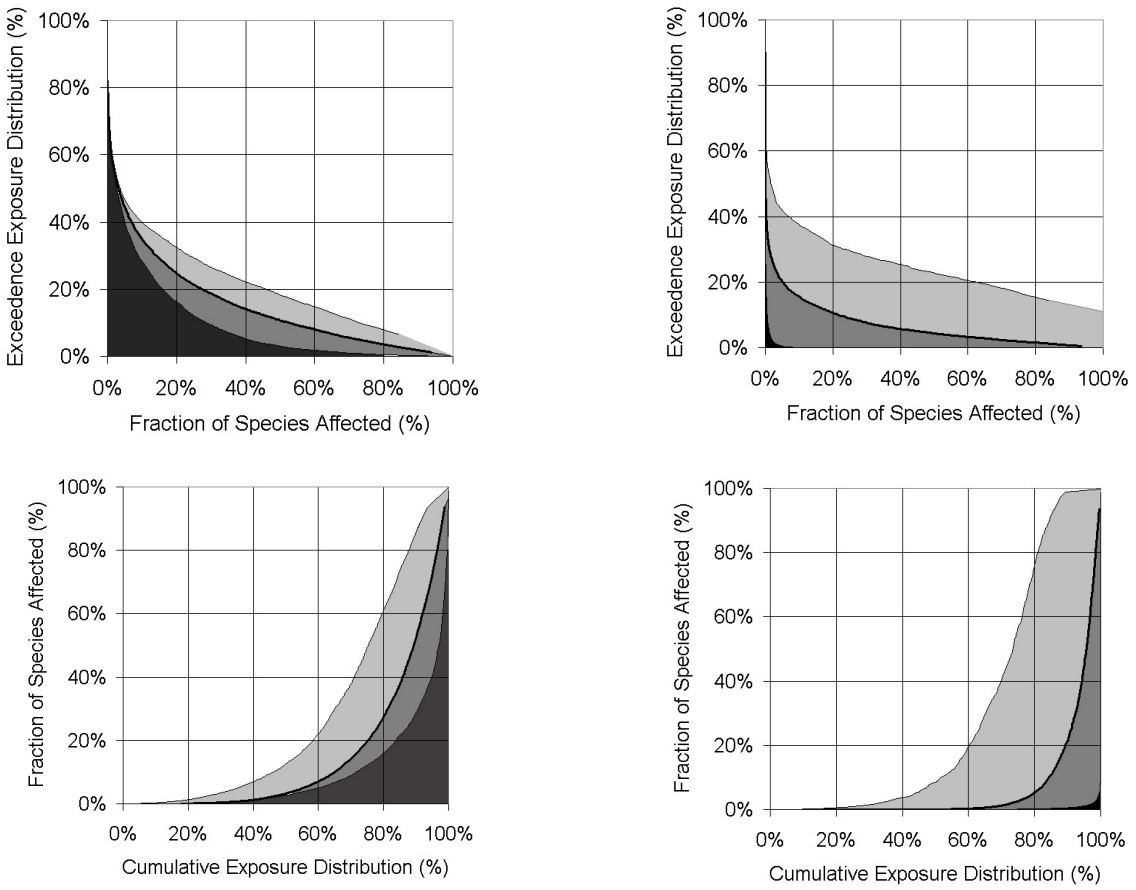


Figure 8: Atrazine JPC (Joint Probability Curve): EPP (Exceedance Profile Plot) on top and CPP (Cumulative Profile Plot) on bottom for the monitoring points in Alveringem (left) and in Aalter (right)

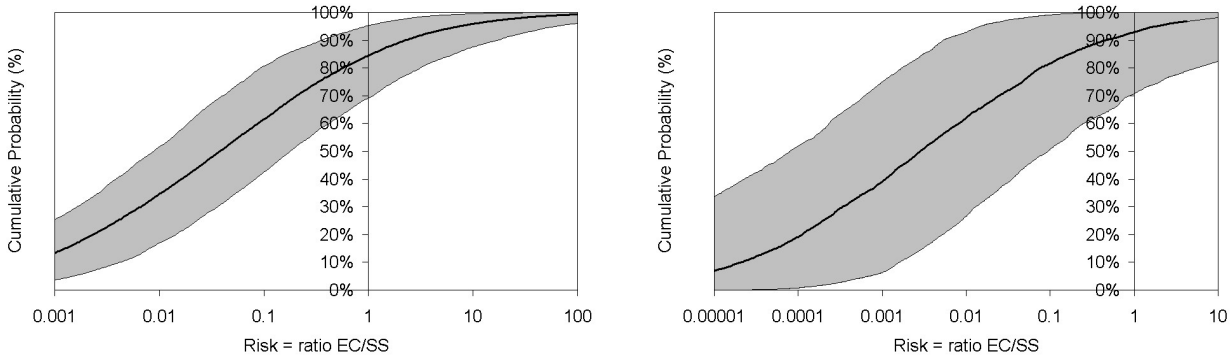


Figure 9: Atrazine probabilistic risk quotient distributions for Alveringem (left) and Aalter (right)

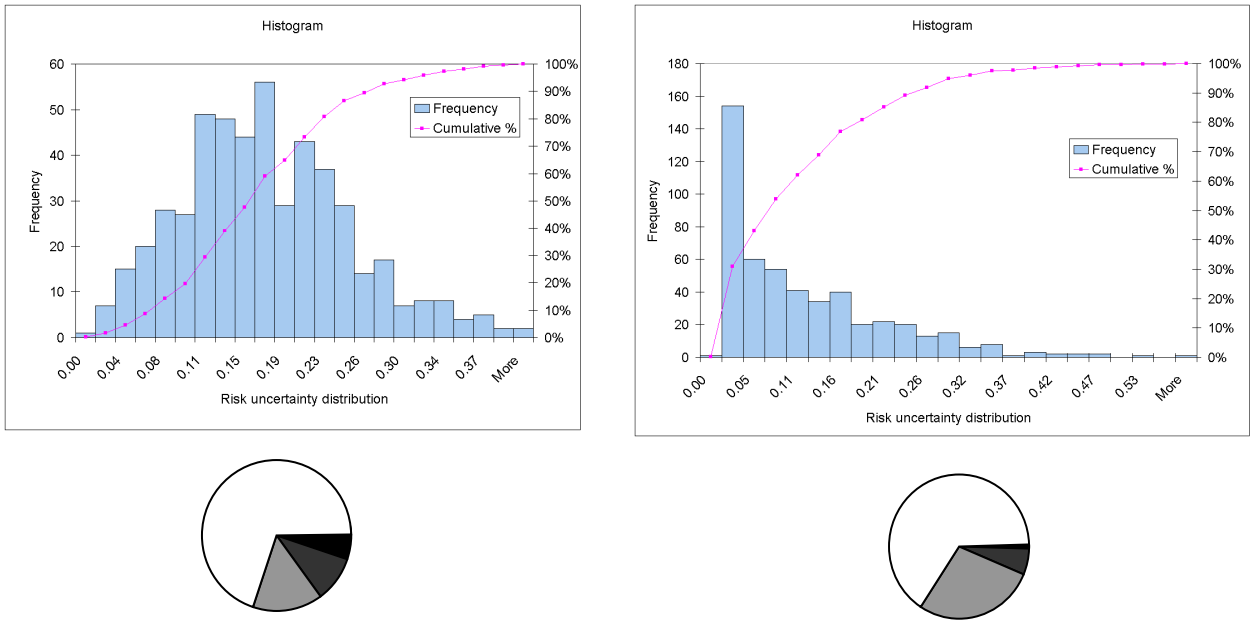


Figure 10: Atrazine risk and its uncertainty distribution (top) and interval as pie diagram (bottom) for the monitoring points in Alveringem (left) and Aalter (right)

Table 1:Probabilistic risks and their uncertainty estimates according to several methods

	Alveringem	Aalter
Lookup table 5.3 in Aldenberg *et al.* (2002)	15 (-)	- (-)
AUC in JPC	14 (5-30)	6 (1-33)
Probabilistic risk quotient method	14 (5-30)	6 (1-33)

4.1.4. Discussion

In the previous two sections, methodological and practical case study results were presented. Here, a discussion is made on the risk characterisation methods and their interpretation.

Consider the example in Figure 1, where the deterministic risk quotients for all different environmental situations are equal to 10 indicating potential risk (since larger than 1). On the other hand, probabilistic risks for the same environmental situations range from low to high risks: 7 to 94% depending on the location and shape of the ECD and the SSD. Both the hypothetical example in Figure 1 and the two case studies illustrate that the probabilistic risk characterisation considers the quantitative information of the full range of the ECD and SSD (including lower SS than HC_5 and higher ECs than the 90^{th} percentile) instead of only considering the upper tail of the ECD and

the lower tail of the SSD. Probabilistic Ecological Risk Assessment (PERA) is capable of differentiating the different environmental situations in Figure 1. The PERA is more realistic and therefore more refined.

The probabilistic risk quotient method takes less time to apply compared to the other methods because it is easy to use and understand. The case studies show it can handle different types of parametric distributions but the methodology can easily be extended for non-parametric distributions as well.

The distribution types of the ECD and SSD determine the shape and location of the probabilistic risk quotient distribution. In the first case study, two distribution types were tested for the SSD. Both distributions had a similar central tendency but differed in the tails, especially in the lower tail (see Figure 5). As a result, the probabilistic risk quotient distributions have the same central tendency but differ in the tails (see Figure 6). So, as long as the probabilistic risk quotient = 1 is around the central tendency of the probabilistic risk quotient distribution, the choice of the distribution type for the SSD (and ECD) is less important. When the probabilistic risk quotient = 1 is situated in the tail of the probabilistic risk quotient distribution, the choice of distribution for the SSD (and ECD) becomes more important.

In the second case study, the monitoring stations in Alveringem and Aalter show that the width of the uncertainty interval on the risk heavily depends on the uncertainty of the ECD. The lack of more EC information results in a larger uncertainty interval for the risk. The first case study shows the same for the SSD. The uncertainty of the lower tail of the Pareto fitted SSD is smaller than the uncertainty of the lower tail of the lognormal fitted SSD. In short, reducing the largest uncertainty source (ECD or SSD) will have most effect on decreasing the uncertainty of the probabilistic risk.

There exist many possibilities to visualise probabilistic risk and its uncertainty distribution/interval: as the AUC in CPP or EPP JPC, as a probabilistic risk quotient distribution, as a histogram or as a pie chart. All have their advantages and disadvantages. The probabilistic risk quotient distribution communicates the underlying idea of the method, being an extension of the deterministic risk quotient method, very well known by risk managers and decision-makers. The pie chart has opportunities for display on geographical maps. The advantages of the JPC will be further explored in Chapter 4.2.

4.1.5. Conclusions

The probabilistic risk quotient method is a probabilistic extension of the well-known and familiar deterministic risk quotient and is capable of estimating a probabilistic risk, defined as the

probability of a randomly selected EC exceeding a randomly selected SS (i.e. a probability instead of a ratio >1/<1) and is in addition, capable of estimating an uncertainty interval representing the sampling error that exists because of the practical inability to collect an infinite number of data. The method can handle all types of parametric and nonparametric distributions and is easy to use and interpret at the same time. Attention was also given to the visualisation of the probabilistic risk and its uncertainty interval. A proposal to present probabilistic risk in the form of a pie chart was made and could be useful when geo-referenced probabilistic risks need to be displayed on a map.

The next chapter will show that, although all current risk characterisation methods are a large improvement compared to the conventional deterministic risk characterisation, the current methods still have some limitations that need further research.

Chapter 4.2

-

Limitations of Current Probabilistic Risk Characterisation Methods

This chapter is going to be published in:

Verdonck F. A. M., T. Aldenberg, J. Jaworska, and P. A. Vanrolleghem. (2003). Limitations of current risk characterisation methods in probabilistic environmental risk assessment. *Environmental Toxicology and Chemistry* 22(9), in press.

Chapter 4.2

Limitations of Current Probabilistic Risk Characterisation Methods

In the previous chapter, it was shown that the calculation of a probabilistic risk can be done in many ways (e.g. Area Under the Curve (AUC) in Joint Probability Curves (JPCs)). However, in this chapter, it will be shown by means of several (hypothetical) examples and some theoretical considerations that the current risk characterisation methods have an integrative character and they focus on the statistical comparison of two distributions without properly considering the underlying interpretation of these distributions. Therefore, a clear environmental interpretation is needed.

4.2.1. Introduction

The goal of Probabilistic Ecological Risk Assessment (PERA) is to estimate the likelihood and the extent of adverse effects occurring in ecological systems due to exposure(s) to substances. It is based on the comparison of an Exposure/Environmental Concentration Distribution (ECD) with a Species Sensitivity Distribution (SSD) derived from toxicity data. The calculation of a probabilistic risk can be done in many ways (see previous chapter and Aldenberg *et al.* (2002)).

In this chapter we focus on Joint Probability Curves (JPCs) (Solomon & Takacs, 2002). JPCs come in two forms (see Chapter 4.1): either as a graph of ECD Exceedence against fraction of species affected (cumulative probabilities of SS) called Exceedence Profile Plot (EPP), or as a graph of fraction of species affected against cumulative probabilities of EC called Cumulative Profile Plot (CPP). Both JPCs represent the same risk curves, they are just different ways of visualisation. The first is more common and is therefore shown here.

An example of the construction of an EPP based on ECD and SSD is given in Figure 1. The dashed curves in the right panel represent thresholds between different types of decisions. These thresholds have to be determined by decision-makers and their position or shape may move depending on the decision to be made. However, currently no quantitative measures exist for these JPC-thresholds.

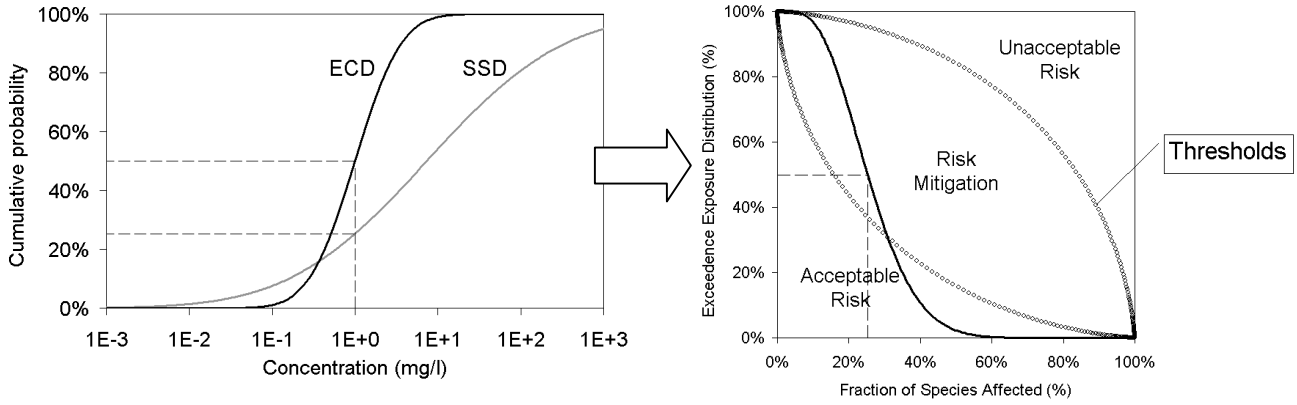


Figure 1: Joint Probability Curve (JPC): an example of an Exceedence Profile Plot (EPP) and thresholds for acceptance

Every data point on the JPC can be easily interpreted (e.g. in Figure 1: for 50% of the time (or in 50% of the locations), 25% of the species will be affected (e.g. a chronic effect on reproduction)), but interpreting and quantifying the entire JPC seems to be more difficult (e.g. in Figure 1, how acceptable or unacceptable is this particular JPC?). Sometimes, the AUC is calculated as a measure of probabilistic risk (see Chapter 4.1).

In this chapter, it will be shown that these current risk characterisation methods have the drawback that they focus on the statistical comparison of two distributions without interpretation of the underlying distributions. Important environmental information and interpretation is lost when only the integrative risk is calculated. Theoretical considerations and simulation studies with hypothetical scenarios will illustrate these shortcomings and show that interpretation of the resulting risks should also be carefully made. Depending on the interpretation of the ECD and the SSD, the interpretation of the resulting risk can be totally different.

4.2.2. Theoretical Considerations

As described above, several probabilistic risk calculation methods are available. The probability of some randomly selected EC exceeding some randomly selected SS has been demonstrated to be a common measure of risk (Aldenberg *et al.*, 2002). This probability of EC exceeding SS is equal to

the probability that the Risk Quotient (RQ = EC/SS) becomes larger than one or that $\log_{10}$(EC/SS) becomes larger than zero since:

Equation 1: $$\text{Risk} = \mathrm{P}(\mathrm{EC} > \mathrm{SS}) = \mathrm{P}\left(\frac{\mathrm{EC}}{\mathrm{SS}} > 1\right) = \mathrm{P}\left(\log_{10}(\mathrm{EC}) - \log_{10}(\mathrm{SS}) > 0\right)$$

Logarithmic transformations were made because the risk can be easily calculated analytically when lognormal distributions are assumed for the ECD and the SSD. The result of a quotient of two lognormal distributions (EC and SS) is again a lognormal distribution. The calculation of the parameters of the risk quotient distribution and the resulting risk is already discussed in Chapter 4.1. The formula for the probabilistic risk in the case of two normal distributions is:

Equation 2: $$P(\log EC - \log SS > 0) = 1 - \Phi_{\mu_{\log EC} - \mu_{\log SS}, \sqrt{\sigma^2_{\log EC} + \sigma^2_{\log SS}}}(0) = \Phi_{0,1}\left(\frac{\mu_{\log EC} - \mu_{\log SS}}{\sqrt{\sigma^2_{\log EC} + \sigma^2_{\log SS}}}\right)$$

where $\Phi_{m,s}(x)$ is the cumulative normal distribution of x with mean m and standard deviation s

In the second panel of Figure 2 in Chapter 4.1, the cumulative log risk quotient distribution $\Phi(\log RQ)$ is shown. The exceedance (reverse cumulative) log risk quotient distribution is $1 - \Phi(\log RQ)$.

This risk formula clearly illustrates the limitations discussed in the introduction. When the difference between the mean EC and the mean SS is fixed, then interchanging the two standard deviations does not change the risk. In other words, a small ECD variance and a SSD variance yield the same risk as that found after exchanging the variances but keeping the same means. However, the interpretation of this could differ, as discussed below.

The RQ method has the advantage of being easy to calculate for lognormal distributions. If the ECD or SSD have a probability distribution that differs from the lognormal one, the risk has to be calculated numerically or by means of simulation, e.g. a Monte Carlo analysis. This is not dealt with here since this will not influence the conclusions.

4.2.3. Hypothetical Case Study

An overview of the hypothetical scenarios studied below can be found in Table 1. In each scenario, the statistical and environmental interpretations are described.

Table 1: Overview of the scenarios for the simulation studies

Scenario	ECD*	SSD*	Statistical interpretation	Environmental interpretation
1	*LN*(0,1)	*LN*(0,1)	• same mean for EC & SS • same variances for EC & SS → ECD = SSD	• same distribution for environmental concentrations and species sensitivity
2	*LN*(0,1)	*LN*(0,5)	• same mean for EC & SS • small variance for EC • large variance for SS	• small range in temporal or spatial environmental concentrations • very sensitive and very insensitive species (broad range)
3	*LN*(0,5)	*LN*(0,1)	• same mean for EC & SS • large variance for EC • small variance for SS	• large range in temporal or spatial environmental concentrations • all species have more or less the same sensitivity
4	*LN*(0,1)	*LN*(8,5)	• mean EC << mean SS • small variance for EC • large variance for SS	• low risk is expected as EC << SS • small range in temporal or spatial environmental concentrations • very sensitive and very insensitive species (broad range)
5	*LN*(0,5)	*LN*(8,1)	• mean EC << mean SS • large variance for EC • small variance for SS	• low risk is expected as EC << SS • large range in temporal or spatial environmental concentrations • all species have more or less the same sensitivity

* Lognormal distribution with parameters mean and standard deviation of the log-transformed data

The three rows in Figure 2 show the ECD, SSD and JPC for the first three scenarios. In each scenario, the risk is an identical 50%. The same results are obtained when using Table 5.3 from Aldenberg *et al.* (2002), which tabulates probabilistic risks by only varying the mean and standard

deviation of the log(ECD) relative to the log(SSD). Clearly, the three scenarios represent different environmental situations and one would expect that they lead to different managerial decisions. This is because environmental effects may differ substantially depending on the interpretation of the ECD or SSD.

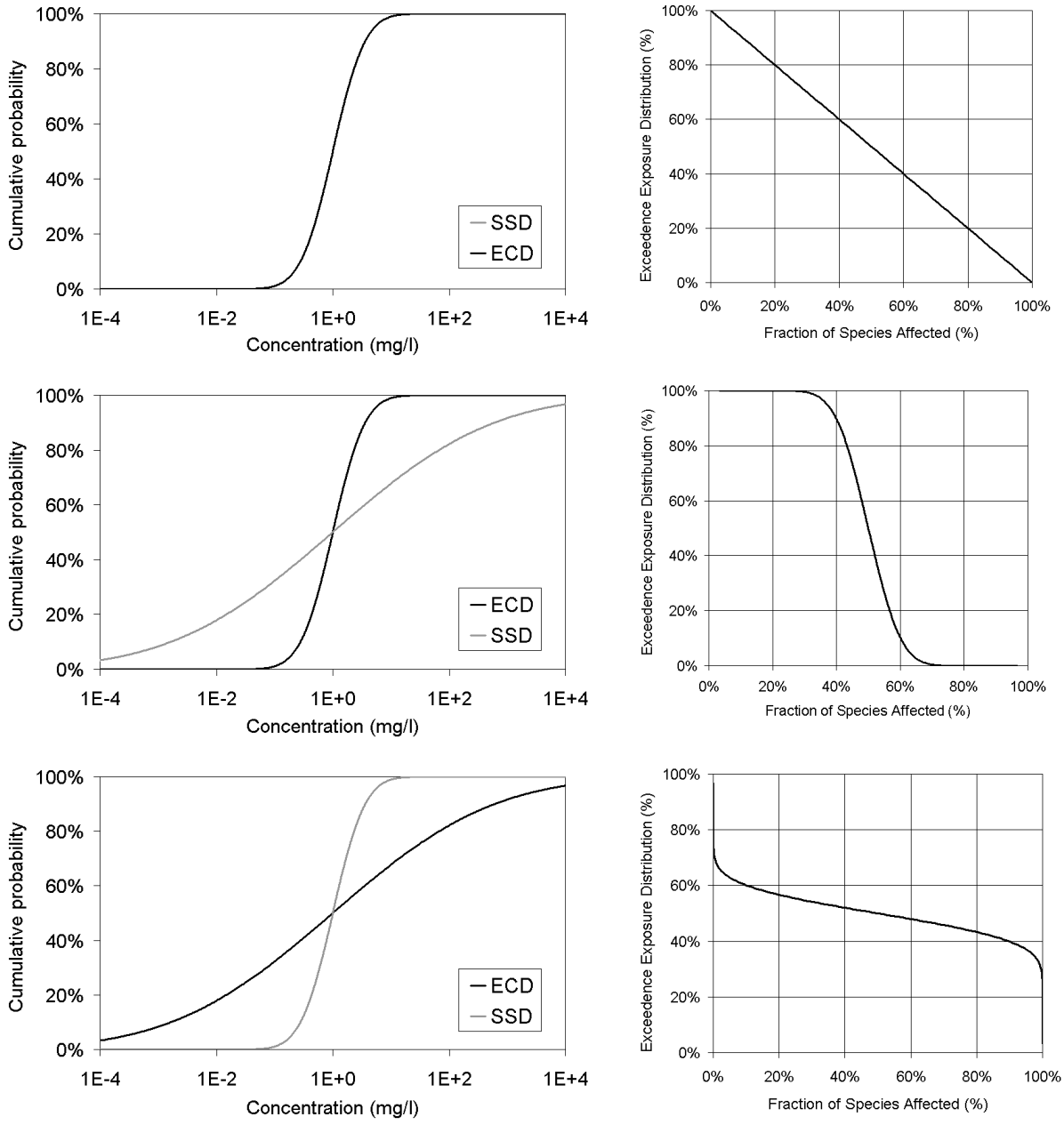


Figure 2: Simulation results: first column shows the ECD (Exposure Concentration Distribution) and SSD (Species Sensitivity Distribution) (on log scale), second column shows the JPC - EPP (Joint Probability Curve – Exceedance Profile Plot); first row: scenario 1, second row: scenario 2; third row: scenario 3

To illustrate this dependence on interpretation, a distinction can, for example, be made between an ECD representing temporal variability or an ECD representing spatial variability. When the ECD represents the temporal variation at one monitoring location, scenario 2 (small temporal EC variance, large SS variance) produces a better environmental outcome than scenario 3 (large temporal EC variance, small SS variance) because in scenario 3 almost all species will die approximately 50% of the time. Now, to assess the effect on the ecosystem, a lot will depend on the recovery time of the organisms. In scenario 2, approximately 50% of the species will die all of the time but the other 50% might survive. When the ECD represents the spatial variation of a chemical, scenario 3 (large spatial EC variance, small SS variance) will arguably lead to a better environmental outcome than scenario 2 (small spatial EC variance, large SS variance) because in scenario 3 all species will die in approximately 50% of the geographical locations while in the other 50% geographical locations, no species are likely to die. This results in more biodiversity. In scenario 2, approximately 50% of all species will die at all locations, leading to lower overall biodiversity.
To further illustrate this dependence on interpretation, a distinction can, for example, also be made between a SSD representing acute toxicity and a SSD representing chronic toxicity. In both interpretations, scenario 2 (small temporal EC variance, large SS variance) produces a better environmental outcome than scenario 3 (large temporal EC variance, small SS variance) because in scenario 3 almost all species will suffer adverse effects (either acute or chronic) approximately 50% of the time. In scenario 2, approximately 50% of the species will suffer adverse effects (either acute or chronic) all of the time but the other 50% will not.

This difference in interpretation of the risk in the different scenarios is also reflected in the shape of the JPC (right column of Figure 2). Probabilistic risks are, like deterministic risks, only comparative measures. Information on the type of risk and the underlying data needs to be considered for proper interpretation. This may be an advantage of probabilistic methods when compared to deterministic risk calculation since probabilistic methods are more transparent and realistic.

Similar results were obtained with scenarios 4 and 5. These results are shown in Figure 3. They both result in an expected risk of 5.8%.

In Figure 4, five JPCs are shown resulting in the same risk (12%). However, it is not straightforward to put thresholds of acceptability. It is shown above that, depending on the interpretation of the ECD (and SSD), one JPC may be concluded to be better or worse than the others (even though they have the same risk). Because of the integrative nature of risk calculation, information leading to interpretation is lost.

