

**STATISTICAL MODELLING OF VARIABILITY AND
UNCERTAINTY IN RISK ASSESSMENT OF
NANOPARTICLES**

Rianne Jacobs

Thesis committee

Promotor

Prof. Dr C.J.F. ter Braak

Personal chair at the Mathematical and Statistical Methods Group
Wageningen University

Co-promotor

Dr H. van der Voet

Researcher statistics at PRI Biometris
Wageningen University and Research Centre

Other members

Prof. Dr A.A. Koelmans, Wageningen University

Dr P.S. Craig, Durham University, UK

Dr F. Verdonck, Arche, Ghent, Belgium

Prof. Dr A.M.J. Ragas, Open University, Heerlen and Radboud University, Nijmegen, The Netherlands

This research was conducted under the auspices of the Graduate School of
Production Ecology & Resource Conservation (PE & RC)

STATISTICAL MODELLING OF VARIABILITY AND UNCERTAINTY IN RISK ASSESSMENT OF NANOPARTICLES

Rianne Jacobs

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Thursday 7 July 2016

at 8:30 a.m. in the Aula.

Rianne Jacobs

Statistical Modelling of Variability and Uncertainty in Risk Assessment of Nanoparticles,
205 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)

With references, with summary in English and Dutch

ISBN 978-94-6257-819-7

DOI 10.18174/382043

ABSTRACT

Engineered nanoparticles (ENPs) are used everywhere and have large technological and economic potential. Like all novel materials, however, ENPs have no history of safe use. Insight into risks of nanotechnology and the use of nanoparticles is an essential condition for the societal acceptance and safe use of nanotechnology.

Risk assessment of ENPs has been hampered by lack of knowledge about ENPs, their environmental fate, toxicity, testing considerations, characterisation of nanoparticles and human and environmental exposures and routes. This lack of knowledge results in uncertainty in the risk assessment. Moreover, due to the novelty of nanotechnology, risk assessors are often confronted with small samples of data on which to perform a risk assessment. Dealing with this uncertainty and the small sample sizes are main challenges when it comes to risk assessment of ENPs. The objectives of this thesis are (i) to perform a transparent risk assessment of nanoparticles in the face of large uncertainty in such a way that it can guide future research to reduce the uncertainty and (ii) to evaluate empirical and parametric methods to estimate the risk probability in the case of small sample sizes.

To address the first objective, I adapted an existing Integrated Probabilistic Risk Assessment (IPRA) method for use in nanoparticle risk assessment. In IPRA, statistical distributions and bootstrap methods are used to quantify uncertainty and variability in the risk assessment in a two-dimensional Monte Carlo algorithm. This method was applied in a human health (nanosilica in food) and an environmental (nano TiO_2 in water) risk context. I showed that IPRA leads to a more transparent risk assessment and can direct further environmental and toxicological research to the areas in which it is most needed.