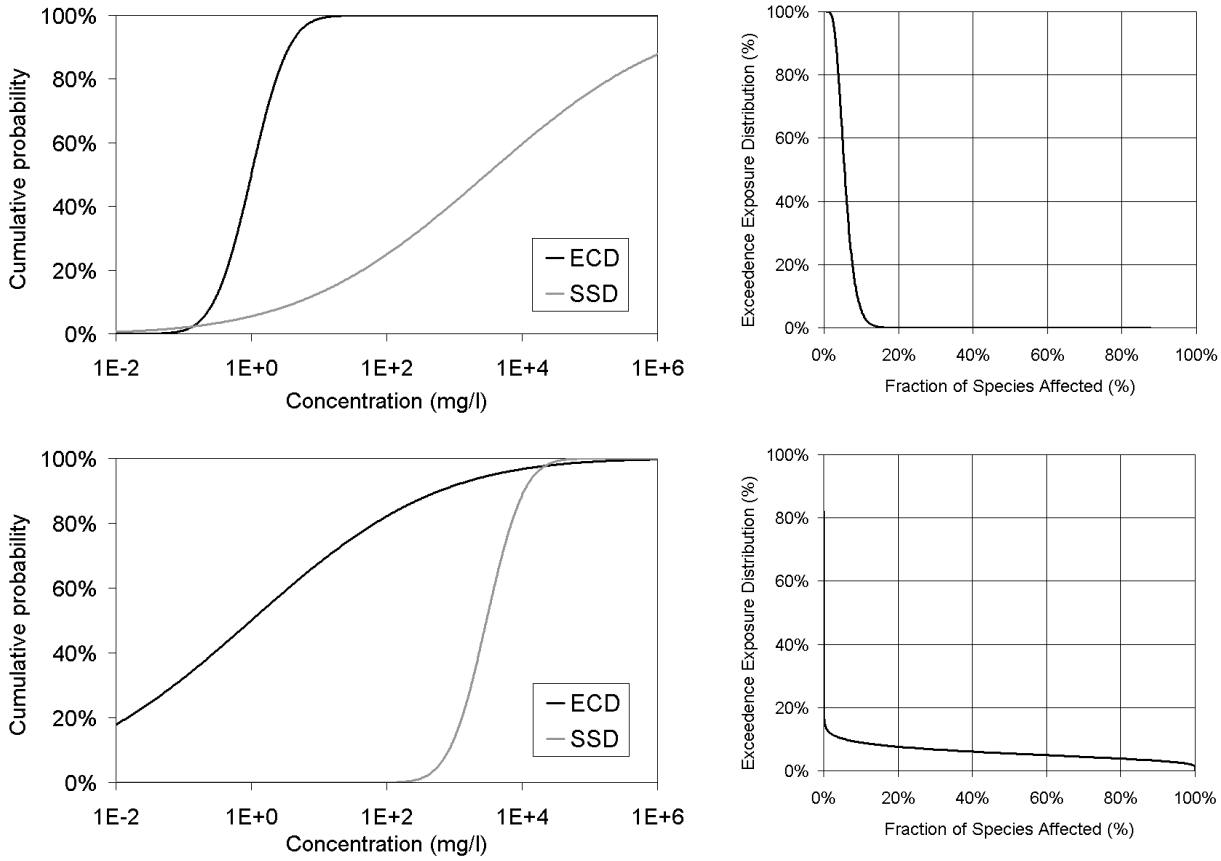


Figure 3: Simulation results of the hypothetical data sets; first column shows the ECD (Exposure Concentration Distribution) and SSD (Species Sensitivity Distribution), second column shows the joint probability curve; first row: scenario 4, second row: scenario 5

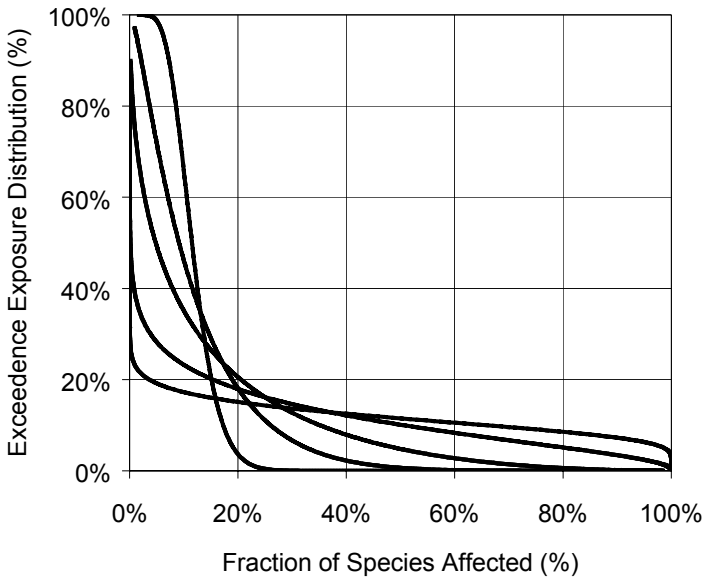


Figure 4: Several JPCs – EPPs (Joint Probability Curves – Exceedance Profiles Plots) all resulting in the same risk (12%)

Clearly, risk is a summary statistic, an integrative measure of the JPC that does not capture all aspects of the shape of the JPC. A potential solution would be to include additional JPC shape parameters. Those may be able to differentiate between several scenarios resulting in the same risk as discussed above. Just as the mean and variance are enough to characterise a normal distribution, means and variances of both ECD and SSD must be sufficient to calculate any shape parameter to characterise the entire JPC. One proposal for such shape parameter is made here. The gravity point of the AUC was calculated for each of the first three scenarios of Table 1. Its x- and y-coordinates (called x_g and y_g) were determined by discretising the AUC and using following formulae:

$$x_g = \frac{\sum\sum x\Delta x\Delta y}{\sum\sum \Delta x\Delta y}$$

$$y_g = \frac{\sum\sum y\Delta x\Delta y}{\sum\sum \Delta x\Delta y}$$

where Δx is an interval of the fraction of species affected
Δy is an interval of the exceedance exposure distribution

The ratio of its y_g- and x_g-coordinate is proposed as shape parameter:

$$Risk\ shape = \frac{\sum\sum y\Delta x\Delta y}{\sum\sum x\Delta x\Delta y}$$

The probabilistic risk and the risk shape parameter for the first three scenarios of Table 1 are shown in Table 2.

Table 2: Probabilistic risks and risk shape parameters (based on the ratio of the coordinates of the gravity point of the area under the curve) for the first three scenarios of Table 1

Scenario	Probabilistic risk (%)	Risk shape parameter
1	50	1
2	50	1.6
3	50	0.6

It is clear from Table 2 that the shape parameter can differentiate the different scenarios. A risk shape parameter equal to one indicates a symmetrical JPC. The closer to one, the more the JPC is symmetrical. If the shape parameter is larger than one, then the AUC is located most in the left part of the JPC. If the shape parameter is smaller than one, then the AUC is located most in the bottom part of the JPC. The corresponding environmental interpretations are described above. In order to

answer the question "how acceptable or unacceptable is the JPC in Figure 1", ranges for these shape parameters will need to be determined by decision-makers based on the underlying interpretation of the ECD and SSD.

4.2.4. Conclusions

Current risk measures, such as the Area Under The Curve (AUC) of a Joint Probability Curve (JPC), contain insufficient information to account for different environmental circumstances (i.e. different interpretations of the ECD and SSD) and to assess potential adverse effects on ecological communities. Therefore, it is recommended to always interpret the risk ecologically. This will force the environmental community to compare SSDs with adequate ECDs. Further research is needed on measures additional to the calculated risk that characterise the shape of the Joint Probability Curve and that has an environmental interpretation (depending on the interpretation of the EC and SS) to help to quantify and manage the risk. A first proposal was made for a risk shape parameter based on the ratio of the coordinates of the gravity point of the JPC.

Part 5

-

Spatio-Temporal Probabilistic Ecological Risk Assessment

Part 5

Spatio-Temporal Probabilistic Ecological Risk Assessment

Different levels of complexity to deal with uncertainty and several types of variability in the exposure and effects assessment can be distinguished. To provide clarity and structure, Figure 1 shows an overview of several tiers (of different level of detail) of Probabilistic Ecological Risk Assessment (PERA).

In the top panel of Figure 1, the deterministic ecological risk assessment is shown. A (random) variable (be it the Exposure Concentration (EC) or the Species Sensitivity (SS)) is considered as a crisp value. Uncertainty is partly ignored, partly considered in assessment or safety factors. The well-known environmental quality standard would fit in this tier (on the effects side). The second panel presents the PERA. It is an extension of the deterministic approach since both the inherent variability and uncertainty (shown as a grey band) is explicitly quantified and assessed (see Parts 3 and 4). However, all types of variability are lumped in a single distribution.

In the next panel/tier, the spatial variability is explicitly accounted for. The random variable (be it the exposure concentration or the species sensitivity) is considered for every spatial location (called here geo-referencing). As a result, the variability distribution no longer represents spatial variability but only lumps temporal and other types of variability. This leads to a large number of geo-referenced distributions but with smaller variances. This tier will be explored in Chapter 5.1.

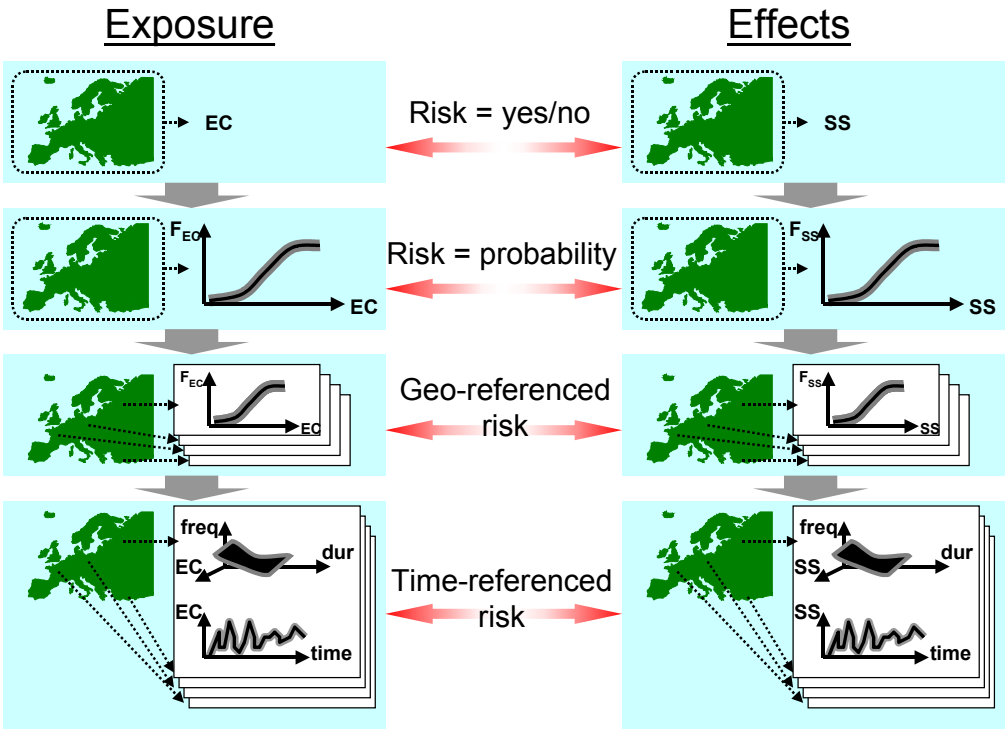


Figure 1: Several tiers of probabilistic environmental risk assessment of chemicals: top: deterministic analysis, below top: probabilistic analysis, above bottom: geo-referenced probabilistic analysis, bottom: time- and geo-referenced probabilistic analysis (EC: Exposure Concentration, SS: Species Sensitivity)

Time-referencing would further increase the level of detail and realism as time-specific information would be accounted for. This is represented in the lower panel of Figure 1. Time related information can be formatted in two ways in an attempt to capture the temporal variability. First, time series can be used as such or second, time series can be summarised into concentration-duration-frequency surfaces. These surfaces are three-dimensional plots with on the three axes the concentration, the duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. The first developments on this tier are explored in Chapter 5.2.

Chapter 5.1

-

Geo-Referencing Probabilistic Ecological Risk Characterisation

Parts of this chapter were published in:

Verdonck F. A. M., Jaworska J., Janssen C. R. & Vanrolleghem P. A. 2002. Probabilistic ecological risk assessment for chemical substances. *Proceedings iEMSs 2002, Integrated Assessment and Decision Support* 1, 144-149. Lugano, Switzerland. 24-27 June 2002.

Verdonck F. A. M., C. R. Janssen, J. Jaworska, and P. A. Vanrolleghem. (in press). Geo-Referencing of Probabilistic Risk of New Chemicals in Rivers. *Water Science & Technology*.

Chapter 5.1

Geo-Referencing Probabilistic Ecological Risk Characterisation

Once the Environmental/Exposure Concentration Distribution (ECD), the Species Sensitivity Distribution (SSD) and the resulting probabilistic risk are determined and characterised by variability and uncertainty (see Parts 3 and 4), all elements are available for a geographically geo-referenced Probabilistic Ecological Risk Assessment (PERA).

5.1.1. Introduction

Currently, the deterministic risk assessment insufficiently accounts for the inherent variability and uncertainty of the Environmental/Exposure Concentration (EC) and the Species Sensitivity (SS). In this respect, it is important to separate variability and uncertainty (see Part 2). Variability represents inherent heterogeneity or diversity in a well-characterised population and is not reducible through further measurement or study. The two most important sources of variability for the EC are spatial and temporal variability. Spatial and temporal variations of chemical concentrations can be captured in a variability distribution, called Exposure Concentration Distribution (ECD). Various SS towards a chemical (i.e. inter-species sensitivity/variability) can also be captured in a variability distribution called Species Sensitivity Distribution (SSD). In Figure 1, the variability distributions are shown as cumulative distribution functions by the black line. Uncertainty represents partial ignorance or lack of perfect information (e.g. sampling or measurement error), and can partly be reduced through further research (Cullen & Frey, 1999). In Figure 1, the uncertainty is shown as a grey band around the variability distribution function. For each percentile of the variability distribution, an uncertainty or confidence interval can be calculated (see Chapter 3.2).

The characterisation of the risk of toxicants to species, when both EC and SS are variable and uncertain, is the central issue in PERA. The resulting risk is no longer a simple ratio of crisp exposure and effects measures but rather a probability (see Chapter 4.1). In addition, the risk probability can be accompanied with a confidence or uncertainty interval (see Figure 1).

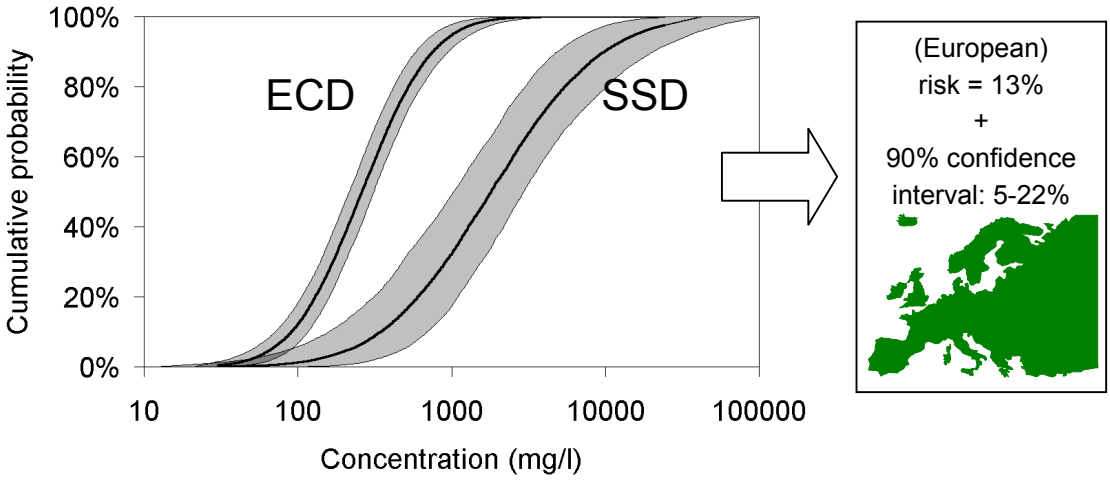


Figure 1: Probabilistic Ecological Risk Assessment (PERA) is based on the comparison of an Exposure Concentration Distribution (ECD) and a Species Sensitivity Distribution (SSD), the grey bands represent 90%-uncertainty bands

Currently, risk assessments, especially those for regulatory decisions, are done for generic situations determined by a set of default values. However for instance, the exposure spatial variability can be quite high. For example in Belgium alone, atrazine concentrations in surface water range from 50 ng/l (detection limit) to more than 1 mg/l (Vandenbroele *et al.*, 2000). This is a range of five orders of magnitude. Also the effects spatial variability can be considerable e.g. because of spatial variations in ecological communities. Until now, the EC and the SS were considered to be independent (see Part 4). However, the spatial variability of the EC and the SS can be such that in some locations, for instance, high SS are more likely to occur if the EC is high (e.g. because of species adaptation). Consequently, incorporating spatial characteristics and potential correlation between EC and SS of the receiving environment could further increase realism of a risk assessment.

The goals of this chapter are two-fold: first, some methodological developments are made to specifically account for the spatial information in the PERA and second, to show the usefulness, feasibility and potential of geo-referenced probabilistic risk for new and existing individual chemicals in several case studies.

5.1.2. Geo-Referenced Framework

The PERA methodology introduced above is well described and further developed in previous chapters. Evidently, it requires that variability distributions (and their uncertainty) are determined. Two different approaches can be used to determine the ECD and the SSD (see proposal framework in Part 2). Data from either measurements in the environment or toxicity tests can be used directly. The alternative is to use prediction or extrapolation models (especially in case of new chemicals). However, these models also need (other) data, which can again be characterised by uncertainty and variability. As a consequence, a distinction should be made between statistical methods for estimating data uncertainty and variability (see Chapter 3.2, 3.3 and 3.4), and methods for propagating uncertainty and variability through mathematical models (such as Monte Carlo analysis, see Chapter 3.1).

In Chapter 4.1, it was discussed how to calculate a probabilistic risk and its uncertainty interval. Among all risk calculation techniques available, one method was selected for all case studies reported here: the probabilistic risk quotient method. The probability of some randomly selected EC exceeding some randomly selected SS can be regarded as a measure of risk (Aldenberg *et al.*, 2002). It can be written as:

$$\text{Risk} = \text{P}(\text{EC} > \text{SS}) = \text{P}\left(\frac{\text{EC}}{\text{SS}} > 1\right) = \text{P}\left(\log_{10}\left(\frac{\text{EC}}{\text{SS}}\right) > 0\right) = \text{P}\left(\log_{10}(\text{EC}) - \log_{10}(\text{SS}) > 0\right)$$

If a geo-referenced risk assessment is to be obtained, each component of the risk assessment should be geo-referenced. Therefore, the three steps in a risk assessment are discussed in a geo-referenced framework: exposure assessment, effects assessment and finally risk characterisation.

5.1.2.1. Geo-Referenced Probabilistic Exposure Assessment (geo-ECD)

Spatial heterogeneities in chemical emission, release, dilution and degradation result in spatially different or geo-referenced ECDs. The ECD can be determined either through monitoring or through modelling.

Monitoring data will be used for case studies 2 and 3. Geo-referenced probabilistic exposure monitoring is more straightforward than modelling. Typically, environmental agencies collect exposure data on a range of chemicals in a number of monitoring locations.

When new, unreleased individual chemicals are assessed as in case study 1, prediction models are the only possibility to determine an ECD. GREAT-ER is such a (aquatic) chemical exposure prediction tool (Feijtel *et al.*, 1997). The system uses a Geographical Information System (GIS) for

data storage and visualisation, combined with simple mathematical models for prediction of chemical fate. GREAT-ER has already built in the idea of refining the exposure assessment by explicitly accounting for the spatial variability (geo-referencing the ECD). Instead of having one lumped ECD for an entire catchment (representing spatial and temporal variability), each river stretch has its own ECD (only representing temporal variability). In other words, the ECD for an entire catchment (ECDtot in the example of Figure 2) has been unlumped into several ECDs (ECD 1, 2 & 3 in the example of Figure 2). A Monte Carlo analysis propagates the temporal variability of the input parameters (such as the river flow).

5.1.2.2. Geo-Referenced Probabilistic Effects Assessment (geo-SSD)

Geo-referenced risk is only useful when both exposure (ECD) and effects (SSD) are geo-referenced (unless the SSD is the same everywhere). Instead of having one SSD for an entire catchment (representing spatial and other types of variability, SSDtot in the example of Figure 2), each river stretch can have its own SSD (SSD A & B in the example of Figure 2). Geo-referencing effects/SSD is still a large, unexplored area.

Many factors influence a geo-SSD. First, several physico-chemical environmental circumstances create ecological niches. These result in different communities and biodiversity. Consequently, different locations have different communities and species presence and determine the geo-SSD. Examples of such trends will be elucidated here using the river continuum theory (Vannote *et al.*, 1980). This theory suggests that the community functions associated with different reaches of a river will differ due to physical changes in the river from headwaters to mouth. For example, in the headwaters (stream orders 1 to 3) with its considerable shading and input of allochtonous material, the productivity to respiration ratio can be less than 1. However, this ratio can increase above 1 in the middle reaches of the river (stream order 4 to 6) where shading is reduced and ample nutrients are present. In the lower reaches, turbidity and depth can limit productivity and the ratio can drop below 1 again. Similarly, the macroinvertebrate functional groups (detritus shredders, collectors, scrapers and predators) shift with the availability of associated resources. Shredders will be prominent in headwaters where there is an abundance of coarse organic particulate matter such as leaves. They will decrease with distance downriver. Grazers will be most abundant in the middle regions where productivity is high. Collectors will be most abundant in the lower reaches. These very general patterns along the gradient from first order to higher order channels are modified in areas of confluence with lower order tributaries.
Second, physico-chemical characteristics are also determining the bioavailability and toxicity of, for instance, metals (Janssen *et al.*, 2000) and other chemicals. Consequently, different locations have different species sensitivities. Note that the same communities can be present at these different locations.
Third, depending on the metal background concentration, biological communities in these different systems may have differentially acclimated/adapted to the natural presence of metal concentrations resulting in varying community sensitivities (Janssen *et al.*, 2000).

Depending on the chemical, one factor will be more important than the other. It is not the goal of this thesis to determine and quantify the most important factors or to propose a framework of how to determine a geo-SSD. Geo-referenced effects are only used here to illustrate the usefulness of a geo-referenced risk. This may lead to more research in this area.

The SSD can be determined either through monitoring or through modelling. In case studies 1 and 3, modelling tools are used. There is no geo-SSD determined in case study 2 due to a lack of data. Field studies are the major source of geo-referenced ecotoxicity data. Unfortunately, these field studies are scarce and expensive. Therefore, effects monitoring is not (yet) a possible source of information.

5.1.2.3. Geo-Referenced Probabilistic Risk Characterisation

Finally, once both a geo-ECD and a geo-SSD are determined, a geo-referenced risk can be calculated for every river stretch in a river basin. Figure 2 shows an example. ECDtot and SSDtot are the lumped, non-geo-referenced exposure and effect distributions (reflecting spatial and other types of variability). By geo-referencing the exposure (ECD 1, 2 & 3 in Figure 2), effects (SSD A & B in Figure 2) and the resulting risk assessment, the spatial variability is explicitly accounted for in each local risk assessment and as a result the risk assessment will be more realistic. The combinations of ECD 1 & SSD A and ECD 2 & SSD B give smaller risks (in comparison with the risk from ECDtot and SSDtot) whereas ECD 3 & SSD A will give a higher risk (in comparison with the risk from ECDtot and SSDtot).

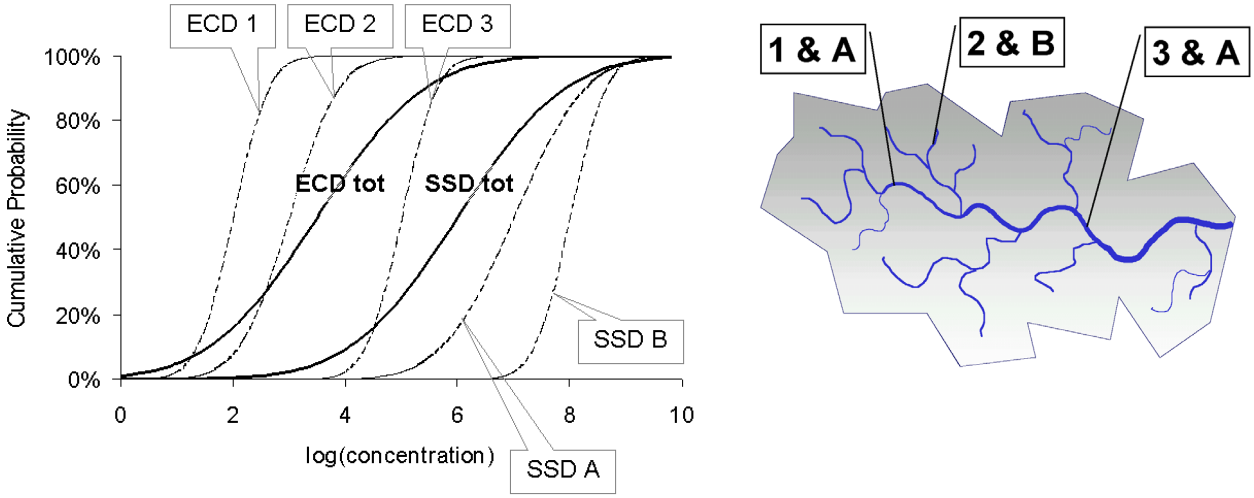


Figure 2: Split up of the lumped ECD and SSD into local ECDs and SSDs

Moreover, such a geo-referenced assessment may reveal a potential dependency between the EC and SS. Until now, the EC and the SS were expected and considered to be independent (see Part 4). However, there are reasons to believe that the spatial variability of the EC and the SS can be such that in some locations, for instance, high SS are more likely to occur if the EC is high. A dependency between EC and SS will influence the probabilistic risk calculation and is therefore important to consider. A geo-referenced assessment automatically solves this potential issue.

In addition, a geo-referenced assessment may also resolve spatial dependencies and autocorrelations of the EC or the SS. The EC at a monitoring location, for example, is influenced by the upstream ECs (measured at upstream monitoring locations). A spatial dependency of the EC or the SS will influence the uncertainty analysis as most variability and uncertainty estimation techniques from Chapter 3.2 assume independent observations. Note that a geo-referenced analysis cannot resolve temporal dependencies but a time-referenced analysis can (see Chapter 5.2).

If needed, these local risks can afterwards again be aggregated to a lumped probabilistic risk, for a catchment for instance. There are two pathways to obtain an aggregated total risk (see Figure 3): either the ECDs and SSDs are first aggregated and then the probabilistic risk is calculated (see pathway 1 in Figure 3) or either the probabilistic risk is calculated in each location and then aggregated (pathway 2).

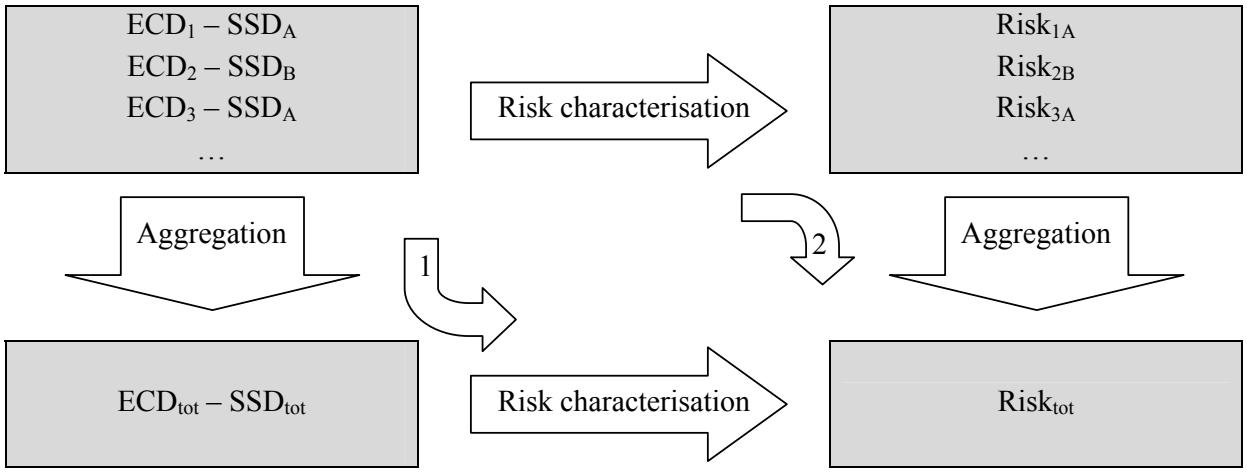


Figure 3: Two pathways to obtain an aggregated total risk

The two pathways give different outcomes because the probabilistic risk calculation is a non-linear operation (see Part 4): the lumped probabilistic risk (right term in the equation below, pathway 2) is not equal to the non-geo-referenced probabilistic risk based on the same spatial information (left term in the equation below, pathway 1).

$$Risk\left(ECD_1 + ECD_2 + \ldots\ ,\ SSD_A + SSD_B + \ldots\right) \neq Risk\left(ECD_1, SSD_A\right) + Risk\left(ECD_2, SSD_B\right) + \ldots$$

It is preferred to first calculate the probabilistic risk and then aggregate (pathway 2) because this is closer to reality.

5.1.3. Case Studies

Three case studies are discussed here to illustrate the potential and usefulness of a geo-referenced PERA. Not all case studies illustrate a "full" geo-referenced PERA. Some do not have a geo-referenced SSD. Others do not consider uncertainty. An overview is summarised in Table 1. The case studies include different regions (Belgium, Sweden) and different chemicals (a detergent chemical, a pesticide and a heavy metal). The ECD or the SSD are sometimes modelled, in other cases determined through monitoring.

Table 1: Overview of Case Studies

Chemical	Region	Exposure (ECD)		Effects (SSD)		Probabilistic Risk
		Model/ Monitoring	Geo?	Model/ Monitoring	Geo?	+ uncertainty?
1. Hypothetical detergent	Rupel basin (B)	Model	Yes	Monitoring	Yes	No
2. Atrazine	Flanders (B)	Monitoring	Yes	Monitoring	No	Yes
3. Cu	Sweden	Monitoring	Yes	Model but only 1 species	Yes	No

5.1.3.1. Case Study on the geo-referenced risk of a detergent chemical in the Rupel basin (Belgium)

5.1.3.1.1. Introduction

The goal of this case study is to show the usefulness, feasibility and potential of geo-referenced probabilistic risk for a new individual chemical in a river basin. The Rupel basin in Belgium was selected as case study area. The Rupel is a tributary to the river Scheldt. The area is 6700 km² large and contains the capital city of Belgium, Brussels (Figure 4). The chemical under study is a new, hypothetical anionic surfactant (to be used in detergents) for widespread use, once regulated. Degradation rates, chronic toxicity values and other parameters were chosen as realistic as possible (based on existing literature information of similar chemicals).

5.1.3.1.2. Determination of geo-ECD

When new, unreleased individual chemicals are assessed, prediction models are the only possibility to determine an ECD. Here, the (aquatic) chemical exposure prediction tool GREAT-ER was used. GREAT-ER 1.0 (Geo-referenced Regional Exposure Assessment Tool for European Rivers) calculates ECDs of consumer "down-the-drain-chemicals" in surface waters, for individual river stretches as well as for entire catchments.

The GREAT-ER project has been approached in a modular way, as previously described in detail in Feijtel *et al.* (1997). In the GIS data manipulation module, input data sourced from several databases and from the hydrology module are transformed into appropriate GIS formats. Geographical segmentation is also performed in this module. The hydrology module combines several hydrological databases with a hydrological model. It provides the GREAT-ER system with the required river flow distributions, flow velocities and river characteristics. The size of the Belgian study area did not allow to apply an advanced hydrological model that needs a lot of input parameters and effort. Therefore, an empirical hydrological model was developed. The applicability of a power function relating flow to the sum of the lengths of all upstream rivers has been demonstrated to be sufficiently accurate (Verdonck *et al.*, 2000).
The waste pathway and river modelling module is used for the prediction of chemical emission, of chemical removal/transformation during conveyance and treatment, and of chemical fate in rivers (Boeije *et al.*, 1997). Chemical fate in wastewater treatment plants and in rivers is described deterministically, with several levels of complexity being available to reflect the available information concerning both the chemical and the environment. For example, removal during sewage treatment can be either on a simple percentage removal basis, or alternatively it can be predicted using the SimpleTreat model (which is currently also used in EUSES (1997)).

An extensive monitoring programme for boron and for Linear Alkylbenzene Sulfonate (LAS) in six European pilot study areas has been performed in order to validate the system. The results illustrate that GREAT-ER can deliver very accurate predictions of chemical concentrations in a river basin, provided reliable input datasets and accurate hydrological and chemical fate models are used (Schowanek *et al.*, 2001).
A Monte Carlo analysis propagates the temporal variability of the input parameters (such as the river flow). Note that the input uncertainty is not considered here. For this, a second order or two-dimensional Monte Carlo would be needed (as demonstrated in Chapter 3.1). Geo-ECDs were predicted using GREAT-ER 1.0 for the Rupel basin (situation wastewater treatment plant infrastructure 1999) as reported in Verdonck *et al.* (2000).

5.1.3.1.3. Determination of "geo-SSD"

In the case study, an empirical approach based on considerable and arguable assumptions is used to determine a geo-SSD to study the environmental effect of a new chemical. The underlying idea of this approach is that a difference should be made between heavily polluted rivers (with small biodiversity) and rivers with a good water quality and large biodiversity. A heavily polluted river is assumed not to contain sensitive species. As a result, the SSD of a heavily polluted river will only contain more resistant species. Consequently, the risk of a new chemical affecting the species present in a heavily polluted river will be lower compared to a river with large biodiversity and sensitive species. This approach was only selected to illustrate geo-risk. Naturally, in the long term, the philosophy may be adhered that all species in all rivers (also the currently polluted ones) should be protected and a large biodiversity should be achieved and maintained.
Geo-SSDs were determined based on the empirical biodiversity approach described above. The Belgian biotic index, expressing species sensitivity and biodiversity of macro-invertebrates and ranging from 0 to 10 (De Pauw *et al.*, 1986), was used as an indicator of the overall biodiversity in the rivers. The biological monitoring network is rather dense (around 300 monitoring points). Only three categories of SS and biodiversity were determined: (1) moderate-good biodiversity (all species were selected in the SSD), (2) moderate-poor biodiversity (the 30% most sensitive species were removed from the SSD, see also graphically SSD A in Figure 2) and (3) poor-very poor biodiversity (the 50% most sensitive species were removed from the SSD, see SSD B in Figure 2). A lognormal distribution was fitted to these three SS data sets.

5.1.3.1.4. Results and discussion

Three tiers were simulated to assess the usefulness of a geo-referenced analysis. In the first tier, a non-geo-referenced PERA was performed (ECDtot & SSDtot of Figure 2). In the second tier, the

ECD was geo-referenced but the SSD was not (ECD 1, 2, 3 … & SSDtot). And in the third tier, both the ECD and SSD were geo-referenced (ECD 1, 2, 3 … & SSD A, B …).

In the first tier, a non-geo-referenced PERA of the new chemical was performed on the entire Rupel basin. The resulting risk is 27%. This can be interpreted as the probability that an Exposure Concentration (EC) from a randomly selected river stretch and a randomly selected day in the year will be larger than a randomly selected Species Sensitivity (SS) from a randomly selected river stretch (with large or small biodiversity). There was no confidence/uncertainty interval calculated for this risk probability.

The results of the geo-referenced PERA of the new, hypothetical but realistic, chemical on the Rupel basin (tier 3) are shown in Figure 4 (without uncertainty intervals). The colour pattern indicates how large the risk is. It can be derived from Figure 4 that spatial risks or geo-risks can vary from 0% to larger than 90%. However, the interpretation of these geo-risks is slightly different. They can be interpreted as the probability (for a particular location) that an Exposure Concentration (EC) from a randomly selected day in the year will be larger than a randomly selected Species Sensitivity (SS). For a particular river stretch, the spatial component is no longer needed in the interpretation because the risks are geo-referenced. Consequently, the risk assessment is more refined compared to the single number (27% in this case study) from a non-geo-referenced approach. Some rivers have higher risk probabilities of affecting the species, others have lower probabilities. These local risks could again be aggregated to a lumped probabilistic risk. Here, the aggregated risk of only the polluted stretches was weighted by the length of each river stretch to resolve any scale-dependencies (methodology, see Boeije *et al.* (2000)) and resulted in a risk probability of 9%.

The rivers under risk can also be studied in more detail to find the underlying causes of higher risks: is wastewater treatment plant infrastructure insufficient, under construction or to be improved, are the rivers under risk of high ecological importance, etc…? Such an analysis may help the decision-maker to avoid approving unsafe chemicals or rejecting safe chemicals. For example, the Nete basin (i.e. the Northern subbasin) has in general a very good wastewater treatment plant infrastructure, good water quality (large biodiversity) and low risks although there are some hot spots with large risks. In the Zenne river (i.e. downstream from Brussels, with bad water quality and low biodiversity), large risks were predicted even when the most sensitive species were already eliminated.

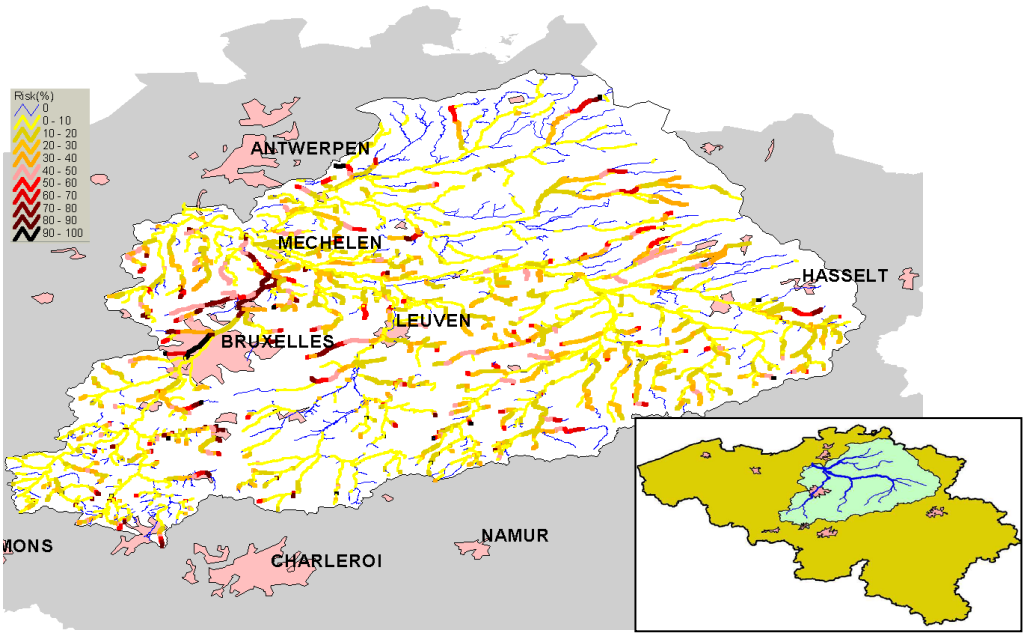


Figure 4: Geo-referenced probabilistic risk in % (no uncertainty intervals visualised) for a new, hypothetical but realistic, chemical on the Rupel basin in Belgium

This can also be seen in risk river profile plots (see Figure 5) where probabilistic risk is plotted against the distance in the river. The three tiers are shown in Figure 5. In the first tier, a PERA is performed on catchment level. The calculated risk is 27% (see above) and can therefore be visualised as a straight line (independent of the location in the river). In the second tier, the geo-risk is calculated based on a geo-ECD and a non-geo-SSD (as in GREAT-ER 1.0). In the last tier, a geo-risk is calculated based on both a geo-ECD and a geo-SSD.

The river profile could be separated into three categories for geo-SSD determination. In the most upstream 35 km of the river, water quality objectives aim for sustainable fish populations. For this, all sensitive and insensitive species were used in the SSD (category 1). Consequently, tier 2 and 3 give the same risk profile. However after 35 km from the source, the biodiversity is moderate to poor (category 2) and as a result, the risks for tier 3 are smaller than for tier 2. The peak at 90 km is caused by a large untreated discharge. But since in the downstream part of the river (last 40 km, almost near the mouth), water quality is so bad resulting in small biodiversity (category 3), adding an additional, new chemical will have less/no effect to the sensitive species since they are not present. Of course, the new chemical will impose additional stress on the existing insensitive species. It must be clear that the factor that was selected in this contribution to determine a geo-SSD should be regarded as an illustration of the methodology and points to the need of further development of a geo-referenced effect analysis (SSD).

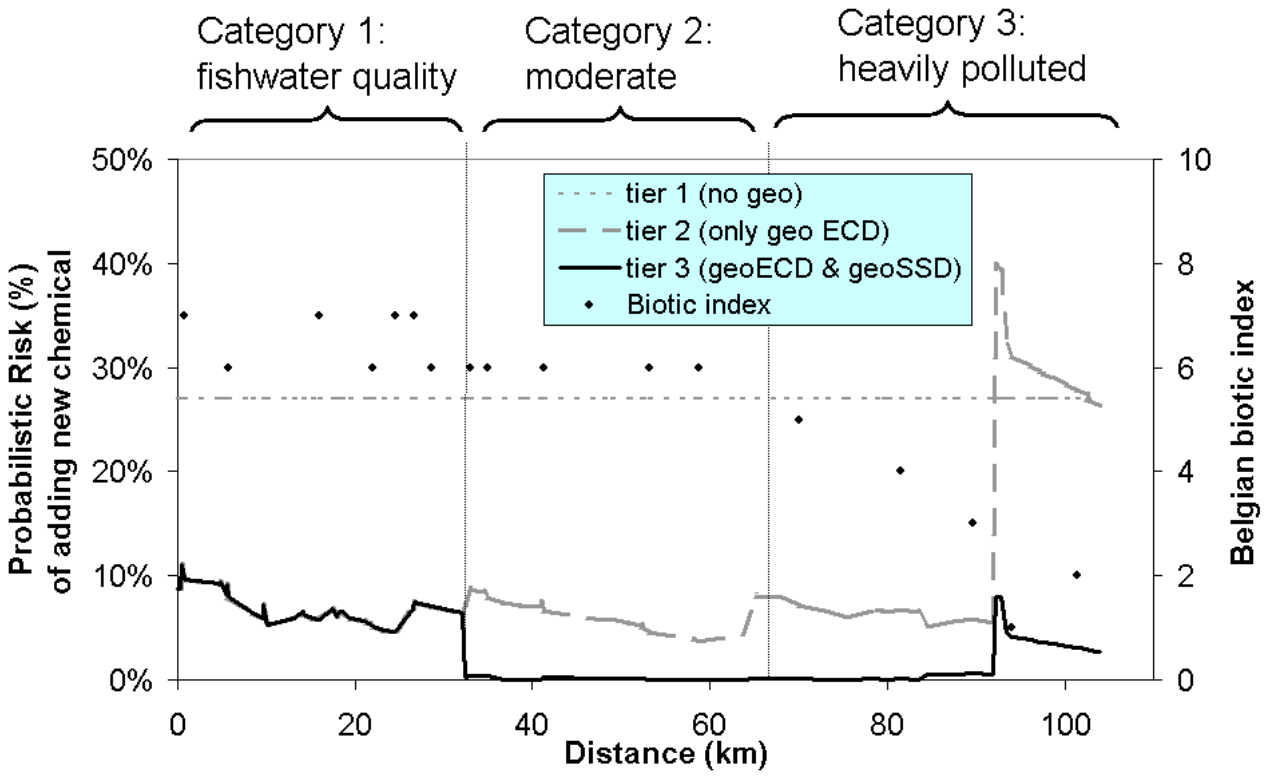


Figure 5: Probabilistic risk (of adding a new chemical) river length profile plot starting in the 'Grote Gete', 'Demer', 'Dijle' and ending in the 'Rupel' in the Rupel basin (Belgium)

An additional bottleneck for geo-referenced PERA may be data availability. The more information one wants to incorporate (to make the PERA more realistic), the more data will be needed. Data acquisition and management is however improving rapidly and will warrant the future application of the proposed methodology.

5.1.3.2. Case Study on the geo-referenced risk of atrazine (a pesticide) in Flemish surface waters (Belgium)

5.1.3.2.1. Introduction

Atrazine is a widely used persistent herbicide. Consequently, it is considered a great potential for groundwater and surface water contamination. Therefore, it is frequently detected. The greatest risk of atrazine runoff occurs shortly after the application because it hasn't had time to adhere to the soil particles and is still at the surface of the soil (Lipishan & Lee, 1996). Probabilistic risks and their

90%-uncertainty intervals were predicted for the pesticide atrazine in the river catchments of Flanders in Belgium.

5.1.3.2.2. Determination of geo-ECD

The ECs were obtained from the Flemish environmental agency (VMM, 2001). Atrazine was (mostly monthly) measured at 134 locations from 1991 till 2000. Only the reliable data from the years 1997 till 2000 were considered. It is assumed that this monitoring network is representative for all rivers in Flanders. The spatially aggregated cumulative empirical distribution function of all ECs is shown in Figure 6. A lognormal distribution was assumed and fitted to all data but the distribution did not fit very well to the data (see grey curve in Figure 6) because the data are left censored i.e. the value 50 ng/l is frequently observed. This value corresponds with the detection limit of atrazine. Censored data can be handled in different ways. Replacing every value below the detection limit with a random number between zero and the detection limit (here 50) is better compared with the detection limit or half of the detection limit (Govaerts *et al.,* 2001). After such correction for censoring, the lognormal distribution now fits very well to the data (see black curve in Figure 6).

The lognormality was also checked for several individual monitoring stations. The ECs were lognormally distributed for most of the monitoring stations based on Kolmogorov-Smirnov statistics, expert knowledge and graphical inspection of normal QQ plots. For each station, a lognormal distribution was fitted to the data. This resulting ECD represents the variation of the concentration (mostly temporal) at that station.

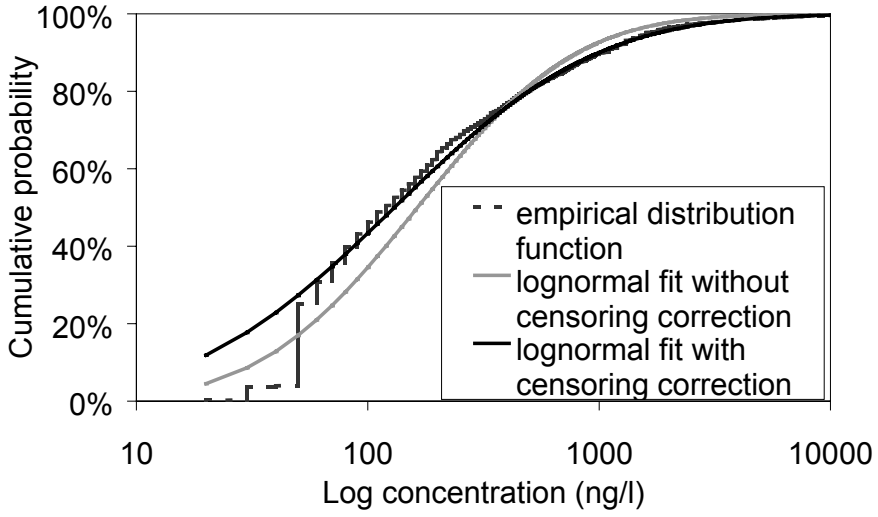


Figure 6: Cumulative probability distribution of atrazine measurements in surface waters in Flanders

5.1.3.2.3. Determination of (non-geo-)SSD

The data set for the atrazine SSD consists of chronic toxicity values (NOEC: No Observed Effect Concentrations) and can be found in Versteeg *et al.* (1999). A lognormal distribution was assumed and fitted to the data. The fit was satisfactory.

5.1.3.2.4. Results and Discussion

In Chapter 4.1, the probabilistic risk was presented as a pie chart as in Figure 7. The entire pie represents 100%. The grey shades indicate how large the probabilistic risk is with a pre-defined certainty. The larger the white slice, the lower the risk is. The more black, the larger the risk is. The larger the grey slices are, the more uncertainty exists on the risk. The example shows that the median risk is 23% (50% certainty) and there is 95% certainty that the risk is smaller than 45%. It can be derived from Figure 8 that visualisation of the probabilistic risk as pie chart promises to be a good communication tool.

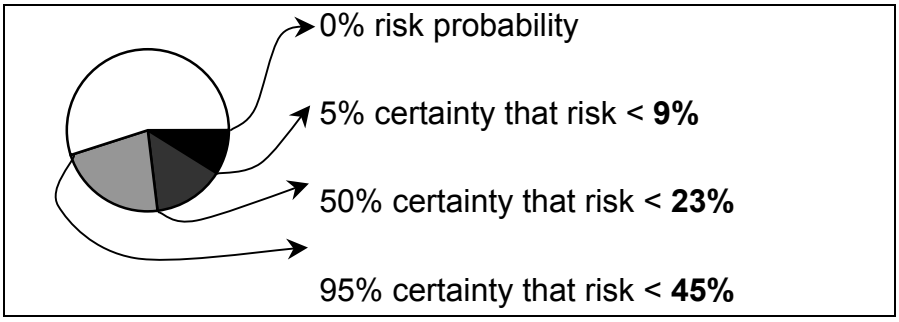


Figure 7: Visualisation of the risk of 23% and its 90%-uncertainty interval

The results of the local PERA of atrazine for all monitoring stations in the river networks of Flanders are shown in Figure 8. The geo-risks can be interpreted as the probability (for a particular location) that an Exposure Concentration (EC) from a randomly selected month in the year will be larger than a randomly selected Species Sensitivity (SS).

The results of the local PERA of atrazine in the basins of Flanders indicate that the predicted atrazine risks are around 1.6% for many monitoring locations. Most of the median risks are smaller than 5% and most of the upper risk 90% uncertainty intervals are smaller than 15%.

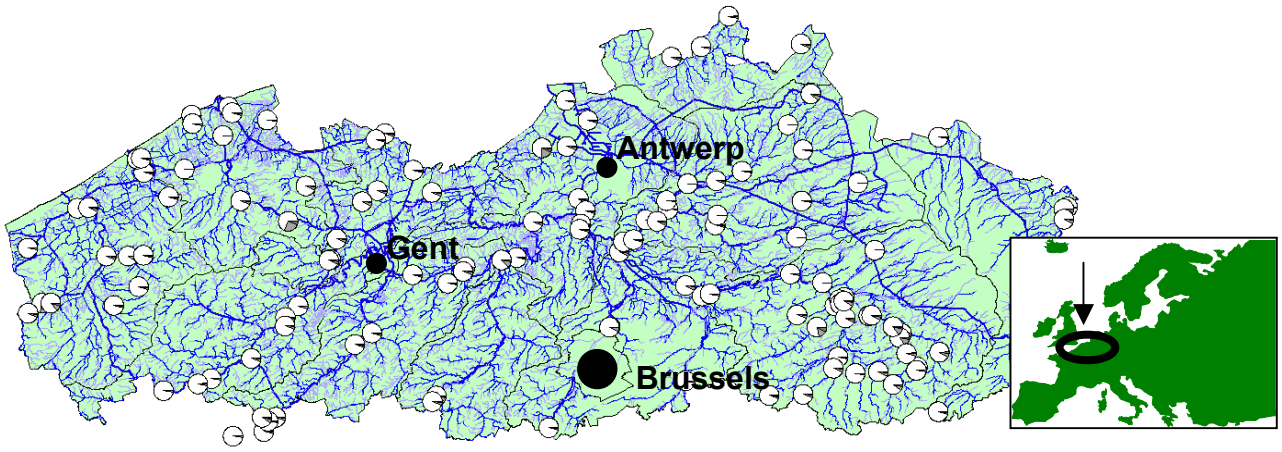


Figure 8: Probabilistic atrazine risk in the catchments of Flanders (Belgium)

Note that geo-referencing the risk is most useful when both the ECD and the SSD are geo-referenced. Here, only the ECD was geo-referenced and the SSD was considered to be the same for every location. But in this case, hot spots could also be located based on the geo-referenced ECDs alone. Moreover in reality, spatial differences lead to different local SSDs (e.g. Janssen *et al.* (2000)).

5.1.3.3. Case Study on the geo-referenced risk of Cu (a metal) in surface waters of Sweden

The third case study illustrates a geo-referenced PERA for Copper (Cu) in the surface waters of Sweden. More information can be found in the MSc. Thesis of De Laender (2003).

5.1.3.3.1. Introduction

Cu is a natural element. It is a vital trace element in our daily diet, helping to ensure the health of body and brain. Like us, animals and plants could not thrive without it. Copper also has many applications: conductor of electricity and heat among non-precious metals, water piping, inhibitor of the growth of bacteria, viruses and fungi, roofing, etc… Because of its diverse applications, it may also increase the natural background concentration in the environment.

Until recently, environmental water quality standards and risk assessment procedures for metals in surface waters were predominantly based on total and/or dissolved metal concentrations (Janssen *et al.*, 2000). However, the importance of bioavailability and toxicity modifying factors like pH, hardness and Dissolved Organic Carbon (DOC), is increasingly being recognized and is a major

contribution to geo-referencing Species Sensitivity (SS). The development of Biotic Ligand Models (BLM) that predict toxicity of metals to fish, invertebrates and algae (e.g. Di Toro *et al.* (2001); De Schamphelaere *et al.* (2002), Heijerick *et al.* (2002)) can be considered as an important step towards a scientifically sound protection of freshwater environments. Hence, the possible use of these models for regulatory purposes is gaining increased interest in both the scientific and the regulatory community.

In this case study, a geo-referenced SS is determined and the powerful combination of GIS and copper bioavailability models (BLM) is demonstrated. The Swedish surface waters were selected because of their large spatial variability in bioavailability modifying factors. The main focus will be on the *Daphnia magna*-BLM (the most advanced chronic BLM, De Schamphelaere and Janssen, unpublished). As *Daphnia magna* is very sensitive to copper, it may be a good model organism for the initial identification of possible environmental risks.

5.1.3.3.2. Determination of geo-ECD

The ECs for Cu were obtained from the Institute of Environmental Analysis of the Swedish University of Agricultural Sciences (SLU, http://info1.ma.slu.se). A lognormal distribution was fitted to the ECs of each location. Consequently, the ECD at each location reflects mostly the temporal variability of the Cu-concentration at that location. More information on the determination of the geo-ECD can be found in De Laender (2003).

5.1.3.3.3. Determination of geo-SS

The main assumption of the BLM is that metal toxicity occurs as the result of free metal ions reacting with binding sites at the organism-water interface (either physiologically active sites, leading to a direct biological response, or transport sites, leading to metal transport into the cell followed by a subsequent, indirect biological response), which is represented as the formation of a metal-biotic ligand complex. The concentration of this metal-biotic ligand complex directly determines the magnitude of the toxic effect, independent of the physical-chemical water characteristics of the test medium (De Schamphelaere *et al.,* 2002).

For the simulation of the BLM, several input parameters are needed: total or dissolved organic carbon (TOC or DOC), pH, alkalinity, temperature, Ca, Mg, Na, K, Cl and SO_4-concentrations. Databases of Swedish surface water characteristics, which fulfil these requirements, were also obtained from the Institute of Environmental Analysis of the Swedish University of Agricultural Sciences (SLU, http://info1.ma.slu.se). Input parameters are monitored in the 'Riksinventeringar' database (about 5000 lakes). The most important outputs of the BLM are No Observed Effect

Concentrations (NOECs) for *Daphnia magna*. These *Daphnia magna* NOECs could, along with NOECs of other species, determine a SSD. The implementation of BLMs for other species is currently under development at Ghent University. This would enable the determination of a geo-referenced SSD. Here, only the sensitivity of one species was assessed (*Daphnia magna*).

The BLM model was then linked to the GIS in order to determine a geo-referenced bioavailability and consequently a geo-referenced SS (De Laender, 2003). Figure 9 shows the spatial variability of the BLM-predicted NOEC values for *Daphnia magna* in Swedish lakes and rivers. It is clear that the SSs are not equal for all sites. BLM-predicted NOEC values were mainly between 60 and 200 µg Cu/L. Typically, differences in SS can be a factor 3 large.

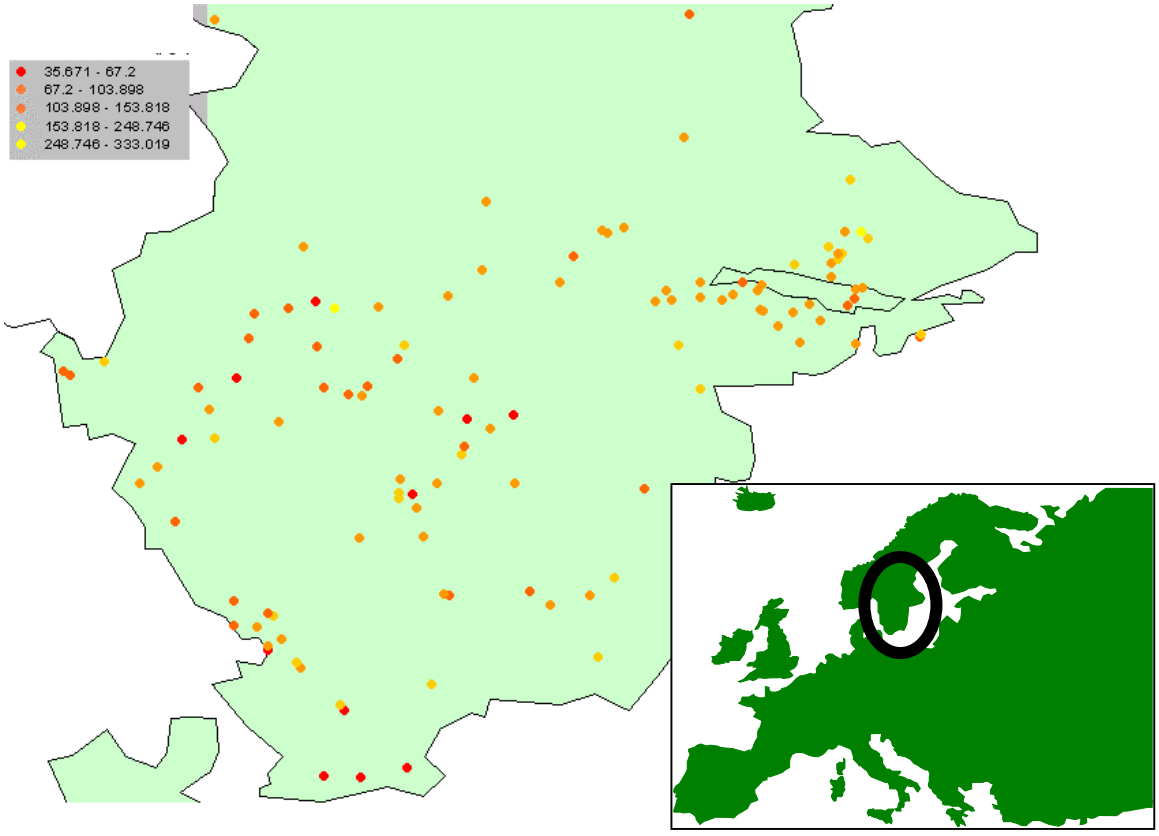


Figure 9: NOECs (No Observed Effect Concentrations) (µg/l) for Daphnia magna in Swedish surface waters

The BLM input parameters are characterised by temporal variability. In Figure 9, average BLM-predicted NOEC values are shown. These are based on average input parameters and variables. But, this temporal variability of the input results in temporally varying bioavailability conditions and consequently temporally varying NOECs. This temporal variability was captured by means of a Monte Carlo simulation. Further details can be found in the second case study of Chapter 3.1. The effects assessment therefore results in a probability distribution of SS for *Daphnia magna*. Note that

this distribution does not reflect inter-species sensitivities as the SSDs in the previous two case studies. This will affect the interpretation of the resulting probabilistic risk. It was not possible to perform Monte Carlo simulations for all Swedish sites in Figure 9 for practical limitations since this would increase the number of manual interventions (i.e. need for additional programming effort). Instead, a subarea around Stockholm was selected. The resulting average NOECs are shown in Figure 10. Some lower percentile (e.g. the 5th percentile) can also be visualized. This lower percentile could be used for environmental quality standard setting. However, it is up to the policy-maker how protective the standards should be.

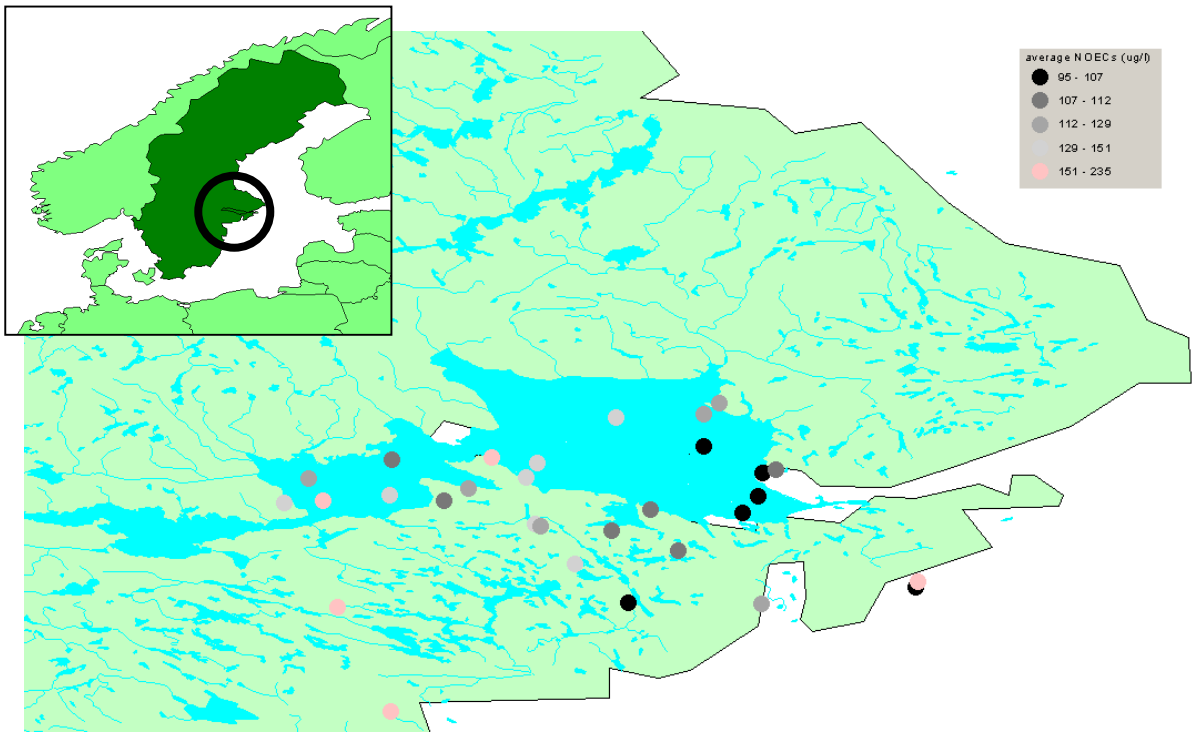


Figure 10: (Average) NOECs (No Observed Effect Concentrations) (µg/l) (from Monte Carlo analysis) for Daphnia magna in the Swedish surface waters around Stockholm

5.1.3.3.4. Determination of Geo-Referenced Probabilistic Risks

Geo-referenced probabilistic risks were determined for Cu on *Daphnia magna* for all monitoring stations in the river networks of Sweden. All locations resulted in a negligible probabilistic risk smaller than 0.002%. Therefore, the results are not shown. The geo-risks are interpreted as the probability (for a particular location) that an Exposure Concentration (EC) from a randomly selected moment in the year will be larger than a Species Sensitivity (SS) from a randomly selected moment in the year. Extending this case study to larger or other parts of Sweden may better illustrate the usefulness of this geo-referenced analysis.

5.1.4. Conclusions

A geo-referenced probabilistic ecological risk assessment framework was developed and illustrated with case studies. Geo-referencing makes the risk assessment more realistic as spatial variability and dependencies of the EC and SS are explicitly accounted for, i.e. less spatial variability is lumped in the probabilistic ecological risk assessment and therefore, this is useful for assessing risk of individual chemicals.

In addition, it was highlighted that geo-referencing effects (species sensitivity distribution) is now still a largely unexplored area but has important potential to further improve probabilistic ecological risk assessments. Geo-referencing PNECs and/or SSDs (in a GIS environment) and incorporating different levels of uncertainty results in a more realistic risk assessment which is preferable to the current practice of using a single (lumped) PNEC or SSD, representing an entire region or country.

Chapter 5.2

-

Towards
Time-referencing Probabilistic Ecological Risk Assessment

Parts of this chapter will be published in:

Verdonck F. A. M., Rousseau D. P. L., Bixio D., Thoeye C. & Vanrolleghem P. A. 2003. Added value of concentration-duration-frequency curves of waste water treatment plant effluent and river water quality. *International Congress on Modelling and Simulation (MODSIM 2003)*. Townsville, Australia, in press.

Chapter 5.2

Towards
Time-referencing Probabilistic Ecological Risk Assessment

As discussed earlier in the introduction of Part 5, different levels of complexity (several tiers of different level of detail in risk assessment) can be distinguished to deal with uncertainty and several types of variability in the exposure and effects assessment. Figure 1 shows a part of that overview. In the top panel, the deterministic ecological risk assessment is shown. The second panel represents the Probabilistic Ecological Risk Assessment (PERA). Time-referencing is represented in the lower panel of Figure 1. Time related information on effects and exposure can be formatted in two ways in an attempt to capture the temporal variability. First, exposure time series can be used as such or second, time series can be transformed into concentration-duration-frequency surfaces. These surfaces are three-dimensional plots with on the three axes the concentration, the duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. Geo-referencing is not considered in this chapter.

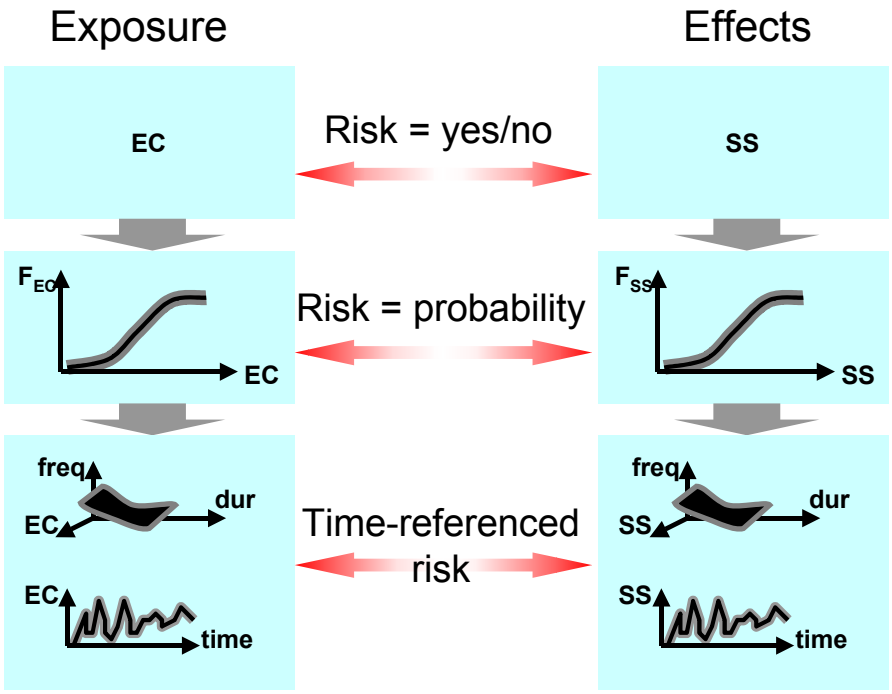


Figure 1: Several tiers of ecological risk assessment of chemicals: top: deterministic risk assessment, below top: probabilistic risk assessment, bottom: time-referenced risk assessment (EC: Exposure Concentration, SS: Species Sensitivity, grey band: 90% uncertainty band)

5.2.1. Introduction

This chapter presents some preliminary, exploratory steps of a time-referenced or dynamic ecological risk assessment. For this, knowledge and information from two converging fields will be used.

First, in the ecological risk assessment of specific chemicals, there is a trend to move from simple, empirical and deterministic approaches to more realistic, complex and also dynamic methods (see Part 2). For example, Milne *et al.* (2000) investigated the influences of exposure duration and frequency on the toxicity of short-term pulses of ammonia to rainbow trout and brown trout. Their results suggest that exposure duration and frequency are both important factors influencing the severity of effects in fish exposed to short-term ammonia peaks. The duration and exceedance of the exposure is presented in e.g. ECOFRAM (1999). But, there are only few developments on the level of the risk characterisation with such concentration-duration-frequency surfaces.

Second, in Urban Pollution Management (UPM): the urban waste water system (sewers, Waste Water Treatment Plant (WWTP) and receiving water) is treated as a single entity in which a change

in one part has implications for the other parts. Compliance with environmental standards is the main objective in UPM (FWR, 1998). For river planning purposes, two approaches are followed:

- River quality standards based on a percentile approach (~ probabilistic approach)
- Fundamental intermittent standards (~ time-referenced approach)

The percentile approach states that the pollutant concentration in the river should, for instance, comply with the given standard for at least 99% of the time (99^{th} percentile). The fundamental intermittent standards are concentration/duration values for a given pollutant that should not be violated more frequently than a specified value. In this way, temporal information is considered.
Only few developments are made on how to combine the exposure with effects assessment to eventually come up with a risk measure.

The goals of this chapter are to explore the possibility and usefulness of concentration-duration-frequency surfaces in a time-referenced ecological risk assessment and to compare them with the less detailed probabilistic methods in a Probabilistic Ecological Risk Assessment (PERA). In addition, it will be illustrated how this method can be used as a decision-support tool in risk assessment practice. The comparison between the two will be done by means of a case study on the effluent of a WWTP.

5.2.2. Methodology

The methodological background of PERA and concentration-duration-frequency surfaces are discussed in this section.

5.2.2.1. Probabilistic Method (Based on Probability Distributions)

Characterisation of the risk of toxicants towards species, when both exposure and effects are variable and uncertain, is the central issue in PERA. The resulting risk is a probability and can be calculated mathematically (see Chapter 4.1). In addition, the risk probability can be accompanied with a confidence or uncertainty interval (e.g. mean risk of 23% and its 90% confidence or uncertainty interval is 15-30%).

In PERA, the two most important sources of variability for the exposure concentration are spatial and temporal variability. Spatial and temporal variations of chemical concentrations in a river or WWTP effluent can be captured in a variability distribution, called Exposure Concentration Distribution (ECD). Various species sensitivities towards a chemical (i.e. inter-species sensitivity/variability) can also be captured in a variability distribution called Species Sensitivity Distribution (SSD). The reader is referred to previous chapters for more background.

5.2.2.2. Time-Referenced Method (Based on Concentration-Duration-Frequency Curves)

In this method, (spatial and) temporal variations of the exposure concentration (in the exposure analysis) and the ecotoxicity data (in the effects analysis) are not summarized into a probability distribution. Instead, the temporal variations are explicitly considered and summarized in concentration-duration-frequency surfaces. Let us consider no spatial variability for simplicity. A concentration-duration-frequency surface is based on a time series analysis and can be determined for both exposure and effects.

5.2.2.2.1. Exposure assessment

A concentration-duration-frequency surface is based on a time series analysis (FWR, 1998). These surfaces are basically histograms of the durations of exceedance of a set of (predefined) concentrations. This results in three-dimensional plots with on the three axes the concentration, the duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. However in practice, only two-dimensional plots are more frequently used. Here, duration-frequency curves are used because current legislation is still mainly based on standards, i.e. the concentration is constant. For ease of use, we shall continue to use the term concentration-duration-frequency curve. In return period analysis of combined sewer overflows, "concentration"-frequency curves are often used for design decisions in urban storm water management. The return period of the exceedance of a variable is found as the reciprocal of its probability of exceedance and the mean number of events per year (Grum & Aalderink, 1999). For water quality objectives, this approach seems less useful since the duration of an exceedance is not taken into account.

The theoretical example of Figure 2 shows two exceedances of the standard. The first and the second exceedance both last 6 hours. The duration of each exceedance is determined and then counted (to obtain a frequency of each exceedance with a particular duration).

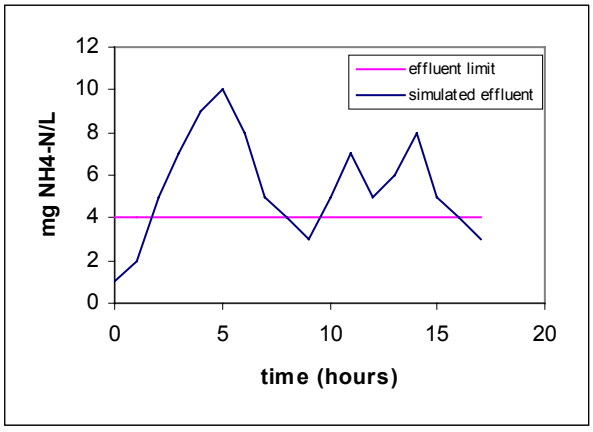


Figure 2: Illustrative effluent time series

Typically, a graph like Figure 3 is then obtained. Each duration interval is represented by its mean duration. From Figure 3, it can be seen that a large number of short exceedances occur (around 3 hours), next to a number of long-lasting exceedances (around 12 hours).

It is clear that these long-lasting exceedances could represent a higher environmental impact. However, this is only an indicative interpretation since these two-dimensional concentration-duration-frequency-curves do not give the severeness of the exceedance. If, for instance, the standard was set at 4 mg/L, an exceedance of 6 mg/L above the standard (as in the first peak of 10 mg/L in Figure 2) has the same weight as an exceedance of e.g. 60 mg/L above the standard would have.

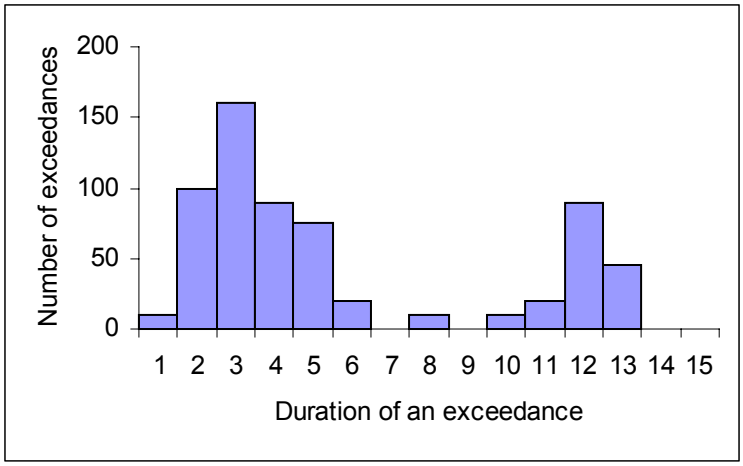


Figure 3: Example of an uncumulated concentration-duration-frequency curve

Concentration-duration-frequency curves are mostly given as cumulative curves. Such cumulative curve is obtained by adding up all exceedances from right to left. The interpretation of these curves changes slightly: instead of looking at the frequency of occurrence of exceedances with length between A and B hours, it now becomes the frequency of occurrence of exceedances lasting longer than A hours. Typically, a curve like Figure 4 is then obtained. The interpretation is as follows: 560 exceedances last two hours or longer while only 170 exceedances last 6 hours or longer. More information can be found in Rousseau *et al.* (2000). Such concentration-duration-frequency curve will be expressed as exposure concentration-duration-frequency curve.

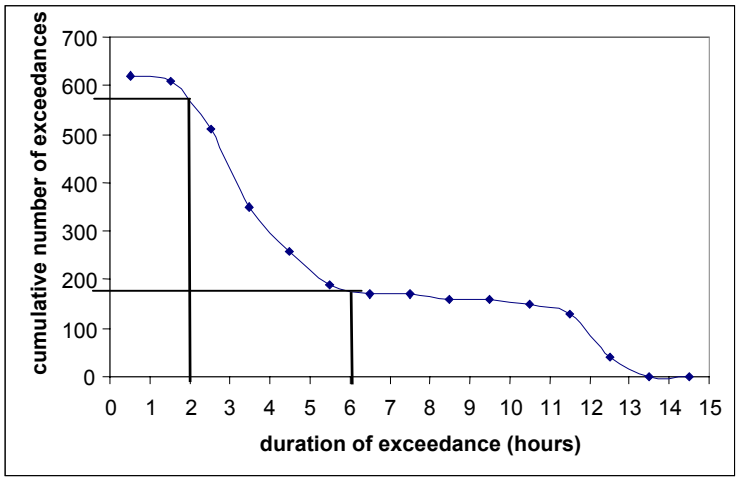


Figure 4: Cumulative concentration-duration-frequency curve

5.2.2.2.2. Effects assessment

Similarly as in the exposure assessment, concentration-duration-frequency curves can be constructed based on ecotoxicity tests for an effects assessment. Such concentration-duration-frequency curve will be expressed as effect concentration-duration-frequency curve. In ecological risk assessment, duration or frequency of an exposure event of a certain chemical is not often used. Usually, a concentration-duration curve is determined (e.g. Karman & Reerink (1998), Bonnomet *et al.* (2002)). An example is given in Figure 5.

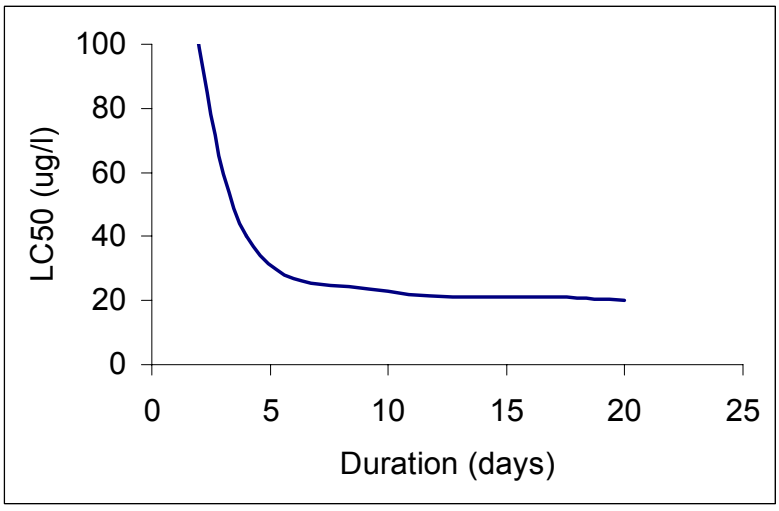


Figure 5: Concentration-duration curve for Daphnia magna exposed to cadmium chloride (based on Bonnomet et al. (2002))

The duration of a toxicity test is typically determined at the duration above, which the toxic concentration (LC_{50} in Figure 5) does not change anymore. In this way, the effect of duration is not explicitly considered but is dealt with in a conservative way.

Frequency-Duration curves and other concentration-duration-frequency curve variants are also explored in ECOFRAM (1999).

5.2.2.2.3. Risk Assessment

Finally, the exposure and effect concentration-duration-frequency curves are overlaid on the same graph. There is no risk if the effect concentration-duration-frequency curve is situated above the exposure concentration-duration-frequency curve. There is potential risk if the effect concentration-duration-frequency curve is situated below the exposure concentration-duration-frequency curve. We are not aware of any quantitative risk measures that integrate the exposure and effect concentration-duration-frequency curves.

5.2.3. Case Study

First, a short description of the case study is given. Then, the determination of the probabilistic ECD and SSD is discussed. Finally, the determination of the exposure and effect concentration-duration-frequency curves is discussed.

5.2.3.1. Problem Formulation

One of the challenges Aquafin NV, the company responsible for Waste Water Treatment Plant (WWTP) infrastructure in Flanders (Belgium), is now facing is to upgrade the patrimonium of old municipal WWTPs. These plants need to be retrofitted towards strict phosphorus and nitrogen removal consents. With this aim, a risk assessment procedure was developed based on a dynamic WWTP model with an uncertainty analysis module (see Chapter 3.1 and Rousseau *et al.* (2001)). The procedure has in practice already proven that capital investment can be reduced by up to 43%, producing savings of more than 1,2 million € (Bixio *et al.*, 2001).

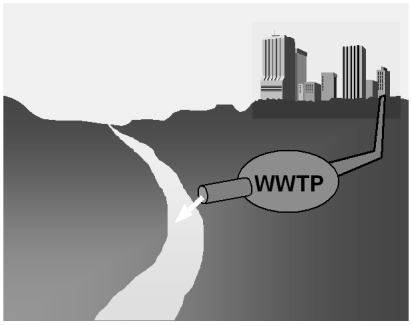


Figure 6: Situation sketch

The two main outputs of the procedure are a probabilistic Exposure Concentration Distribution (ECD) and an exposure concentration-duration-frequency curve. Both are accompanied with an uncertainty or confidence band.

In this perspective, a simplified case study is worked out here. The effect of total ammonia in the effluent of the WWTP in Hove (Belgium) (Bixio *et al.*, 2002) on the aquatic salmonid community in the receiving river "Bautersembeek" is studied (see Figure 6). Only the direct, acute toxic effects on salmonid fish populations are studied while indirect effects as eutrophication and chronic effects are not dealt with. Note that there are no combined sewer overflows considered either. The choice of this case study was driven by the need to illustrate the time-referenced concept for a realistic WWTP for which a model was available.

A key factor is the chemical speciation of ammonia. In aqueous solution, ammonia primarily exists in two forms, un-ionised ammonia (NH_3) and ammonium ion (NH_4^+), which are in equilibrium with each other according to:

$$NH_4^+ \leftrightarrow NH_3 + H^+$$

$$K = \frac{[NH_4^+][H^+]}{[NH_3]}$$

The individual fractions of un-ionised ammonia and ammonium vary markedly with temperature and pH. The mechanisms of the effects are poorly understood, but the pH dependence strongly suggests that joint toxicity of un-ionised ammonia and ammonium ion is an important component (EPA, 1999). Un-ionised ammonia is much more toxic than the ammonium ion. This is not surprising because it is a neutral molecule and thus is able to diffuse across the epithelial membranes of aquatic organisms much more readily than the charged ammonium ion. Ammonia is unique among regulated pollutants because it is an endogenously produced toxicant that organisms have developed various strategies to excrete, which is in large part by passive diffusion of un-ionised ammonia from the gills. High external un-ionised ammonia concentrations reduce or reverse diffusive gradients and cause the build-up of ammonia in gill tissue and blood (EPA, 1999).

Because of the importance of un-ionised ammonia, it became a convention in the scientific literature to express ammonia toxicity in terms of un-ionised ammonia concentrations, and water quality criteria and standards followed this convention. However, there are reasons to believe that the ammonium ion can also contribute significantly to ammonia toxicity under some conditions (especially pH) (EPA, 1999). Therefore, all concentrations will here be expressed as Total Ammoniacal Nitrogen (TAN).

5.2.3.2. Dynamic Waste Water Treatment Plant (WWTP) Simulation and Dilution Model + Uncertainty Analysis

The total ammonia ECD and concentration-duration-frequency curves in the river are both based on the effluent time series of the WWTP Hove. These effluent time series are predicted by means of a dynamic WWTP simulation model (activated sludge model No 1 and Tackacs model for clarification). More details can be found in Bixio *et al.* (2002). No dilution is assumed to convert the total ammonia probability distribution and concentration-duration-frequency curve of the WWTP effluent to the river. This assumption was made because the river flow is mainly determined by the effluent discharge flow of the WWTP and for conservative reasons. Obviously, a more realistic approach would be to account for the upstream river flow and total ammonia concentration time series, and as a result to obtain time series of dilution factors. Nevertheless, the main goal of this paper is to show the possibilities and potential of both proposed applications. This assumption is therefore acceptable in this perspective.

The Monte Carlo simulation takes into account both parameter and input uncertainty, in this way dealing with the difficulties to estimate model parameters and taking into account the inherent uncertainty in specific processes. As a consequence, an uncertainty or confidence band expressing the prediction uncertainty due to the uncertainty of the input variables and parameters will accompany the two resulting model outputs, namely the effluent ECD and concentration-duration-frequency curves. More information on this can be found in Chapter 3.1, Rousseau *et al.* (2001) and Rousseau *et al.* (2000).

5.2.3.3. Probabilistic Ecological Risk Assessment

5.2.3.3.1. Probabilistic Exposure Assessment

The probabilistic ECD was determined based on a time series analysis. It can simply be done by summarizing the time series as shown in Figure 7. The effluent ECD is then interpreted as the temporal variability of the effluent concentration of the chemical under study.

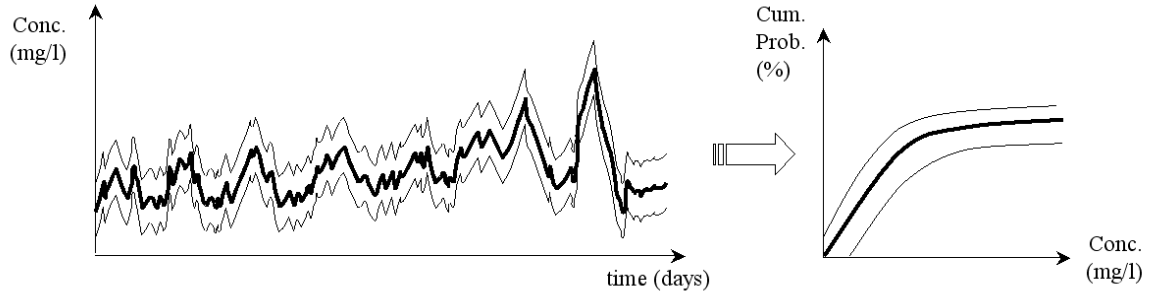


Figure 7: Aggregation of an uncertain time series into an uncertain Exposure Concentration Distribution (ECD)

In the left part of Figure 8, the total ammonia ECD in the river (predicted by the simulation model) is shown as a cumulative distribution function by the black line. The 90% uncertainty band (resulting from the Monte Carlo analysis) is shown in Figure 8 as a grey band around the ECD. Point A in Figure 8 (indicated by arrow) on this distribution can be interpreted as follows: 60% of the total ammonia concentrations are lower than 10 mg TAN/l. A 90% confidence or uncertainty interval on that is 8-13.2 mg TAN/l.

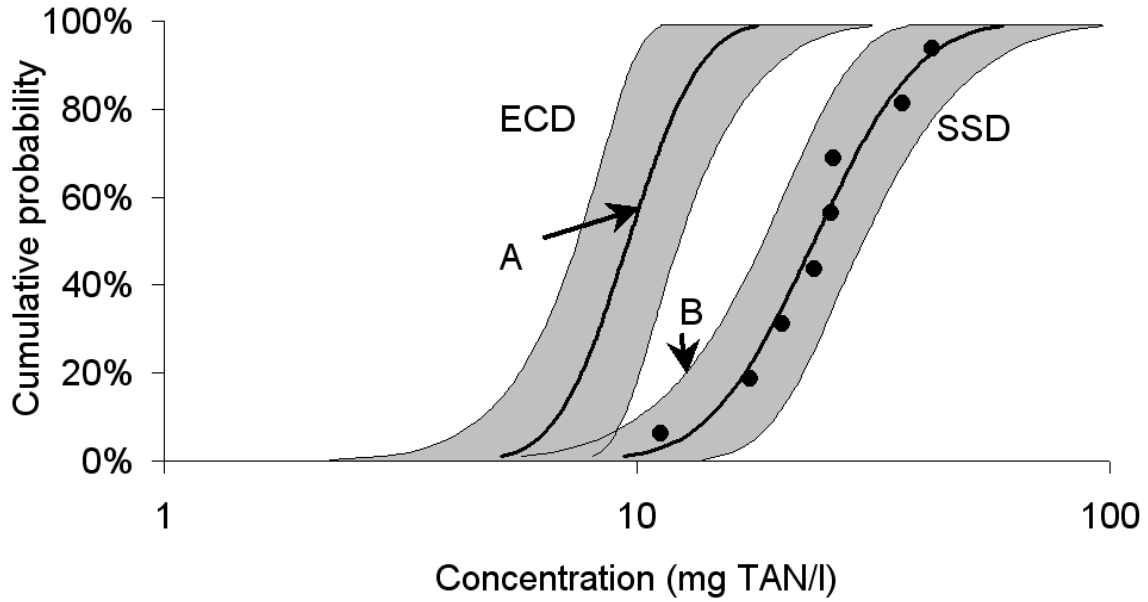


Figure 8: The Exposure Concentration Distribution (ECD) and the salmonid acute Species Sensitivity Distribution (SSD) of Total Ammonia Nitrogen (TAN) downstream of the waste water treatment plant of Hove

5.2.3.3.2. Probabilistic Effects Assessment

Acute ecotoxicity data were collected from EPA (1999). The assessment endpoints are LC_{50}s, which are Lethal Concentrations at which 50% of the organisms will die. Only data on salmonid species were considered: *Salvelinus fontinalis* (Brook trout), *Salmo trutta* (Brown trout), *Oncorchynchus aquabonita* (Golden trout), *Oncorchynchus clarki* (Cutthroat trout), *Oncorchynchus gorbuscha* (Pink salmon), *Oncorchynchus kisutch* (Coho salmon), *Oncorchynchus mykiss* (Rainbow trout) and *Oncorchynchus tshawytscha* (Chinook salmon). The LC_{50}s were expressed as Total Ammonia Nitrogen (TAN) at pH 8 (equivalent to EPA (1999)).

Various species have different sensitivities towards a chemical (i.e. inter-species sensitivity/variability). These differences are captured in a variability distribution called a Species Sensitivity Distribution (SSD). In the right part of Figure 8, the (salmonid) SSD is shown as a cumulative distribution function (black line). The sampling uncertainty is shown as a 90% confidence band and was determined by a parametric bootstrap method (see Chapter 3.2). Point B in Figure 8 (indicated by arrow) can be interpreted as follows: there is 95% certainty that 80% of the fish species will not be affected (50% lethality) at a total ammonia concentration lower than 12.6 mg TAN/l.

5.2.3.3.3. Probabilistic Ecological Risk Assessment

For this case study, a PERA was conducted based on the ECD obtained from dynamic model predictions and the SSD of salmonid species (with LC_{50} endpoints) in Figure 8. The resulting probabilistic risk is 8.4%, and may be interpreted as the probability that a randomly selected exposure concentration exceeds a randomly selected species sensitivity. This may not be acceptable for a water manager. This is especially true when its accompanying 90% uncertainty interval, based on parameter and input uncertainty in the WWTP model and sampling uncertainty of the ecotoxicity tests, is also considered. Risk is expected to range between 0.3-28%. This means that a water manager is 95% certain that the risk is smaller than 28%. This may not be accepted, especially because the LC_{50} endpoints are not suitable for sustainable salmonid fishery.

There are two possibilities to refine this risk assessment. First, the uncertainty could be reduced by either performing more toxicity tests that will result in a smaller uncertainty band on the SSD. Or, the uncertainty on the input parameters of the WWTP model could be reduced by collecting more samples or knowledge. This could be a reasonable option if the risk of 8.4% is around an acceptable level. Indeed, reducing uncertainty does not make the risk decrease. Second, one could make the risk assessment more realistic by explicitly considering duration and frequency of an exposure exceedance into the analysis. This is discussed in the next section.

5.2.3.4. Time-Referenced Risk Assessment (Based on Concentration-Duration-Frequency-curves)

In this method, temporal variations of the exposure concentration (in the exposure assessment) and the ecotoxicity data (in the effects assessment) are not summarized in probability distributions. Instead, the temporal variations are integrated into concentration-duration-frequency curves. A concentration-duration-frequency curve is based on a time series analysis and can be determined for both exposure and effects.

5.2.3.4.1. Concentration-Duration-Frequency Exposure Assessment

Concentration-duration-frequency curves result in three-dimensional plots with on the three axes the concentration, the duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. However in practice, two-dimensional duration-frequency curves are more frequently used (see methods section). The considered critical concentration was set at 4 mg TAN/l.

The dynamic model and successive time series analysis resulted in an exposure concentration-duration-frequency curve as shown in Figure 10. The small uncertainty band indicates that input uncertainty had a small effect on the output curve. Point A in Figure 10 can be interpreted as follows: there is 95% certainty that there are 7.2% of the total time exceedances of a total ammonia concentration of 4 mg TAN/l lasting for 2 hours or longer.

5.2.3.4.2. Concentration-Duration-Frequency Effects Assessment

Similarly to the exposure assessment, concentration-duration-frequency curves can be constructed for an effects assessment based on ecotoxicity tests. Such concentration-duration-frequency curve will be named an effect concentration-duration-frequency curve. Such curves for sustainable salmonid fishery were given in the urban pollution management manual (FWR, 1998). The data can also be found in Table 1. The return period, the duration (expressed in hours) and the concentration (expressed as mg NH_3-N/l) are shown. A three-dimensional representation is shown in Figure 9.

The return period was expressed as a monthly frequency (in percent). A return period of three months means every three months an exceedance or four times an exceedance in 12 months (4/12 = 33.33%). The choice of calculating the frequency based on months is justified by the fact that the resolution of the data is only one month.

Table 1: Concentration-frequency-duration data for un-ionised ammonia (mg TAN/l) for an ecosystem suitable for sustainable salmonid fishery (FWR, 1998)

Return period	Frequency by month (%)	Duration (h)		
		1	6	24
1 month	100	3.25	1.25	0.90
3 months	33.333	4.75	1.75	1.25
1 year	8.333	5.25	2.00	1.50

Since the effect is assessed in terms of total ammonia, these un-ionised ammonia concentrations need to be transformed into Total Ammoniacal Concentration (TAN) values using the ammonia equilibrium equation. However, the equilibrium constant is dependent on the temperature. The temperature was set at 11°C. This is the average river water temperature in the Bautersembeek.

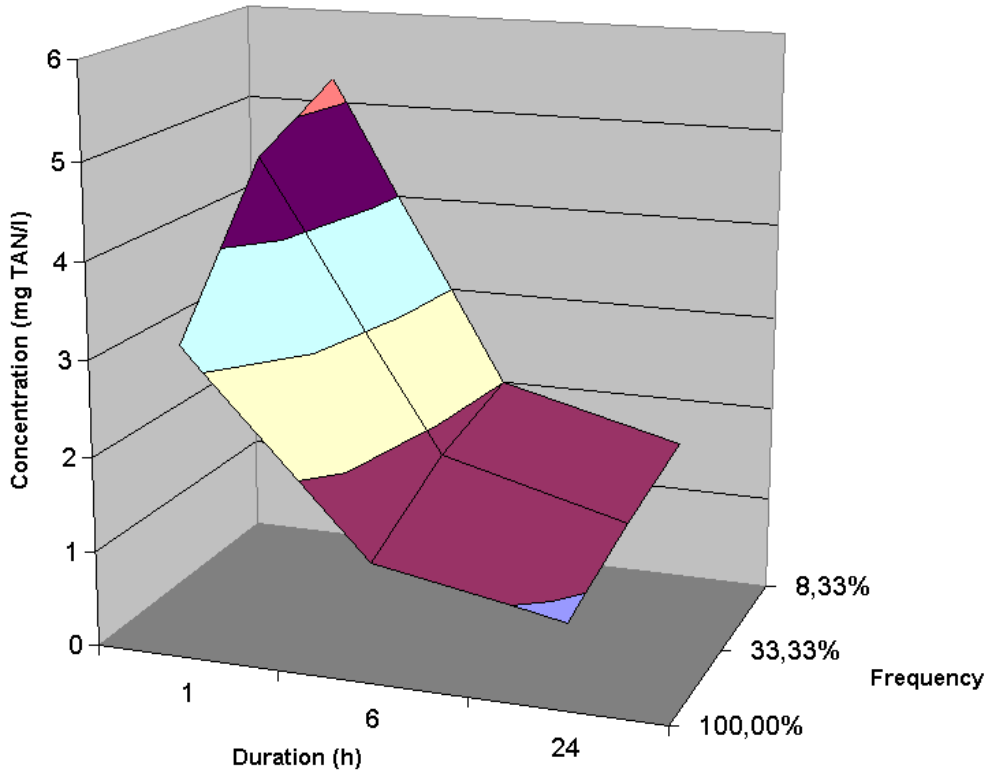


Figure 9: Effect concentration-duration-frequency surface for sustainable salmonid fishery

Equivalent to the exposure assessment, two-dimensional duration-frequency curves were derived from this three-dimensional effect concentration-duration-frequency surface. The concentration was, as in the exposure concentration-duration-frequency determination, set at 4 TAN mg/l (= 0.08 mg NH_3-N/l). This curve is a slice from that three-dimensional plot and can be determined through linear interpolation (Figure 9). The resulting curve is shown in Figure 10. Note that no 90%-

uncertainty band could be determined because of an insufficient number of data points. Point B can be interpreted as follows: there are no adverse effects for salmonid species if the total ammonia concentration of 4 mg TAN /l is exceeded with a frequency of less than 33.3% and with a duration of 2.25 hours or longer.

5.2.3.4.3. Time-Referenced Risk Assessment

Finally, the exposure and effect concentration-duration-frequency curves are overlaid on the same graph in Figure 10.

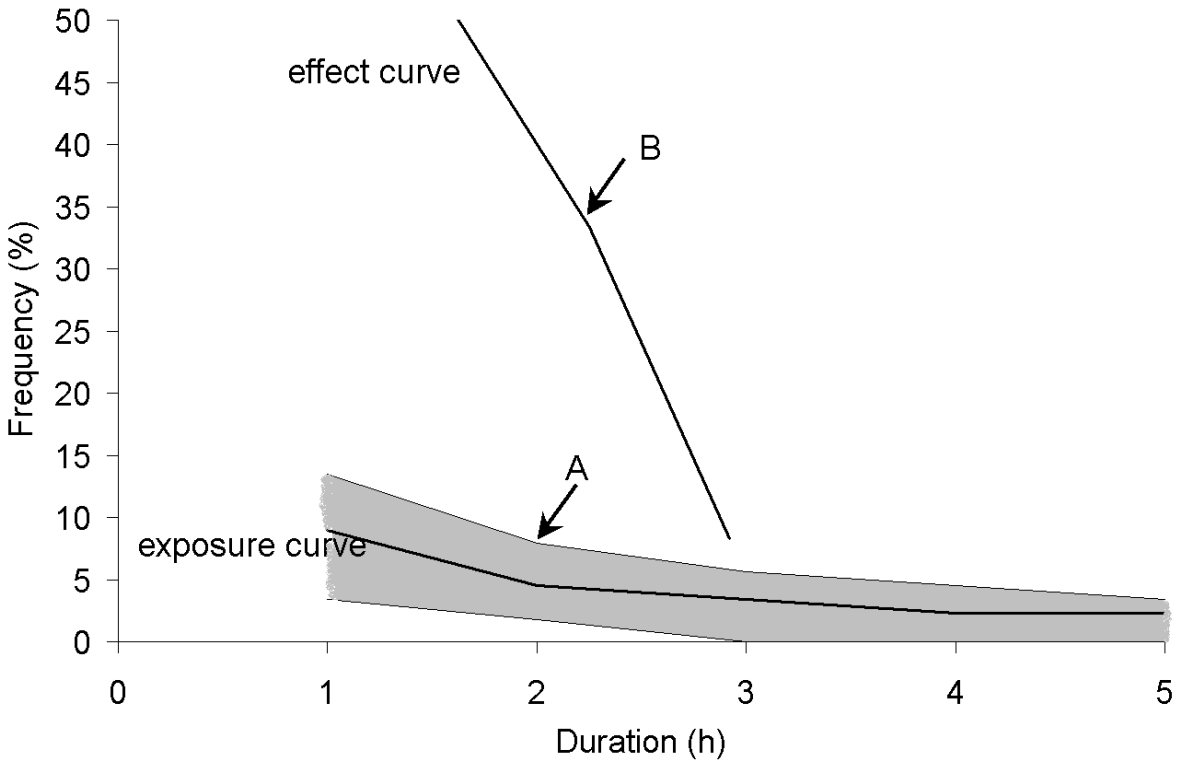


Figure 10: Concentration-duration-frequency exposure and effects curves (the exposure curve also has an 90%-uncertainty band)

This analysis shows that there is no risk involved in the situations of exceedances shorter than 3 hours. The salmonid species can even stand more frequent exceedances (again for durations shorter than 3 hours). This conclusion is very useful for the operation of the WWTP. It stimulates the operator to fix a problem at the plant within three hours. Or more general, short-lasting calamities that occur for example daily are less important in comparison with long-lasting calamities that occur only once. Unfortunately, no assessment can be made for durations longer than 3 hours because no

ecotoxicity data are available. Such situations could occur in cases of equipment failure that take relatively long to repair.

This time-referenced method is expected to be more realistic and refined compared to the probabilistic method. However, this assessment only partly confirms this. The presented two-dimensional analysis insufficiently supports this statement because the critical concentration was only set at one value (namely 4 mg TAN/l). More concentrations and their corresponding duration-frequency curves should be constructed. This could not be accomplished due to a lack of more effect concentration-duration-frequency data and simulations. Figure 11 shows how the exposure and effects concentration-duration-frequency curves would shift in case the critical concentration would increase. The effect concentration-duration-frequency curve would shift to lower durations and lower frequencies if one extrapolates Figure 9. The exposure concentration-duration-frequency curve would shift to higher frequencies for short-lasting durations and to lower frequencies for long-lasting durations. This can easily be shown in the example of Figure 2 when the concentration (effluent limit) is increased for instance from 4 to 6 mg/L in Figure 2.

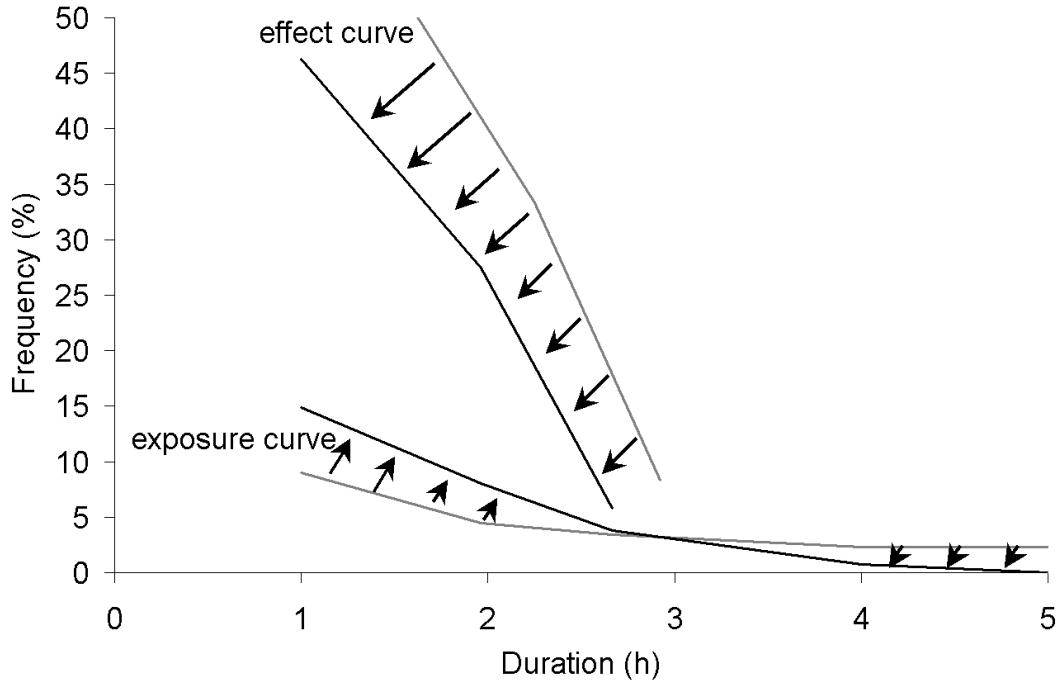


Figure 11: Shift in exposure and effect concentration-duration-frequency curves when the critical concentration increases

5.2.4. Conclusions and Further Research

The risk of a chemical was determined using two methods: a purely probabilistic method in which exposure and effects are considered as overall time-lumped probability distributions and a time-referenced method in which the duration and frequency of exceeding a standard of exposure and effects are considered. The probabilistic method resulted in a risk of 8.4% (0.3-28% is a 90% confidence interval). No quantitative risk measure could be calculated for the time-referenced method. There is no risk for exceedances lasting smaller than 3 hours at a concentration level of 4 mg TAN/l. Both methods can be used for decision-support in risk assessment practice.

The time-referenced method is expected to be more realistic and refined compared to the probabilistic method. However, this assessment only partly confirms this. The comparison of both methods by means of a case study mainly indicated that (1) further research should be undertaken to collect more data to build more extensive three-dimensional effects and exposure concentration-duration-frequency surfaces. (2) The three-dimensional concentration-duration-frequency surfaces should be compared and assessed (instead of the duration-frequency curves alone). This is illustrated in a fictitious example of Figure 12. In this way, the magnitude of an exceedance would also be considered. (3) There is also a need for quantitative measures to characterise time-referenced risks from such three-dimensional concentration-duration-frequency surfaces.

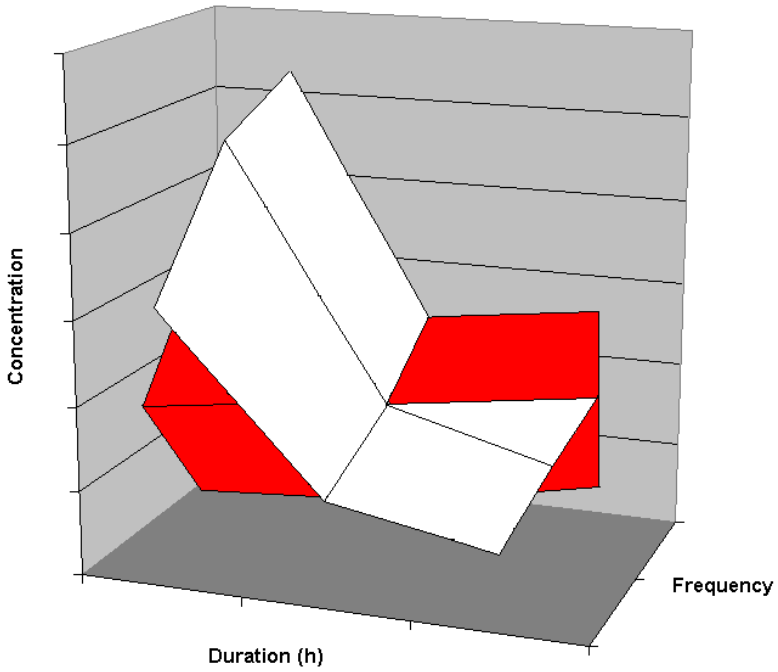


Figure 12:Fictitious example of exposure and effects concentration-duration-frequency surfaces

Part 6

-

General Conclusions & Further Research

Part 6

General Conclusions & Further Research

Environmental pollution due to release of chemicals has led governments to develop new laws and regulation that puts constraints on chemical emissions. These are based on environmental quality standards and environmental/ecological risk assessment. The key question to be answered is: "What is the likelihood (i.e. probability) of adverse effects occurring to exposed ecological systems due to exceedance of a toxicity level by an environmental concentration?". This dissertation studied and developed a range of statistical techniques needed to answer this question with a risk probability and an uncertainty or confidence interval rather than with the current black white "yes, maybe / no" answer which the conventional risk assessments provide. For this, a trade-off had continuously to be made between accurate, good but usually more complex statistical techniques and easy-to-use, -understand and -apply but usually less good statistical techniques.

After all, in a Probabilistic Ecological Risk Assessment (PERA), the exposure concentration and species sensitivity are treated as random variables taken from probability distributions (respectively Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD)) which are combined to give a risk probability. In environmental quality standard setting, a 5^{th} percentile is calculated from the SSD. This concentration is also known as the HC_5 (Hazardous Concentration at 5%).

Next, the conclusions and further research topics are discussed in more detail in two sections: PERA and geo- (and time-)referencing of PERA.

6.1. Probabilistic Ecological Risk Assessment Framework

A PERA framework was proposed and is repeated in Figure 1. Two different approaches can be used to determine the ECD and the SSD. Data from either measurements in the environment or toxicity tests can be used directly (see Figure 1, right side). The alternative is to use prediction or extrapolation models, especially in case of new chemicals (see Figure 1, left side). In practice, exposure models are more common for ECD determination and effect data are more common for SSD determination. Obviously, this may shift in the future.

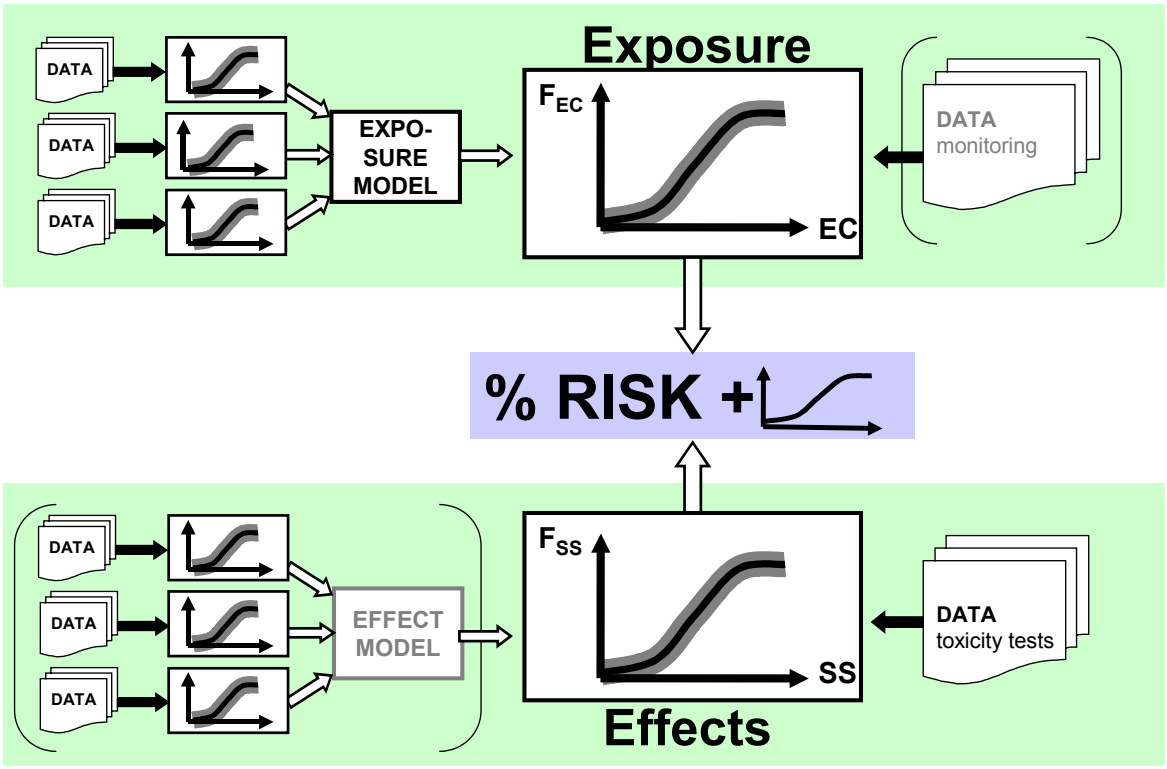


Figure 1:Proposed framework of Probabilistic Ecological Risk Assessment (PERA) with EC: Exposure Concentration and SS: Species Sensitivity

Several statistical methods are needed to obtain a probabilistic risk. A distinction was made between methods for:

- Propagating uncertainty and variability through mathematical models (open arrows in Figure 1, used in the effects and exposure modelling but also in the risk characterisation): Monte Carlo simulation, first order analysis, probability bounds analysis or analytical techniques.
- Estimating data uncertainty and variability (full arrows in Figure 1, used in the ECD and SSD estimation based on measured data but also for estimating variability and uncertainty of input parameters and variables needed for exposure and effect modelling): Bootstrap, methods from classical and Bayesian statistics, maximum likelihood estimation.

Mainly statistical weaknesses of the current approaches in PERA were addressed in this dissertation. An overview of the solutions provided is given below.

6.1.1. Correct Application of Probabilistic Methods

It was shown that some existing probabilistic methods were applied wrongly or insufficiently correct. In a bootstrap analysis, the resample size should not exceed the sample size. No theoretical background was found that supports this and simulation examples showed that when the resample size exceeds the sample size logical and statistical inconsistencies arise.
In Monte Carlo analysis, any correlations between the inputs should be considered. Vose's (1996) 'cardinal rule of risk analysis modelling' is "Every iteration of a risk analysis model must be a scenario that could physically occur". If uncertainty and/or variability is considered in a Monte Carlo analysis, probability distributions should be assigned to all relevant input variables and parameters, or at least to the most sensitive ones. If not, the output distribution will be more difficult to interpret.

6.1.2. Importance of Interpretation

It was shown that interpretation of all probability distributions in a PERA framework should be made carefully. Insufficient or wrong interpretations often lead to confusion, misleading results, incomprehension and make PERA an unattractive technique for policy-makers.

In Monte Carlo simulations, separation of uncertainty and variability and the correct application of Monte Carlo analysis simplify the interpretation of a model's output distribution of interest. A case study showed that the NOEC (No Observed Effect Concentration) of Cu will be larger than 75 μg/l for 80% of the time for a lake in Sweden. This is a quite different result than being 80% certain that the NOEC will be larger than 75 μg/l. A first order or one-dimensional Monte Carlo simulation can only propagate variability or uncertainty, but not both at the same time without having difficulties with interpreting the output distribution. For this, a second order or two-dimensional Monte Carlo simulation is needed.

A probabilistic risk should be interpreted as the probability that a random selected exposure or environmental concentration will exceed a species sensitivity. Examples have shown that the same risk probability can represent different environmental conditions (e.g. depending on whether the ECD represents spatial or temporal variability). Therefore, it is suggested to always include as much information as possible in the answer to the key question described above: indicate what type of variability the ECD or the SSD represents (geo- or time-referenced), what endpoint was used … For example, there is 14% probability that the environmental/exposure concentration of atrazine

randomly selected from a day in the year at the "Westsluisbeek" in Alveringem will exceed the chronic threshold NOEC of a randomly selected aquatic species. A 90% uncertainty or confidence interval for this risk is 5-30%.

6.1.3. Comparison and Validity of Several Probabilistic Estimation Techniques

The reliability of several uncertainty and variability estimation methods at sample size 20 and smaller was compared and assessed by means of simulation and case studies. The considered methods display varying robustness and accuracy in determining lower confidence limits of the HC_5. Differences between methods are for a large part determined by the choice of the probability distribution: parametric or nonparametric, threshold or non-threshold distributions (see 6.1.4). The most suitable methods to estimate lower end percentiles such as the 5^{th}-percentile were found to be the parametric approaches from classical and Bayesian statistics, and nonparametric bootstrapping (using the interpolated empirical distribution function and the Hazen plotting system).

Some nonparametric methods should not be used for estimating low percentiles given a small sample size. All resampling techniques (basic bootstrap) showed they were rather arbitrary and inaccurate because they are bounded by the smallest data point.

For estimating 5^{th}-percentiles of small sample sizes, the Hazen plotting and the mean plotting are both used in literature but one should be aware that both plottings give different results (a factor of 2 was observed here) at low sample sizes.

Further research on robust estimation techniques will be needed to reduce the influence of outliers as these were observed to have a large influence on the HC_5 (in the case studies, a factor of 4 and 7 was observed for respectively parametric and nonparametric methods).
And more research is also needed to deal with spatial and temporal autocorrelations of environmental concentration observations in uncertainty and variability estimation techniques.

6.1.4. Parametric (Threshold or Non-Threshold) or Nonparametric Methods

Throughout this dissertation, parametric and nonparametric methods were often used in parallel. Results are very sensitive to the choice of the method (for the HC_5 estimation, a factor of 5 difference was observed when results from different methods were compared). Consequently, the importance of a proper use of distribution selection methods should not be underestimated. Statistical tests, graphical exploration and expert knowledge can help in identifying the appropriate distribution.

Parametric methods assume that the data come from a fixed form underlying distribution. This assumption enables them to work with smaller sample sizes. Nonparametric methods rely on the data points themselves. This makes them less vulnerable to deviations from certain distribution assumptions but more vulnerable to deviations in the data points since for calculating HC_5, the nonparametric methods are almost insensitive to upper outliers and very sensitive to lower outliers. For calculating HC_5, parametric methods are more sensitive to upper outliers but less sensitive to lower outliers compared with the nonparametric methods.

As described above, the two approaches have their advantages and disadvantages and their use depends on the expert's opinion, the problem formulation, the goals and the sample size. For calculating the HC_5, it was found that preference should be given to parametric methods when the sample size is below 10 and preference should be given to nonparametric methods when the sample size is very large (e.g. 50). For the intermediate sample sizes, either parametric or nonparametric techniques can be used or maybe a combination of the two could be used as demonstrated in Figure 2. The 5^{th} percentile could be calculated as a weighted sum of the 5^{th} percentiles calculated parametrically and nonparametrically. The weights can be assigned based on expert knowledge or as demonstrated in Figure 2.

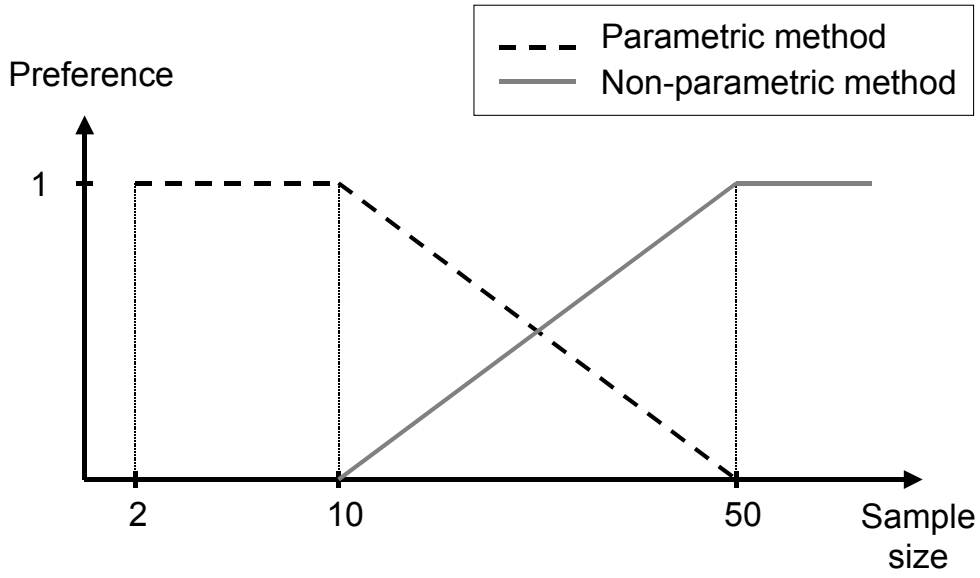


Figure 2: Potential user's preference to parametric or nonparametric methods based on sample size

6.1.5. 'Clarifications' of Particular Issues

Several chapters and sections in this dissertation tried to clarify particular issues that are in literature often presented incompletely or confusing. The two most important issues for this were the sample

size determination problem and how to deal with several (hierarchical) levels of variability and uncertainty.

Two important considerations were illustrated needed for sample size determination. The general methodology is first to determine the scientific reliability and accuracy of the HC_5 estimation method, representativeness of the data set and second to specify the desired level of precision needed, by the policy-maker or risk manager.

Based on simulation studies to assess the reliability and accuracy of the statistical methods for small sample sizes, it can be summarised that sample size 2 is the absolute minimum for all parametric methods and sample size 10 is the absolute minimum for the nonparametric bootstrap based on Hazen plotting and the interpolated EDF. It was also shown that the minimum sample size for nonparametric bootstrap based on EDF (resampling) is much larger than 25. The findings should be considered when the HC_5 estimation method is selected. Once an appropriate method is selected, the level of precision can be specified on an absolute or relative scale by the decision-maker and the minimum sample size can then be determined.

Several (statistical) methods were proposed to account for hierarchical variability (i.e. inter-laboratory variability, inter- and intra-species sensitivity) and hierarchical uncertainty (mainly sampling uncertainty). Their analysis by means of simulation and case studies led to three conclusions. First, of all studied techniques, the hierarchical bootstrap method was found to be the most accurate and precise method and the only method with a scientifically reliable interpretation. Second, hierarchical, non-hierarchical and conventional methods were found to produce similar results. This indicates that the conventional method does not perform as bad as one would expect based on the fact that it is ignoring underlying information. Third, the non-hierarchical method seems to be most conservative for the simulations performed here, but should not be used for confidence interval estimation. Further testing and research on (non-)hierarchical methods is however needed to generalise these conclusions.

Note that not all sources of variability and uncertainty were considered. Rather, this should be viewed as a step forward in revealing, quantifying and propagating more sources of uncertainty and variability in a scientifically defendable manner. The more sources are quantified; the better the decision-maker can assess the reliability of a risk assessment outcome.

Among all the sources of uncertainty in risk assessment, sampling uncertainty was considered and quantified most. It might be useful to also include other sources of uncertainty in order to obtain a more complete uncertainty analysis. Two important sources were not discussed here but could be studied in further research. First, the extrapolation from laboratory to field ecotoxicity effects is currently considered in an assessment or safety factor. Either, these assessment factors could be considered as an uncertain random variable as initiated by Roelofs (2001) and Monte Carlo

simulations could propagate the uncertainty. Or, the change in location and shape or type of the SSD could be studied in more detail.
Second, the model structure uncertainty of exposure and effect models but also the distribution type uncertainty of the SS or EC was not quantified and assessed. There are currently no statistical techniques that are able to quantify model uncertainty. Typically, validation studies or quantification of model complexities can help comparing several models but no absolute measures exist that would allow the integration of the model uncertainty with, for example, the parameter uncertainty into a total output uncertainty interval.

6.1.6. Probabilistic Ecological Risk Characterisation

The probabilistic risk quotient method is a probabilistic extension of the well-known and familiar deterministic risk quotient and is capable of estimating a probabilistic risk, defined as the probability of a randomly selected EC exceeding a randomly selected SS (i.e. a probability instead of a ratio >1/<1) and is in addition, capable of estimating an uncertainty interval representing the sampling error that exists because of the practical inability to collect an infinite number of data. The method can handle all types of parametric and nonparametric distributions and is easy to use and interpret at the same time.

Several examples and case studies have proven that the probabilistic risk characterisation considers the quantitative information of the full range of the ECD and SSD (including lower SS than HC_5 and higher ECs than the 90^{th} percentile) instead of only considering the upper tail of the ECD and the lower tail of the SSD. Consequently, several issues on calculating tail percentiles (such as HC_5) can be omitted because they are no longer needed in the risk characterisation.

The current risk measures, such as the Area Under The Curve (AUC) of a Joint Probability Curve (JPC), contain insufficient information to account for different environmental circumstances (i.e. different interpretations of the ECD and SSD) and to assess potential adverse effects on ecological communities. Therefore, it is recommended to always interpret the risk ecologically. This will force the environmental community to compare SSDs with adequate ECDs.
Further research is needed on measures additional to the calculated risk that characterise the shape of the JPC and that have an environmental interpretation (depending on the interpretation of the EC and SS) to help to quantify and manage the risk. Only a first proposal was made for a risk shape parameter based on the ratio of the coordinates of the gravity point of the JPC.

6.2. Geo- (and Time-)Referencing of Probabilistic Ecological Risk Assessment

The PERA can be considered as a part of a tiered framework. Figure 3 shows an overview of several tiers (of different level of detail) of PERA. PERA (represented in the second panel) is an extension of the conventional ERA (in top panel) since both the inherent variability and uncertainty (shown as a grey band) is explicitly quantified and assessed. However, all sources of variability are eventually lumped in one single distribution.

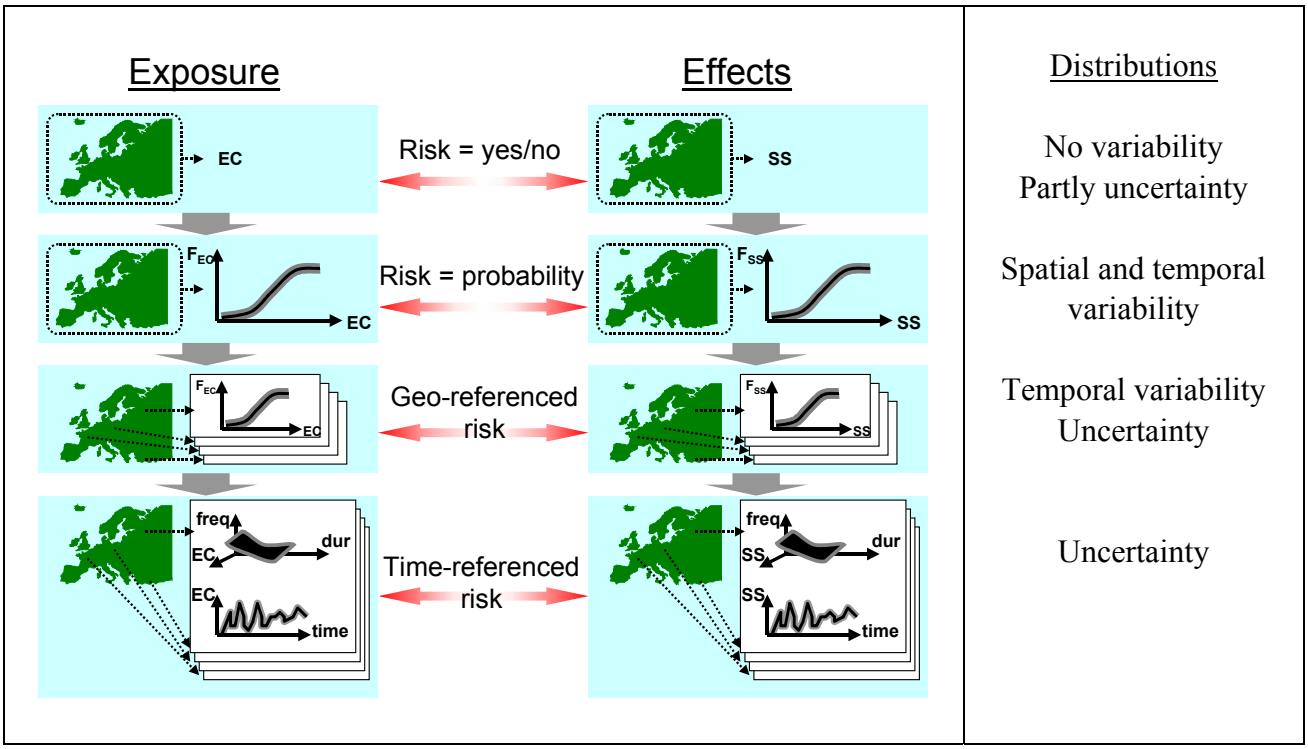


Figure 3: Different tiers in ERA (Ecological Risk Assessment), (EC: Exposure Concentration, SS: Species Sensitivity)

A geo-referenced PERA framework was developed and illustrated with case studies (represented in the third panel of Figure 3). Geo-referencing makes the risk assessment more realistic as spatial variability and dependencies of the EC and SS are explicitly accounted for, i.e. less spatial variability is lumped in the probabilistic ecological risk assessment and therefore, this is useful for assessing risk of individual chemicals. Probabilistic risk assessment therefore delivers a more transparent, realistic and non-conservative approach to estimate risks.

In addition, it was highlighted that geo-referencing effects (SSD) is now still a largely unexplored area but that it has important potential to further improve probabilistic ecological risk assessments. Geo-referencing PNECs (Predicted No Effect Concentration) and/or SSDs (in a GIS environment)

and incorporating different levels of uncertainty results in a more realistic risk assessment which is preferable to the current practice of using a single (lumped) PNEC or SSD, representing an entire region or country.

Time-referencing could further increase the level of detail and realism, as time-specific information would be accounted for. This is represented in the lower panel of Figure 3. Time related information can be formatted in two ways in an attempt to capture the temporal variability. First, time-series can be used as such or second, time-series can be translated into concentration-duration-frequency surfaces. These surfaces are three-dimensional plots with on the three axes the concentration, the duration of an exceedance above a particular concentration and the frequency of an exceedance above a particular concentration with a particular duration. However, the case study (based on duration-frequency curves) here only partly confirms the improvement. Further research should be undertaken to collect more data to build more extensive three-dimensional effects and exposure concentration-duration-frequency surfaces. The three-dimensional should be compared and assessed (instead of the duration-frequency curves alone as in this dissertation). In this way, the magnitude of an exceedance would also be considered. There is also a need for quantitative measures to characterise time-referenced risks.

In the presented tiered framework and many modelling studies, it is always important to balance the model complexity and model accuracy. Figure 4 shows the conceptual relationship between the accuracy and the model/framework complexity. The more complex the model, the more accurate predictions will be up to a certain complexity. After that, the accuracy does not increase anymore with increasing complexity. It remains constant or may even decrease. Three (or four) zones can be distinguished depending on the slope or rate of increase in realism versus increase in complexity. In zone I, this slope is high. In zone II, this is slope is still positive although smaller compared to zone I. In zone III, the slope is zero (in a fourth zone, the rate may be negative).

In order to assess the added-value of this work, all tiers were added to the conceptual relationship in Figure 4. The PERA was proven to be more complex but also more realistic and accurate compared to the conventional, deterministic risk assessment (in one of the case studies, conventional ERA resulted in risk whereas PERA resulted in negligible risk). The geo-referenced PERA was also proven to be more complex and realistic compared to the non-geo-referenced PERA although the slope is smaller (in one case study, risk dropped by a factor 3). A time-referenced PERA could not be situated exactly in this conceptual relationship. Further research will have to show whether the increase in accuracy is worthwhile compared to the increase in complexity.

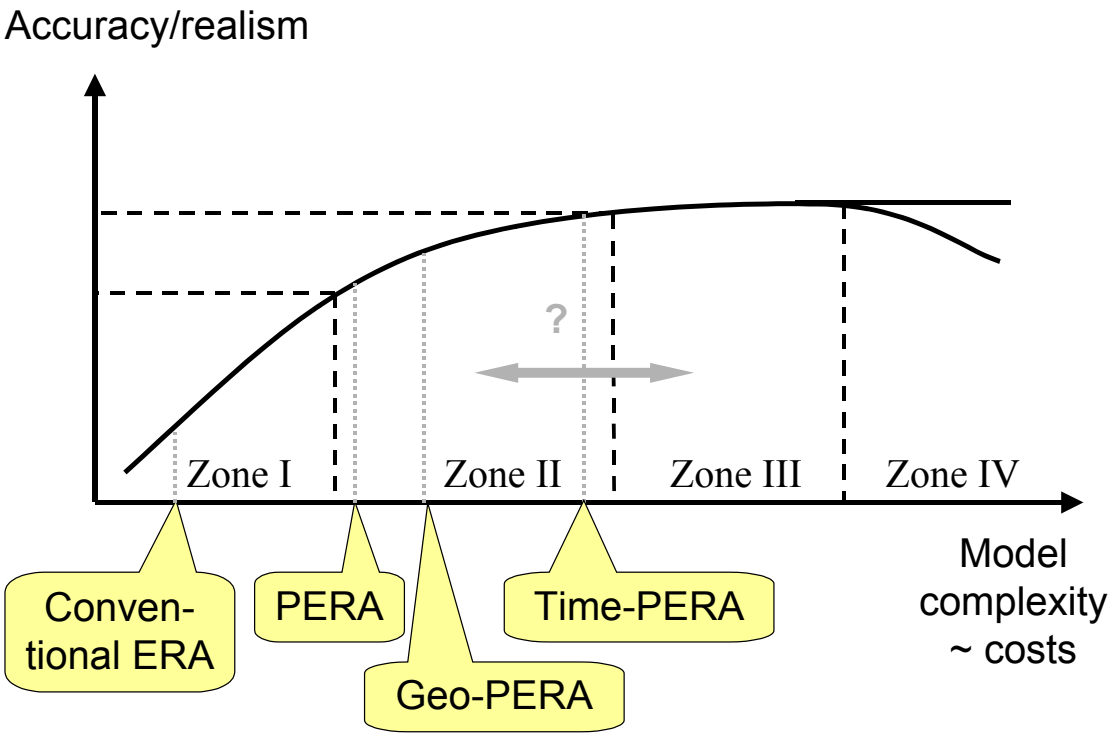


Figure 4: Conceptual complexity-accuracy relationship and the location of the conventional ERA (Risk Assessment), PERA (Probabilistic Ecological Risk Assessment), Geo-PERA and Time-PERA

6.3. Validation

The proposed geo-referenced PERA can not be validated because there exists no 'instrument' that can measure probabilistic risk. Note that the same can be said about the conventional, deterministic risk assessment. However, it is possible to validate components/steps of the geo-referenced PERA. The ECD, for example, could be validated. But since the main goal of this thesis is to evaluate statistical techniques, only the techniques were validated by means of several simulation studies. The underlying risk assessments principles as for example, using the SSD as a way to set environmental standards and protect ecological communities is not studied here.

6.4. Concluding Thought

"The greatest risk in life is to not take any risk at all". The time has come that policy-makers realise that chemicals will always induce a certain risk as probabilistic ecological risk assessments tells this to the policy-makers in contrast to the current conventional methods, which conceal this because they only provide a "yes/no" risk answer. The key question for the policy maker now becomes: "what level of risk is acceptable?". This more complex assessment will not simplify the decision but at least, it will be closer to reality.

Bibliography

Aldenberg T. & Jaworska J. S. 2000. Estimation of the hazardous concentration and fraction affected for normally distributed species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46, 1-18.

Aldenberg T., Jaworska J. S. & Traas T. P. 2002. Normal species sensitivity distributions and probabilistic ecological risk assessment. Pages 49-102 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers, London, UK.

Aldenberg T. & Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, 25, 48-63.

Asante-Duah K. 1998. Risk management in environmental management. A guide for managing chemical contamination problems. John Wiley & Sons, New York, USA. 515.

Baker J. L. 1994. Aquatic dialogue group: pesticide risk assessment and mitigation. Society of Environmental Toxicology and Chemistry, Pensacola, USA.

Bauwens W., Vanrolleghem P. A. & Smeets M. 1996. An evaluation of the efficiency of the combined sewer - wastewater treatment system under transient conditions. *Water Science & Technology*, 33(2), 199-208.

Berthouex P. W. & Brown L. C. 1994. Statistics for environmental engineers. Lewis publishers.

Bickel P. J. & Freedman D. A. 1981. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6), 1196-1217.

Bickel P. J., Gotze F. & van Zwet W. R. 1997. Resampling fewer than n observations: gains, losses and remedies for losses. *Statistica Sinica*, 7, 1-31.

Bixio D., Parmentier G., Boonen I. & Thoeye C. 2001a. Renovation of the WWTP Hove. *Med. Fac. Landbouww. Univ. Gent*, 66(3a), 103-110.

Bixio D., Parmentier G., Rousseau D., Verdonck F. A. M., Meirlaen J., Vanrolleghem P. A. & Thoeye C. 2001b. Intergrating risk analysis in the design/simulation of activated sludge systems. *Proceedings 74th Annual Conference and Exposition of Water Environment Federation (WEFTEC)*. Atlanta, VS. 14-18 October 2001.

Bixio D., Parmentier G., Rousseau D., Verdonck F. A. M., Meirlaen J., Vanrolleghem P. A. & Thoeye C. 2001c. A quantitative risk analysis tool for design/simulation of wastewater treatment plants. *2nd IWA World Water Congress*. Berlin, Germany. 15-19 October 2001.

Bixio D., Rousseau D., Verdonck F. A. M., Meirlaen J., Thoeye C. & Vanrolleghem P. A. 2002. A quantitative risk analysis tool for design/simulation of wastewater treatment plants. *Water Science & Technology*, 46(4-5), 301-307.

Blom G. 1958. Statistical estimates and transformed beta-variables. John Wiley & sons, New York, USA.

Boeije G. 1999. Chemical fate prediction for use in geo-referenced environmental exposure assessment. PhD. Ghent University, Gent. 225

Boeije G., Vanrolleghem P. A. & Matthies M. 1997. A geo-referenced aquatic exposure prediction methodology for 'down-the-drain' chemicals. *Water Science & Technology*, 36(5), 251-258.

Boeije G. M., Wagner J.-O., Koormann F., Vanrolleghem P. A., Schowanek D. & Feijtel T. C. J. 2000. New PEC definitions for river basins applicable to GIS-based environmental exposure assessment. Contribution to GREAT-ER #8. *Chemosphere*, 40, 255-265.

Bonnomet V., Duboudin C., Magaud H., Thybaud E., Vindimian E. & Beauzamy B. 2002. Modeling explicitely and mechanistically median lethal concentration as a function of time for risk assessment. *Environmental Toxicology and Chemistry*, 21(10), 2252-2259.

Box G. E. P. & Tiao G. C. 1973. Bayesian inference in statistical analysis. Addison Wesley (1992, Wily Classics, New York).

Bros W. E. & Cowell B. C. 1987. A technique for optimizing sample size (replication). *J. Exp. Mar. Biol. Ecol.*, 114, 63-71.

Burmaster D. E. 1997. Technical and policy memorandum. Computations illustrating variability and uncertainty. Alceon corporation, Cambrige.

Burmaster D. E. & Anderson P. D. 1994. Priniciples of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Analysis*, 14(4), 477-481.

Burmaster D. E. & Bloomfield L. R. 1996. Mathematical properties of the risk equation when variability is present. *Human and Ecological Risk Assessment*, 2(2), 348-355.

Burmaster D. E. & Wilson A. M. 1996. An introduction to second-order random variables in human health risk assessments. *Human and Ecological Risk Assessment*, 2(4), 892-919.

Campbell K. R., Bartell S. M. & Shaw J. L. 2000. Characterising aquatic ecological risks from pesticides using a diquat dibromide case study. II. Approaches using quotients and distributions. *Environmental Toxicology and Chemistry*, 19(3), 760-774.

Cardwell R. D., Brancato M. S., Toll J., DeForest D. & Tear L. 1999. Aquatic ecological risks posed by tributyltin in United States surface waters: pre-1989 to 1996 data. *Environmental Toxicology and Chemistry*, 18(3), 567-577.

Cardwell R. D., Parkhust B. R., Warren-Hicks W. & Volosin J. S. 1993. Aquatic ecological risk. *Water Environment and Technology*, 18, 567-577.

Cowan C. E., Versteeg D. J., Larson R. J. & Kloepper-Sams. 1995. Integrated approach for environmental assessment of new and existing substances. *Regulatory Toxicology and Pharmacology*, 21, 3-31.

Cox C. D. 2000. Integration of uncertainty and water quality considerations in the design of activated sludge plants. *WEFTEC 2000*.

Crane M. & Newman M. C. 2000. What level of effect is a No Observed Effect? *Environmental Toxicology and Chemistry*, 19, 516-519.

Cullen A. C. & Frey H. C. 1999. Probabilistic techniques in exposure assessment. A handbook for dealing with variability and uncertainty in models and inputs. Plenum, New York. 335.

Davison A. C. & Hinkley D. V. 1997. Bootstrap methods and their application. Cambridge University Press. 582.

De Laender F. 2003. Geografisch gerefereerde chemicalien toxiciteit. MSc Thesis. (in dutch). Ghent University, Gent.

De Pauw N., Roels D. & Fontoura A. P. 1986. Use of artificial substrates for standardized sampling of macroinvertebrates in the assessment of water quality by the Belgian Biotic Index. *Hydrobiologia*, 100, 153-168.

De Schamphelaere K., Heijerick D. G. & Janssen C. R. 2002. Refinement and field validation of a biotic ligand model predicting acute copper toxicity to *Daphnia magna*. *Comparative Biochemistry and Physiology Part C*, 133, 243-258.

Deksissa T. & Vanrolleghem P. A. 2001. Dynamic exposure assessment and modelling. *Med. Fac. Landbouww. Univ. Gent*, 66(4), 239-244.

Di Mauro C., Hoogerwerf M. & Sinke A. J. C. 2000. A GIS-based risk assessment model: application on sites contaminated by chlorinated solvents and petroleum hydrocarbons (BTEX). *Proceedings 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4)*. Banff, Alberta, Canada. September 2 - 8, 2000.

Di Toro D. M., Allen H. E., Bergman H. L., Meyer J. S., Paquin P. R. & Santore R. C. 2001. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. *Environmental Toxicology and Chemistry*, 20(10), 2383-2396.

Dunn A. J., Frodsham D. A. & Kilroy R. V. 1998. Predicting the risk to consent compliance of new sewage treatment works. *Water Science & Technology*, 38(3), 7–14.

Duvall S. E. & Barron M. G. 2000. A screening level probabilistic risk assessment of mercury in Florida Everglades food webs. *Ecotoxicology and Environmental Safety*, 47, 298-305.

ECOFRAM. 1999. ECOFRAM Aquatic and Terrestrial Final Draft Reports. US Environmental Protection Agency.

EEC. 1996. Technical Guidance Document in Support of Commission Directive 96/67/EEC on Risk Assessment of New Notified Substances and Regulation (EC) no. 1488/94 on Risk Assessment of Existing Substance (Parts I, II, III and IV). EC Catalogue Nos. CR-48-96-001, 002, 003, 004-EN-C, Office for official publications of the European Community, Luxembourg, Luxembourg.

Efron B. 1982. The jackknife, the bootstrap and other resampling plans, Philadelphia, Pennsylvania.

Efron B. & Tibshirani R. J. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

EPA. 1997. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001, Environmental protection Agency.

EPA. 1999a. Report of the workshop on selecting input distributions for probabilistic assessments. EPA/630/R-98/004, U.S. Environmental Protection Agency, Washington, DC.

EPA. 1999b. Update of ambient water quality criteria for ammonia. EPA-822-R-99-014, United States Environmental Protection Agency, Washington, D.C., USA.

EPA. 2001. Risk assessment guidance for superfund (RAGS): Volume III - Part A. Process for conducting probabilistic risk assessment. US Environmental Protection Agency.

Erickson R. J. & Stephan C. E. 1988. Calculation of the final acute value for water quality criteria for aquatic organisms. PB88-214994, National technical information service, Springfield, VA, USA.

EUSES. 1997. European Uniform system for the evaluation of substances (EUSES). *in*. European Chemical Bureau, Ispra, Italy.

EU-TGD. 1995. Environmental risk assessment of new and existing substances. Technical Guidance Document., Brussels, Belgium.

EU-TGD. 1996. Technical guidance document in support of commission directive 93/67/EEC on risk assessment for new notified substances and commission regulation (EC) N° 1488/94 on risk assessment for existing substances., Brussels.

Fedra K. 1998. Integrated risk assessment and management: overview and state of the art. *Journal of Hazardous Materials*, 61(5-22).

Feijtel T., Boeije G., Matthies M., Young A., Morris G., Gandolfi C., Hansen B., Fox K., Holt M., Koch V., Schröder R., Cassini G., Schowanek D., Rosemblom J. & Niessen H. 1997. Development of a geography-referenced regional exposure assessment tool for European rivers - GREAT-ER. *Chemosphere*, 34(11), 2351-2373.

Ferson S., Moore D. R. J., Estes T. L., O'Connor R. J., Van den Brink P. & Verdonck F. A. M. in press. Bounding uncertainty analysis. *in* A. Hart, editor. Application of uncertainty analysis to ecological risks of pesticides. SETAC, Pensacola, FL, USA.

Forbes V. E., Calow P. & Sibly R. M. 2001. Are current species extrapolation models a good basis for ecological risk assessment? *Environmental Toxicology and Chemistry*, 20(2), 442-447.

Foster G. D., Lippa K. A. & Miller C. V. 2000. Seasonal concentrations of organic contaminants at the fall line of the Susquehanna River basin and estimated fluxes to Northern Chesapeake bay, USA. *Environmental Toxicology and Chemistry*, 19(4), 992-1001.

Fox B. L. 1999. Strategies for Quasi-Monte Carlo. Kluwer Academic Publishers, Boston. 392.

Frey H. C. & Burmaster D. E. 1997. Methods for characterizing variability and uncertainty: comparison of bootstrap simulation and likelihood-based approaches..

Frey H. C. & Li S. 2001. Quantification of variability and uncertainty in stationary gas-fueled internal combustion engine NOx and total organic compounds emission factors. *Proceedings, Annual Meeting of the Air & Waste Management Association*. Orlando, Florida, USA.

Frey H. C. & Rhodes D. S. 1998. Characterization and simulation of uncertain frequency distributions: Effects of distribution choice, variability, uncertainty, and parameter dependence. *Human and Ecological Risk Assessment*, 4(2), 423-468.

Fuchsman P. C., Duda D. J. & Barber T. R. 1999. A model to predict threshold concentrations for toxic effects of chlorinated benzenes in sediment. *Environmental Toxicology and Chemistry*, 18(9), 2100-2109.

FWR. 1998. Urban Pollution Management Manual - 2nd edition. FR/CL0009, Foundation for Water Research, Marlow, Bucks, UK.

Giddings J. M., Hall L. W. & Solomon K. R. 2000. Ecological risks of diazinon from agricultural use in the Sacramento-San Joaquin river basins, California. *Risk Analysis*, 20(5), 545-571.

Giesy J. P., Solomon K., Coats J. R., Dixon K. R., Giddings J. M. & Kenaga E. E. 1999. Chlorpyrifos, ecological risk assessment in North American aquatic environments. *Reviews in Environmental Contamination and Toxicology*, 160, 1-129.

Giri B. S., Karimi I. A. & Ray M. B. 2001. Modeling and Monte Carlo simulation of TCDD transport in a river. *Water Research*, 35(5), 1263-1279.

Girling A. E., Tattersfield L., Mitchell G. C., Crossland N. O., Pascoe D., Blockwell S. J., Maund S. J., Taylor E. J., Wenzel A., Janssen C. R. & Jüttner I. 2000. Derivation of Predicted No-Effect Concentrations for Lindane, 3,4-Dichloroaniline, Atrazine and Copper. *Ecotoxicology and Environmental Safety*, 46, 148-162.

Govaerts B., Beck B., Lecoutre E., le Bailly C. & Vanden Eeckhaut P. 2001. From monitoring data to regional distributions: a practical methodology applied to water risk assessment. 0127, Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Grist E. P. M., Leung K. M. Y., Wheeler J. R. & Crane M. 2002. Better bootstrap estimation of hazardous concentration thresholds for aquatic assemblages. *Environmental Toxicology and Chemistry*, 21(7), 1515-1524.

Grum M. & Aalderink R. H. 1999. Uncertainty in return period analysis of combined sewer overflow effects using embedded Monte Carlo simulations. *Water Science & Technology*, 39(4), 233-240.

Gumbel E. J. 1958. Statistics of extremes. Wiley Classics Library, John Wiley and Sons, New York, USA.

Haas C. N. & Trussell R. R. 1998. Frameworks for assessing reliability of multiple, independent barriers in potable water reuse. *Water Science & Technology*, 38(6), 1-8.

Hammersley J. M. & Morton K. W. 1964. Monte Carlo methods. Chapman & Hall, New York, USA.

Hart A. 2001. Probabilistic risk assessment for pesticides in Europe: implementation & research needs. *European workshop on probabilistic risk assessment for the environmental impacts of plant protection products*. The Netherlands.

Havens P. L., Cryer S. A. & Rolston L. J. 1998. Tiered aquatic risk refinement: case study - at-plant applications of granular chlorpyrifos to corn. *Environmental Toxicology and Chemistry*, 17(7), 1313-1322.

Hazen A. 1914. Storage to be provided in impounding reservoirs for municipal water supply. *Transactions of the American society of civil engineers*, 77, 1539-1640.

Heijerick D. G., De Schamphelaere K. A. C. & Janssen C. R. 2002. Biotic ligand model development predicting Zn toxicity to the alga *Pseudokirchneriella subcapitata*: possibilities and limitations. *Comparative Biochemistry and Physiology Part C*, 133, 207-218.

Henze M., Grady Jr C. P. L., Gujer W., Marais G. v. R. & Matsuo T. 1987. Activated Sludge Model N° 1, IAWQ Scientific and Technical Report N° 1. IAWQ, London, Great-Britain.

Hoffman F. O. & Hammonds F. S. 1994. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*, 14(5), 707-712.

Hopkin S. P. 1993. Ecological implications of "95% protection levels" for metals in soil. *Oikos*, 66, 137-141.

Iman R. L. & Conover W. J. 1982. A distribution-free approach to inducing rank order correlation among input variables. *Commun Statist-Simula Computa*, 11(3), 311-334.

Isnard P., Flammarion P., Roman G., Babut M., Bastien P., Bintein S., Esserméant L., Férard J. F., Gallotti-Schmitt S., Saouter E., Saroli M., Thiébaud H., Tomassone R. & Vindimian E. 2001. Statistical analysis of regulatory ecotoxicity tests. *Chemosphere*, 45, 659-669.

Jager T., Vermeire T. G., Rikken M. G. J. & van der Poel P. 2001. Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere*, 43, 257-264.

Jagoe R. H. & Newman M. C. 1997. Bootstrap simulation of community NOEC values. *Ecotoxicology*, 6, 293-306.

Janssen C. R. 2001. Personal communication. colin.janssen@ugent.be. *in*, Gent, Belgium.

Janssen C. R., De Schamphelaere K., Heijerick D., Muyssen B., Lock K., Bossuyt B., Vangheluwe M. & Van Sprang P. 2000. Uncertainties in the environmental risk assessment of metals. *Human and Ecological Risk Assessment*, 6, 1003-1018.

Janssens P. H. M., Heuberger P. S. C. & Sanders R. 1992. UNCSAM 1.1, a software package for sensitivity and uncertainty analysis, manual. Report nr. 959101004, National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.

Karman C. C. & Reerink H. G. 1998. Dynamic assessment of the ecological risk of discharge of produced water from oil and gas producing platforms. *Journal of Hazardous Materials*, 61, 43-51.

Kooijman S. A. L. M. 1987. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Res*, 21, 269-276.

Kooijman S. A. L. M. & Bedaux J. J. M. 1996. The Analysis of Aquatic Toxicity Data. VU University Press, Amsterdam, The Netherlands.

Kops S. G. T. & Vanrolleghem P. A. 1996. An evaluation of methodologies for uncertainty analysis in biological wastewater treatment. *9th European Junior Scientist Workshop*, 6. Kilve, UK. April 11-14, 1996.

Law A. M. & Kelton W. D. 1991. Simulation modeling and analysis, Second Edition edition. McGraw-Hill, Inc., Boston, USA. 454.

Lecoutre E. 2001. Utilisation des données du monitoring de produits chimiques dans les rivières pour déterminer des concentrations régionales. UCL, Louvain-la-Neuve. 129

Lipishan B. & Lee S. 1996. The Atrazine Herbicide. *in*.
MacLeod M., Fraser A. J. & Mackay D. 2002. Evaluating and expressing the propagation of uncertainty in chemical fate and bioaccumulation models. *Environmental Toxicology and Chemistry*, 21(4), 700-709.
Mancini J. L. 1983. A method for calculating effects, on aquatic organisms, of time varying concentrations. *Water Research*, 17(10), 1355-1362.
Manly B. F. J. 1992. Bootstrapping for determining sample sizes in biological studies. *Journal of Exp. Mar. Biol. Ecol.*, 158, 189-196.
Manz M., Weissflog L. & Schüürmann G. 1999. Ecotoxicological hazard and risk assessment of heavy metal contents in agricultural soils of central Germany. *Ecotoxicology and Environmental Safety*, 42, 191-201.
Matthies M. & Berding V. 2001. Probabilistic uncertainty analysis for EUSES multimedia regional distribution model. *SETAC Europe 11th Annual Meeting*. Madrid, Spain.
Maund S. J., Travis K. Z., Hendley P., Giddings J. M. & Solomon K. R. 2001. Probabilistic risk assessment of cotton pyrethroids: V. Combining landscape-level exposures and ecotoxicological effects data to characterize risks. *Environmental Toxicology and Chemistry*, 20(3), 687-692.
Melching C. S. 1995. Reliability estimation. Pages 69-118 *in* V. P. Singh, editor. Computer models of watershed hydrology. Water Resources Publications, Littleton, Colorado, USA.
Miaoulis G. & Michener R. D. 1976. An introduction to sampling. Kendall/Hunt, Dubuque, IA, USA.
Milne I., Seager J., Mallet M. & Sims I. 2000. Effects of short-term pulsed ammonia exposure on fish. *Environmental Toxicology and Chemistry*, 19(12), 2929-2936.
Milton A. & Stegun I. A. 1970. Handbook of mathematical functions. Dover, New York. 947.
Moore D. R. J., Sample B. E., Suter G. W., Parkhust B. R. & Teed R. S. 1999a. A probabilistic risk assessment of the effects of methylmercury and PCBs on Mink and Kingfishers along east Fork Poplar Creek, Oak Ridge, Tennessee, USA. *Environmental Toxicology and Chemistry*, 18(12), 2941-2953.
Moore D. R. J., Sample B. E., Suter G. W., Parkhust B. R. & Teed R. S. 1999b. Risk-based decision making: the East Fork Poplar Creek case study. *Environmental Toxicology and Chemistry*, 18(12), 2954-2958.
Moore D. R. J., Warren-Hicks W., Parkhust B. R., Teed R. S., Baird R. B., Berger R., Denton D. L. & Pletl J. J. 2000. Intra- and intertreatment variability in reference toxicant tests: implications for whole effluent toxicity testing programs. *Environmental Toxicology and Chemistry*, 19(1), 105-112.
Newman M. C., Ownby D. R., Mézin L. C. A., Powell D. C., Christensen T. R. L., Lerberg S. B. & Anderson B.-A. 2000. Applying species-sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient number of species. *Environmental Toxicology and Chemistry*, 19(2), 508-515.
NIST. 2001. NIST Handbook 148: Dataplot Reference Manual. Volume 2: LET subcommands & library functions. National Institute of Standards & Technology.
NRC. 1983. Risk assessment in the federal government: managing the process. National Academy Press, Washington DC.
Pawlisz A. V., Moore D. R. J. & Hart A. 2001. Application of 2nd order Monte Carlo and probability bounds analysis to estimate risk to birds exposed to chlorpyrifos in apple orchards. *SETAC Europe 11th Annual Meeting*. Madrid, Spain.

Poletika M. N., Woodburn K. B. & Henry K. S. 2002. An ecological risk assessment for chlorpyrifos in an agriculturally dominated tributary of the San Joaquin river. *Risk Analysis*, 22(2), 291-308.

Portielje R., Hvitved-Jacobsen T. & Schaarup-Jensen K. 2000. Risk analysis using stochastic reliability methods applied to two cases of deterministic water quality models. *Water Research*, 34(1), 153-170.

Posthuma L., Suter II G. W. & Traas T. P., editors. 2002. Species sensitivity distributions in ecotoxicology. Lewis publishers.

Rai S. N., Krewski D. & Bartlett S. 1996. A general framework for the analysis of uncertainty and variability in risk assessment. *Human and Ecological Risk Assessment*, 2(4), 972-989.

Regan H. M., Sample B. E. & Ferson S. 2002. Comparison of deterministic and probabilistic calculation of ecological soil screening levels. *Environmental Toxicology and Chemistry*, 21(4), 882-890.

Reichert P. 1997. On the necessity of using imprecise probabilities for modelling environmental systems. *Water Science & Technology*, 36(5), 149-157.

Reinert K. H., Giddings J. M. & Judd L. 2002. Effects analysis of time-varying or repeated exposures in aquatic ecological risk assessment of agrochemicals. *Environmental Toxicology and Chemistry*, 21(9), 1977-1992.

Ritter A. M., Shaw J. L., Williams W. M. & Travis K. Z. 2000. Characterizing aquatic ecological risks from pesticides using a diquat dibromide case study. I. Probabilistic exposure estimates. *Environmental Toxicology and Chemistry*, 19(3), 749-759.

Roelofs J. J. W. 2001. Uncertainty in the determination of ecological no-effect concentrations. Reports Environmental Studies nr. 194, Faculty of Science, Mathematics and Informatics, Nijmegen, The Netherlands.

Roman G., Isnard P. & Jouany J.-M. 1999. Critical Analysis of Methods for Assessment of Predicted No-Effect Concentration. *Ecotoxicology and Environmental Safety*, 43, 117-125.

Röpke B., Bach M. & Frede H. G. 2002. DRIPS - A DSS estimating the input quantity of pesticides in German river basins. *Proceedings iEMSs 2002, Integrated Assessment and Decision Support* 1, 150-155. Lugano, Switzerland. 24-27 June 2002.

Rousseau D., Verdonck F. A. M., Carrette R., Thoeye C., Meirlaen J. & Vanrolleghem P. A. 2000a. Risk assessment tool for the design of waste water treatment plants. *Med. Fac. Landbouww. Univ. Gent*, 65(4), 219-223.

Rousseau D., Verdonck F. A. M., De Pauw D. & Vanrolleghem P. A. 2000b. Risk assessment project: Technical manual for the RAP server 1.2 & add-in to WEST. Ghent University, BIOMATH department, Gent, Belgium.

Rousseau D., Verdonck F. A. M., Moerman O., Carrette R., Thoeye C., Meirlaen J. & Vanrolleghem P. A. 2001. Development of a risk assessment based technique for design/retrofitting of WWTPs. *Water Science & Technology*, 43(7), 287-294.

Sakov A. & Bickel P. J. 2000. An edgeworth expansion for the m out of n bootstrapped median. *Stat Prob Lett*, 49, 217-223.

Scholze M., Boedeker W., Faust M., Backhaus T., Altenburger R. & Grimme L. H. 2001. A general best-fit method for concentration-response curves and the estimation of low-effect concentrations. *Environmental Toxicology and Chemistry*, 20(2), 448-457.

Schowanek D., Fox K., Holt M., Schröder F. R., Koch V., Cassani G., Matthies M., Boeije G., Vanrolleghem P. A., Young A., Morris G., Gandolfi C. & Feijtel T. C. J. 2001. GREAT-ER: a new tool for management and risk assessment of chemicals in river basins. Contribution to GREAT-ER #10. *Water Science & Technology*, 43(2), 179-185.

Schwacke L. H., Vorr E. O., Hanssen L. J., Wells R. S., Mitchum G. B., Hohn A. A. & Fair P. A. 2002. Probabilistic risk assessment of reproductive effects of polychlorinated biphenyls on bottlenose dolphins (*Tursiops Truncatus*) from the southeast United States coast. *Environmental Toxicology and Chemistry*, 21(12), 2752-2764.

Shao Q. 2000. Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. *Environmetrics*, 11, 583-595.

Shieh J.-N., Chao M.-R. & Chen C.-Y. 2001. Statistical comparisons of the no-observed-effect concentration and the effective concentration at 10% inhibition (EC10) in algal toxicity tests. *Water Science & Technology*, 43(2), 141-146.

Smit M. G. D., Hendriks A. J., Schobben J. H. M., Karman C. C. & Schobben P. M. 2001. The variation in slope of concentration-effect relationships. *Ecotoxicology and Environmental Safety*, 48, 43-50.

Snedecor G. W. 1959. Statistical methods. The Iowa State College Press, Ames, Iowa, USA.

Solomon K. 2000. Probabilistic ecological risk assessment for pesticides: Concepts of risk - deterministic and probabilistic risk assessment. *Third SETAC World Congress*. Brighton, UK. 21 May 2000.

Solomon K., Baker D. B., Richards R. P., Dixon K. R., Klaine S. J., La Point T. W., Kendall R. J., Weisskopf C. P., Giddings J. M., Giesy J. P., Hall L. W. & Williams W. M. 1996. Ecological risk assessment of atrazine in North American surface waters. *Environmental Toxicology and Chemistry*, 15(4), 31-76.

Solomon K., Giesy J. P. & Jones P. 2000. Probabilistic risk assessment of agrochemicals in the environment. *Crop Protection*, 19, 649-655.

Solomon K. R. 1996. Overview of recent developments in ecotoxicological risk assessment. *Risk Analysis*, 16, 627-633.

Solomon K. R. & Takacs P. 2002. Probabilistic risk assessment using species sensitivity distributions. Pages 285-314 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers.

Spanjers H., Vanrolleghem P. A., Nguyen K., Vanhooren H. & Patry G. G. 1998. Towards a simulation benchmark for evaluating respirometry-based control strategies. *Water Science & Technology*, 37(12), 219-226.

Stephan C. E., Mount D. I., Hansen D. J., Gentile J. H., Chapman G. A. & Brungs W. A. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. US EPA, Duluth MN, US.

Suter II G. W. 2002. North American history of species sensitivity distributions. Pages 11-17 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers, London, UK.

Takada H., Mutoh K., Tomita N., Miyadzu T. & Ogura N. 1994. Rapid removal of Linear Alkylbenzenesulfonates (LAS) by attached biofilm in an urbun shallow stream. *Water Research*, 28(9), 1953-1960.

Takada H., Ogura N. & Ishiwateri R. 1992. Seasonal variations and Modes of riverine input of organic pollutants to the coastal zone: 1. Flux of detergent-derived pollutants to Tokyo Bay. *Environment Science & Technology*, 26, 2517-2523.

The Cadmus Group I. 1996. Aquatic ecological risk assessment software user's manual, version 2.0. RP91-AER1, Water Environment Research Foundation, Alexandria, VA, USA.

Van Der Hoeven N. 2001. Estimating the 5-percentile of the species sensitivity distributions without any assumptions about the distribution. *Ecotoxicology*, 10, 25-34.

van Leeuwen C. J. 1990. Ecotoxicological effects assessment in The Netherlands, recent developments. *Environmental Management*, 14, 779-792.

Van Sprang P., Verdonck F. A. M. & Vanrolleghem P. A. 2002. Probabilistic risk assessment of Zn in European surface waters. EURAS bvba, Gent, Belgium.

Van Straalen N. M. 1990. New methodologies for estimating the ecological risk of chemicals in the environment. *Proceedings 6th Int. IAEG Congress*, 165-173. Rotterdam, The Netherlands.

Van Straalen N. M. & Denneman C. A. J. 1989. Ecotoxicological evaluation of soil quality criteria. *Ecotoxicological Environmental Safety*, 18, 241-251.

Van Straalen N. M. & van Leeuwen C. J. 2002. European history of species sensitivity distributions. Pages 19-34 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers, London, UK.

Vandenbroele M., Janssen C. R., Bogaert S. & Le Roy D. 2000. Hefbomen voor een beleid gericht op duurzame ontwikkeling: uitwerking van evaluatiecriteria voor pesticiden in relatie tot de milieufiscaliteit. HL/DD/011, Laboratorium voor Milieutoxicologie en Aquatische Ecologie,Ghent University (in dutch), Gent, Belgium.

Vanhooren H., Meirlaen J., Amerlinck Y., Claeys F., Vangheluwe H. & Vanrolleghem P. A. 2003. WEST: Modelling biological wastewater treatment. *Journal of Hydroinformatics*, 5, 27-50.

Vega M. M., Urzelai A. & Angulo E. 1999. Minimum data required for deriving soil quality criteria from invertebrate ecotoxicity experiments. *Environmental Toxicology and Chemistry*, 18(6), 1304-1310.

Verdonck F. A. M., Boeije G. M., Schowanek D. R., Vanrolleghem P. A. & Force E. G.-E. T. 2000. Geography-referenced regional exposure tool for European rivers (GREAT-ER): A case study for the Rupel basin. *Proceedings 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4)*. Banff, Alberta, Canada. September 2 - 8, 2000.

Versteeg D. J., Belanger S. E. & Carr G. J. 1999. Understanding single-species and model ecosystem sensitivity: data-based comparison. *Environmental Toxicology Chemistry*, 18(6), 1329-1346.

VMM. 2001. personal communication. *in*. Flemisch Environmental Agency, Aalst, Belgium.

von Sperling M. 1996. Design of facultative ponds based on uncertainty analysis. *Water Science & Technology*, 33(7), 41-47.

Vose D. 1996. Quantitative risk analysis. A guide to Monte Carlo simulation modeling. John Wiley & sons. 317.

Wagner C. & Løkke H. 1991. Estimation of ecotoxicological protection levels for NOEC toxicity data. *Water Research*, 25, 1237-1242.

Warren-Hicks W., Carbone J. P. & Havens P. L. 2002a. Using Monte Carlo techniques to judge model prediction accuracy: validation of the pesticide root zone model 3.12. *Environmental Toxicology and Chemistry*, 21(8), 1570-1577.

Warren-Hicks W., Parkhust B. R., Qian S. & Kellar P. 2002b. Accounting for toxicity test variability in evaluating test results. *Proceedings 75th Annual Conference and Exposition of Water Environment Federation (WEFTEC)*. USA.

Warren-Hicks W. J. & Moore D. R. J. 1995. Uncertainty analysis in ecological risk assessment. SETAC, Pensacola, US. 275.

Warren-Hicks W. J., Parkhust B. R. & Butcher J. B. 2002c. Methodology for aquatic ecological risk assessment. Pages 345-382 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers.

Warren-Hicks W. J., Parkhust B. R., Moore D. R. J., Teed R. S., Baird R. B., Berger R., Denton D. L. & Pletl J. J. 2000. Assessment of whole effluent toxicity test variability: partitioning sources of variability. *Environmental Toxicology and Chemistry*, 19(1), 94-104.

Wheeler J. R., Grist E. P. M., Leung K. M. Y., Morritt D. & Crane M. 2002. Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin*, 45, 192-202.

Summary

Environmental pollution of toxic substances has led governments to develop new laws and regulation that puts constraints on these chemical emissions. These are based on environmental quality standards and environmental/ecological risk assessment. The key question to be answered is: "What is the likelihood (i.e. probability) of adverse effects occurring to exposed ecological systems due to exceedance of a toxicity level by an environmental concentration?". The goal of ecological risk assessment is to estimate this likelihood. It is based on the comparison of a predicted or measured exposure/environmental concentration with a 'no effect concentration' based on a set of (acute or chronic) toxicity test results (i.e. testing species sensitivity).

This PhD dissertation studied and developed a range of statistical techniques needed to answer the key question with a risk probability and an uncertainty or confidence interval rather than with the current black white "yes, there is potential risk / no risk" answer which the conventional risk assessments provide. Such answers may mislead stakeholders to think that ecological risks are simple black and white issues.

After all, in a Probabilistic Ecological Risk Assessment (PERA), the exposure concentration and species sensitivity are treated as random variables taken from probability distributions (respectively Exposure Concentration Distribution (ECD) and Species Sensitivity Distribution (SSD)) which are combined to give a risk probability. In this way, the inherent variability and uncertainty of the environmental concentration and the species sensitivity is accounted for. Variability represents inherent heterogeneity or diversity in a well-characterised population. Uncertainty represents partial ignorance or lack of perfect information about poorly characterised phenomena or models (e.g. sampling or measurement error). PERA therefore delivers a more transparent, realistic and non-conservative approach to estimate risks. It is recognised in literature that probabilistic methods would improve the environmental evaluation of chemicals, if appropriate action is taken to address their potential weaknesses.

Some of these current (mainly statistical) weaknesses in probabilistic ecological risk assessment are addressed in this PhD dissertation. Most of them deal with misuse of existing techniques (e.g. Monte Carlo analysis, bootstrap), reliability of statistical techniques at small sample size, the lack of consensus on which method or model or what sample size to use, misinterpretation of probability distributions (e.g. output of Monte Carlo analysis), inappropriately or insufficiently dealing with uncertainty or variability (e.g. one- versus two-dimensional Monte Carlo analysis), discussions on how to calculate probabilistic risk…

It was shown that interpretation of all probability distributions in a PERA framework should be made carefully. In Monte Carlo simulations, separation of uncertainty and variability and the correct application of Monte Carlo analysis simplify the interpretation of a model's output distribution of interest. A case study showed that the exposure concentration of total ammonia nitrogen in a Flemish river will be larger than 10 mg/l for 40% of the time. This is a quite different result than being 40% certain that the exposure concentration will be larger than 10 mg/l.

A probabilistic risk should be interpreted as the probability that a random selected exposure or environmental concentration will exceed a species sensitivity. Examples were developed that show that the same risk probability can represent different environmental conditions (e.g. depending on whether the ECD represents spatial or temporal variability). Therefore, it is suggested to always include as much information as possible in the answer to the key question described above: indicate what type of variability the ECD or the SSD represents (geo- or time-referenced), what endpoint was used …

Throughout this dissertation, parametric and nonparametric methods were often compared. The results appear to be very sensitive to the chosen method. The proper use of distribution selection methods was stressed as well. Statistical tests, graphical exploration and expert knowledge can help in identifying the appropriate distribution. To calculate a lower percentile (e.g. the 5^{th} percentile), it was found that preference should be given to parametric methods when the sample size is below 10 while preference should be given to nonparametric methods when the sample size is large (e.g. 50). For the intermediate sample sizes, either parametric or nonparametric techniques can be used or maybe a combination of the two could be used.

Several examples and case studies have proven that the probabilistic risk characterisation considers the quantitative information of the full range of the ECD and SSD (including lower SS than its 5^{th} percentile and higher ECs than the 90^{th} percentile) instead of only considering the upper tail of the ECD and the lower tail of the SSD as in traditional risk assessment. Consequently, several issues on calculating tail percentiles can be omitted because they are no longer needed in the risk characterisation.

Finally, attention was focused on the fact that the probability distributions in probabilistic risk assessment can be wide due to large spatial (and temporal) variability. Instead of lumping all the sources of variability into one probability distribution, spatial and/or temporal differences and dependencies between EC and SS can be explicitly accounted for in a respectively geo- and/or time-referenced analysis (or spatial-temporal analysis). In this way, the risk assessment becomes more realistic as more information is taken into account. This was confirmed by several case studies.

Samenvatting

Milieuverontreiniging van chemische stoffen heeft ertoe geleid dat overheidsinstanties nieuwe wetten hebben ingevoerd die restricties leggen op deze chemische emissies. Deze restricties zijn gebaseerd op milieunormen en milieu-/ecologische risicoanalyse. De te beantwoorden hamvraag is: "Wat is de kans op nadelige effecten op blootgestelde organismen of ecologische systemen door overschrijding van een toxiciteitsniveau door een milieuconcentratrie?". De doelstelling van ecologische risicoanalyse is deze kans in te schatten. Risicoanalyse is gebaseerd op de vergelijking van een voorspelde of gemeten milieu- of blootstellingsconcentratie met een 'geen effect concentratie' gebaseerd op een set van (acute of chronische) toxiciteitstestresultaten (d.i. het testen van soortengevoeligheid).

Deze thesis bestudeerde en ontwikkelde verschillende statistische technieken die nodig zijn om de hamvraag te kunnen beantwoorden met een kans en een onzekerheids- of betrouwbaarheidsinterval i.p.v. het huidige zwart-wit antwoord "ja, er is een potentieel risico / geen risico" dat door de conventionele risicoanalyses wordt gegeven. Dergelijke antwoorden kunnen beleidsbeslissers of risicomanagers doen misleiden door hen te doen denken dat ecologische risico's eenvoudige zwart-wit problemen zijn.

Immers, in een Probabilistische Ecologische RisicoAnalyse (PERA) worden de blootstellings-concentratie en de soortengevoeligheid behandeld als random variabelen en voorgesteld door probabiliteitsdistributies (in het vakjargon respectievelijk *Exposure Concentration Distribution* (ECD) en *Species Sensitivity Distribution* (SSD) genoemd) die gecombineerd in een risico-probabiliteit of -kans resulteren. Op deze manier wordt de intrinsieke variabiliteit en onzekerheid van de blootstellingsconcentratie en de soortengevoeligheid in rekening gebracht. Variabiliteit stelt de instrinsieke heterogeniteit of diversiteit voor in een goed karakteriseerde populatie. Onzekerheid stelt de gedeeltelijke onwetenheid of het gebrek aan perfecte informatie over weinig gekende fenomenen of modellen voor (bvb. meetfouten). Daarom is PERA een meer transparante, realistische en niet-conservatieve manier om risico's te schatten. Het is ook erkend in de literatuur dat probabilistische methoden de milieurisicoanalyse van chemicaliën zou verbeteren, indien geschikte actie wordt ondernomen om hun potentiële zwakheden aan te pakken.

Sommige van deze huidige (voornamelijk statistische) zwakheden in PERA worden behandeld in deze scriptie. Het merendeel van hen handelen over verkeerd gebruik van bestaande technieken (bvb. Monte Carlo analyse, bootstrap), de betrouwbaarheid van statistische technieken voor kleine datasets, het gebrek aan consensus over welke methode of model of welke datasetgrootte te gebruiken, verkeerde interpretatie van probabiliteitsdistributies (bvb. de uitkomst van een Monte Carlo analyse), onjuist of onvoldoende aanpakken van variabiliteit en onzekerheid (bvb. één- versus twee-dimensionele Monte Carlo analyse), discussies over de berekeningswijze van een probabilistisch risico, …

Het is aangetoond dat interpretatie van alle probabiliteitsdistributies in een PERA framework met de nodige voorzichtigheid dient te gebeuren. In de correcte toepassing van Monte Carlo simulaties vereenvoudigt de scheiding van onzekerheid en variabiliteit de interpretatie de output distributie van een model. Een gevallenstudie toonde aan dat de blootstellingsconcentratie van totale ammonium stikstof in een Vlaamse rivier groter zal zijn dan 10 mg/l gedurende 40% van de tijd. Dit is een vrij verschillend resultaat in vergelijking met het resultaat dat men meer dan 40% zeker is dat de blootstellingsconcentratie groter zal zijn 10 mg/l.

Een probabilistisch risico wordt geïnterpreteerd als de kans dat een random geselecteerde blootstellings- of milieuconcentratie een random geselecteerde soortengevoeligheid zal overschrijden. Voorbeelden werden echter ontwikkeld die aantonen dat hetzelfde probabilistisch risico verschillende milieusituaties kan voorstellen (bvb. afhankelijk of de blootstellingsdistributie ruimtelijke of temporele variabiliteit voorstelt). Daarom wordt er gesuggereerd om altijd zoveel mogelijk informatie toe te voegen bij het antwoord op de hierboven beschreven hamvraag: aan-duiden welk type van variabiliteit de blootstelling en effecten voorstellen (al dan niet geografisch of tijdsgerelateerd), …

Doorheen de scriptie werden parametrische en niet-parametrische methoden vaak met elkaar vergeleken. De resultaten bleken sterk afhankelijk te zijn van de gebruikte methode. Het juiste gebruik van distributieselectiemethoden werd dan ook benadrukt. Statistische testen, grafische verkenning en expertkennis kunnen helpen om de geschikte distributie te identificeren. Er werd gevonden dat, om een staartpercentiel te berekenen (bvb. de 5^{de} percentiel), voorkeur zou moeten gaan naar parametrische methoden wanneer de datasetgrootte lager is dan 10 terwijl voorkeur zou moeten gaan naar niet-parametrische methoden wanneer de datasetgrootte groot is (bvb. 50). Voor de tussenliggende datasetgroottes kunnen ofwel parametrische of niet-parametrische technieken gebruikt worden ofwel misschien een combinatie van de twee.

Verschillende voorbeelden en gevallenstudies hebben bewezen dat probabilistische risico-karakterisatie de kwantitatieve informatie van het volledige bereik van de ECD en SSD beschouwt (inclusief soortengevoeligheden kleiner dan de 5^{de} percentiel en blootstellingsconcentraties groter dan de 90^{ste} percentiel). Bijgevolg kunnen verschillende problemen bij het berekenen van staart-percentielen worden weggelaten omdat ze niet langer nodig zijn in de risicokarakterisatie.

Tenslotte werd aandacht besteed aan het feit dat de probabiliteitsdistributies in PERA grote varianties kunnen hebben door een grote ruimtelijke (of temporele) variabiliteit. In plaats van alle bronnen van variabiliteit samen te gooien in één distributie, kunnen ruimtelijke en/of tijdelijke verschillen en mogelijke afhankelijkheden tussen blootstelling en effecten expliciet in rekening gebracht worden in een respectievelijk geografisch- en/of tijdsgerelateerde analyse. Op deze manier wordt de risicoanalyse realistischer aangezien meer informatie mee in rekening wordt gebracht. Dit werd bevestigd door verschillende gevallenstudies.

C u r r i c u l u m V i t a e

ir. Frederik VERDONCK

Hoekskensstraat 73
B-9080 Lochristi
09 / 342 06 42 - 0474 / 44 66 05
frederik.verdonck@biomath.rug.ac.be
27 jaar (° 16 juni 1976 te Gent), samenwonend, Belg

Diploma's

1999 – 2002 **Doctoraatsopleiding in de Toegepaste Biologische Wetenschappen**

Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen, Universiteit Gent, Gent.

1994 – 1999 **Bio-ingenieur in de milieutechnologie**

Geslaagd met onderscheiding aan de Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen, Universiteit Gent, Gent.

Scriptie (in samenwerking met *Procter & Gamble*): 'Toepassing van een geografisch gerefereerd regionaal blootstellingsmodel voor chemicaliën op het Rupelstroombekken', Vakgroep voor Toegepaste Wiskunde, Biometrie en Procesregeling (BIOMATH), Promotor: Prof. dr. ir. P. Vanrolleghem.

1988 – 1994 Secondair onderwijs: **Moderne Humaniora Wetenschappelijke A**

Geslaagd als eerste van de klas met net geen grote onderscheiding aan het Glorieux Instituut (nu: EDUGO), Gent-Oostakker.

Werkervaring

Sinds Nov 1999 **Universiteit Gent, Vakgroep voor Toegepaste Wiskunde, Biometrie en Procesregeling (BIOMATH), Gent.** IWT-bursaal. Doctoraat over "Geografisch gerefereerde probabilistische ecologische risicoanalyse".

Jan 2000 – Nov 2000 **International Water Association (IWA).** Wetenschappelijke secretaris. Mede-organisator van het internationaal symposium WATERMATEX ("System analysis and computing in water quality management"), 18-20 September 2000, Gent.

Aug 1999 - Okt 1999 **Procter & Gamble, European Technical Center, Environmental Safety Group, Strombeek-Bever.** Environmental engineer/GIS expert.

Juli-Aug 1987-1997 **Bloemistrij VERDONCK-NAUDTS, Lochristi.** (Vakantiejob). Hulp op het tuinbouwbedrijf.

Prijzen

31 juni 1994	1e prijs voor het vak wiskunde, Gent-Oostakker
23 juni 2000	finalist voor WEL prijs water (met eindejaarsthesis 1999), Aalst
22 oktober 2001	1e prijs voor de beste poster op B-IWA bijeenkomst, Brussel
9 oktober 2002	Beste poster op PhD-symposium, Gent

Journal refereeing

Water Research
Environmental Toxicology & Chemistry
Human and Ecological Risk Assessment
Environmental Pollution
Environmental Science & Technology
Computers & Industrial Engineering

Publicaties

Verdonck F. (1999) Toepassing van een geografisch gerefereerd blootstellingsmodel voor chemicaliën op het Rupelstroombekken. Scriptie. Universiteit Gent, Faculteit Landbouwkundige en toegepaste biologische wetenschappen.

Verdonck F., Boeije G., Vanrolleghem P.A., Feijtel T.C.J. & Schowanek D. (2000) Toepassing van GREAT-ER, een blootstellingsmodel voor chemicaliën, op het Rupelstroombekken. @WEL4 - Water.

Verdonck F., Boeije G., Vanrolleghem P.A., Feijtel T.C.J. & Schowanek D. (2000) Toepassing van GREAT-ER, een blootstellingsmodel voor chemicaliën, op het Rupelstroombekken. Chemie magazine, 2, 75-79.

Rousseau D., Verdonck F., Moerman O., Carrette R., Thoeye C., Meirlaen J., Vanrolleghem P.A. (2001) Development of a risk assessment based technique for design/retrofitting of WWTPs. Water Science & Technology, 43 (7), 287-294.

De Clercq B., Meirlaen J., Verdonck F. & Vanrolleghem P.A. (2001) An overview of the posters presented at WATERMATEX 2000. I. New models / integrated urban wastewater systems / time series analysis. Water Science & Technology, 43 (7), 377-380.

Verdonck, F., Jaworska, J. & Vanrolleghem, P.A. (2001). Comments on species sensitivity distributions sample size determination based on Newman et al. (2000). SETAC Globe 2(3), Learned Discourse, 22-24.

Verdonck F.A.M., Jaworska J., Thas O. & Vanrolleghem P.A. (2001) Determining environmental standards using bootstrapping, Bayesian and maximum likelihood techniques: a comparative study. Analytica chimica acta, 446 (1-2), 429-438.

Verdonck F.A.M. & Vanrolleghem P.A. (2002). Milieurisicoanalyse in een onzekere wereld. Het ingenieursblad, 8-9, 43-47.

Bixio D., Parmentier G., Rousseau D., Verdonck F., Meirlaen J., Peter A. Vanrolleghem & Thoeye C. (2002) A quantitative risk analysis tool for design/simulation of wastewater treatment plants. Water Science & Technology, 46 (4-5), 301-307.

Verdonck F.A.M., Deksissa T., Matamoros D. & Vanrolleghem P.A. (ingestuurd) Dealing with variability in chemical exposure modelling in rivers. Journal of Environmental Pollution.

Verdonck F.A.M., Aldenberg, T., Jaworska, J. & Vanrolleghem P.A. (in druk). Limitations of current risk characterization methods in probabilistic ecological risk assessment. Environmental Toxicology and Chemistry.

Verdonck F.A.M., Jaworska J., Janssen C.R. & Vanrolleghem P.A. (in druk) Geography Referencing Probabilistic Risk Of Chemicals In Rivers. Water Sience and Technology.

Congressen, studiedagen (orale presentaties) .....

Verdonck F.A.M., Boeije G.M., Schowanek D.R., Vanrolleghem P.A. & ECETOC GREAT-ER Task Force (2000) Geography-referenced regional exposure tool for European rivers (GREAT-ER): A case study for the Rupel basin. Proceedings 4th International Conference on Integrating Conference on Integrating GIS and Environmental Modeling (GIS/EM4), 2-8 september 2000, Banff, Alberta, Canada.

Rousseau D., Verdonck F., Moerman O., Carrette R., Thoeye C., Meirlaen J. & Vanrolleghem P.A. (2000) Development of a risk assessment based technique for design/retrofitting of WWTPs. Preprints WATERMATEX 2000 Symposium, 18-20 september 2000, Gent, België & Water Science & Technology, 43 (7), 287-294.

Verdonck F. & Vanrolleghem P.A. (2000) Statistical tools in risk assessment. COST-meeting, 21-22 september 2000, Aartselaar, België.

Verdonck F., Jaworska J. & Vanrolleghem P.A. (2000) Determining safe environmental standards using bootstrapping, Bayesian and maximum likelihood techniques: a comparative study, CAC-2000, 7th International Conference on Chemometrics in Analytical Chemistry, 16-20 oktober 2000, Antwerpen, België.

Verdonck F.A.M., Jaworska J., Janssen C.R. & Vanrolleghem P.A. (2001) Comparing statistical techniques for uncertainty assessment of species sensitivity distributions: effect of sample size. 11th Annual Meeting of SETAC-Europe, 6-10 mei 2001, Madrid, Spanje.

Bixio D., Parmentier G., Rousseau D., Verdonck F., Meirlaen J., Vanrolleghem P.A. and Thoeye C. (2001) Intergrating risk analysis in the design/simulation of activated sludge systems. In: Proceedings 74th Annual Conference and Exposition of Water Environment Federation (WEFTEC). 14-18 oktober 2001, Atlanta, VS.

Verdonck F.A.M., Janssen C.R., Thas O., Jaworska J. & Vanrolleghem P.A. (2001) Probabilistic environmental risk assessment. PhD symposium, Med. Fac. Landbouww. Univ. Gent, 66(4), 13-19. 10 oktober 2001, Gent, België.

Bixio D., Rousseau D., Verdonck F., Thoeye C., Meirlaen J., Peter A. Vanrolleghem. (2001) A quantitative risk analysis tool for design/simulation of wastewater treatment plants. In: Proceedings 2nd IWA World Water Congress. 15-19 oktober 2001, Berlijn, Duitsland & Water Science & Technology, 46 (4-5), 301-307.

Verdonck F.A.M., Deksissa T., Matamoros D. & Vanrolleghem P.A. (2002) Dealing with variability in chemical exposure modelling in rivers. In: Proceedings Seminar on Exposure and Effects, Modelling in Environmental Toxicology. C.5. 4-8 februari 2002, Antwerpen, België.

Verdonck F., Janssen C.R., Jaworska J., Thas O. & Vanrolleghem P.A. (2002) Accounting for hierachical variability in species sensitivity distributions. 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.

Verdonck F.A.M., Thas O., Jaworska J. & Vanrolleghem P.A. (2002) Added value of a hierarchical bootstrap model in environmental standard setting. In: Proceedings TIES 2002, Annual Conference of The International Environmetrics Society. 110. 18-22 juni 2002, Genua, Italië.

Verdonck F.A.M., Jaworska J., Janssen C.R. & Vanrolleghem P.A. (2002) Probabilistic ecological risk assessment framework for chemical substances. In: Proceedings iEMSs 2002, Integrated Assessment and Decision Support Vol. 1, 144-149, 24-27 juni 2002, Lugano, Zwitserland.

Verdonck F.A.M., Jaworska J., Janssen C.R. & Vanrolleghem P.A. (2002) Geography Referencing Probabilistic Risk Of Chemicals In Rivers. In: Proceedings 1st IWA Young Researchers Conference, 33-40, 9-10 september 2002, Cranfield, Verenigd Koninkrijk.

Vanrolleghem, PA. & Verdonck, F.A.M. (2002) Appetizer on concepts and state of the art in chemical (ecological) risk assessment. COST meeting, Innsbrück, Oostenrijk.

Verdonck FA.M., Janssen C.R. & Vanrolleghem P.A. (2002) Potential of ecological informatics in probabilistic risk assessment of chemicals in rivers. Proceedings Ecological informatics applications in water management conference, 18, 6-7 november 2002, Gent, België.

Verdonck FA.M., Deksissa T., De Laender F., De Schamphelaere K.A.C., Matamoros D., Vandenberghe, V., Vincke S., Janssen, C.J. & Vanrolleghem P.A. (2003). Uncertainty and variability in spatio-temporal probabilistic risk modelling. 13th Annual Meeting of SETAC-Europe, 27 april – 1 mei 2003, Hamburg, Duitsland.

Verdonck F.A.M., Rousseau D., Bixio D., Thoeye C. & Vanrolleghem P.A. (2003). Added value of concentration-duration-frequency curves of wastewater plant effluent quality. International Congress on Modelling and Simulation, 13-17 juli 2003, Townsville, Australië.

Congressen, studiedagen (poster presentatie) .....

Verdonck F., Boeije G., Schowanek D. & Vanrolleghem P.A. GREAT-ER: Geo-referenced Regional Exposure Assessment Tool for European Rivers: a case study for the Rupel basin (Belgium)

- 1st ESRI BeLux USER DAYS, 26-27 november 1998, Wemmel, België.
- Final workshop GREAT-ER, 15 maart 1999, Arona - Milaan, Italië.
- 9th Annual Meeting of SETAC-Europe, 25-29 mei 1999, Leipzig, Duitsland.
- 13th forum for applied biotechnology, 22-23 september 1999, Gent, België. (Med. Fac. Landbouww. Univ. Gent, 64/5a. 225-228).
- 2th ESRI BeLux USER DAYS, 18-19 november 1999, Wemmel, België.

Verdonck F. & Schowanek D. Application and validation of the Rupel catchment in GREAT-ER.

- BIWA (IAWQ-Belgium) "Happy hour", 14 maart 2000, Brussel, België.
- 3th SETAC World Congress, 21-25 mei 2000, Brighton, Verenigd Koninkrijk.

Verdonck F., Jaworska J., Thas O. & Vanrolleghem P.A. Uncertainty techniques in environmental risk assessment.

- PhD symposium, 11 oktober 2000, Gent, België. (Med. Fac. Landbouww. Univ. Gent, 65/4, 247-252).

Rousseau D., Verdonck F., Carrette R., Thoeye C., Meirlaen J. & Vanrolleghem P.A. Risk assessment tool for the design of waste water treatment plants.

- PhD symposium, 11 oktober 2000, Gent, België (Med. Fac. Landbouww. Univ. Gent, 65/4, 219-223).
- BIWA (IWA-Belgium) "Happy hour", 19 februari 2001, Brussel, België.

Verdonck F.A.M., Janssen C., Schowanek D., Jaworska J. & Vanrolleghem P.A. Geography referenced probabilistic risk assessment, a case study for the Rupel basin.

- Ecologisch Onderzoek in het Scheldebekken, 29-30 maart 2001, Brussel, België.
- BIWA (IWA-Belgium) "Happy hour", 22 oktober 2001, Brussel, België. (1e prijs voor beste poster)

Verdonck F.A.M., Jaworska J., Janssen C.R. & Vanrolleghem P.A. (2002). Methodologies to determine risk of chemicals in rivers under data uncertainty.

- IWA 3rd World Water Congress, 7-12 April 2002, Melbourne, Australia. (Proceedings).
- 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.

Rottiers, A., Verdonck F. & Schowanek, D. Evaluation of the Rupel River Basin Implementation in the GREAT-ER Model through Boron Monitoring.

- BIWA (IWA-Belgium) "Happy hour", 18 maart 2002, Brussel, België.
- 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.

Verdonck F.A.M., Thas O. & Vanrolleghem P.A. Characterising inter-laboratory variability in environmental standard setting using weighted hierarchical bootstrapping.

- 9th Annual Meeting of the Belgian Statistical Society, 12-13 oktober 2001, Oostende, België.
- 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.

Deksissa T., Verdonck F., Vanrolleghem P.A. Dynamic fate model in river: A case study for the fate of LAS in river Lambro.

- 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.
- PhD symposium, 9 oktober 2002, Gent, België.

De Pauw D.J.W., Carvalho G., Nopens I., Verdonck F.A.M., Meirlaen J. & Vanrolleghem P.A.. WEST, a general tool for dynamic modelling and simulation of biodegradation processes.

- 12th Annual Meeting of SETAC-Europe, 12-16 mei 2002, Wenen, Oostenrijk.

Verdonck F.A.M., Matamoros, D., Deksissa, T., Vandenberghe, V. & Vanrolleghem P.A. Spatio-temporal probabilistic environmental modelling.

- PhD symposium, 9 oktober 2002, Gent, België (Med. Fac. Landbouww. Univ. Gent, 67/4, 205-208).

Verdonck F.A.M., Aldenberg T., Jaworska J., Thas O. & Vanrolleghem P.A. Limitations of current probabilistic risk calculation methods

- 10th Annual Meeting of the Belgian Statistical Society, 18-19 oktober 2002, Kerkrade, Nederland.
- 13th Annual Meeting of SETAC-Europe, 27 april – 1 mei 2003, Hamburg, Duitsland.

De Laender F., Verdonck F.A.M., Deschamphelaere K.A.C., Janssen C.R. & Vanrolleghem P.A. Geography-referenced bioavailability modelling in risk assessment: a case study of copper in Swedish surface waters

- 13th Annual Meeting of SETAC-Europe, 27 april – 1 mei 2003, Hamburg, Duitsland.
- 7th Int. Conf. on the biogeochemistry of trace elements (ICOBTE), 15-17 juni 2003, Uppsala, Zweden

Congressen, studiedagen, cursussen (deelname).....

Tijdens mijn bio-ingenieur- & doctoraatsopleiding werd steeds gestreefd naar een ideale mix van toegepaste wiskunde (met nadruk op modellering, statistiek,…) en 'bio-wetenschappen' (met nadruk op milieu).

19-21 okt 1998	GREAT-ER & ARCINFO: privécursus in het kader van mijn thesis, University of Osnabrück, Institute of Environmental Systems Research, Osnabrück, Duitsland.
21 april 1999	Neveneffecten van gewasbeschermingsmiddelen op mens en milieu in perspectief geplaatst. Studiedag TI-KVIV, Tervuren, België.
okt-jun 1998-1999	Keuzevakken in mijn laatste jaar: grondwatermodellering, grondwaterkwaliteit, beslissingsondersteunende technieken in het milieubeheer, bioprocesregeling, statistische data-analyse, Universiteit Gent, Gent, België.
20 januari 2000	Environmental modelling: an uncertain future? Interuniversitaire Buitenlandse Francquileerstoel. Leuven, België.
9 maart 2000	Pharmaceuticals in the environment. Studiedag TI-KVIV, Brussel, België.
10-12 april 2000	Workshop on Statistical Inference, Diepenbeek, België.
21 mei 2000	Probabilistic ecological risk assessment for pesticides, Third SETAC World Congress, Brighton, Verenigd Koninrijk.
19-21 juni 2000	World Engineers' Convention. Environment, Climate, Health. EXPO2000. Hannover, Duitsland.
18-20 sept 2000	Internationaal Symposium over "Systems analysis and computing in water quality management" (WATERMATEX), Gent, België.
27 oktober 2000	Seminarie VITO-methodologie voor de risicobeoordeling van milieuverontreiniging. Praktijkvoorbeelden van een geïntegreerde benadering. Gent, België.
16 november 2000	New trends in mathematical modeling and numerical methods. Leuven, België.
15 februari 2001	Werkgroep ecotoxicologie en statistiek, Bilthoven, Nederland.
17 mei 2001	Forum voor wateronderzoek in Vlaanderen, studiedag Vlaams Water Netwerk, Brussel, België.
9-13 juli 2001	Course on Quantitative methods in ecotoxicology by Newman, Antwerpen, België.
24-25 sept 2001	15th forum for applied biotechnology, Gent, België.
26-27 nov 2001	Doctoral course on survival analysis by Dave Harrington, , Gent, België.
29 november 2001	Studiedag bosomvorming, Gent, België.
30 november2001	Doctoral course on robust statistical methods, Leuven, België.
2de semester 2002	Doctoral course in Trends in Research and Development: Risk analysis and quality management, Gent, België.
24 februari – 1 maart 2002	Uitgenodigd op SETAC Pellston workshop on the application of uncertainty analysis to ecological risks of pesticides, Pensacola, Florida, USA.
5 april 2002	Workshop on resampling methods by Philip Good, Gent, België.
2de semester 2002	Doctoral course in Trends in Research and Development: Experimental design for bio-process engineering, Gent, België.
2 december 2002	Doctoral course on longitudinal data, Leuven, België.
20 januari 2003	Symposium on Environmental Statistics, Gent, België.
31 januari 2003	Waterforum over watersysteemkennis, Brussel, België.
22 juni 2003	Beyond Monte Carlo: tutorial workshop on imprecise probabilities, Brussel, België.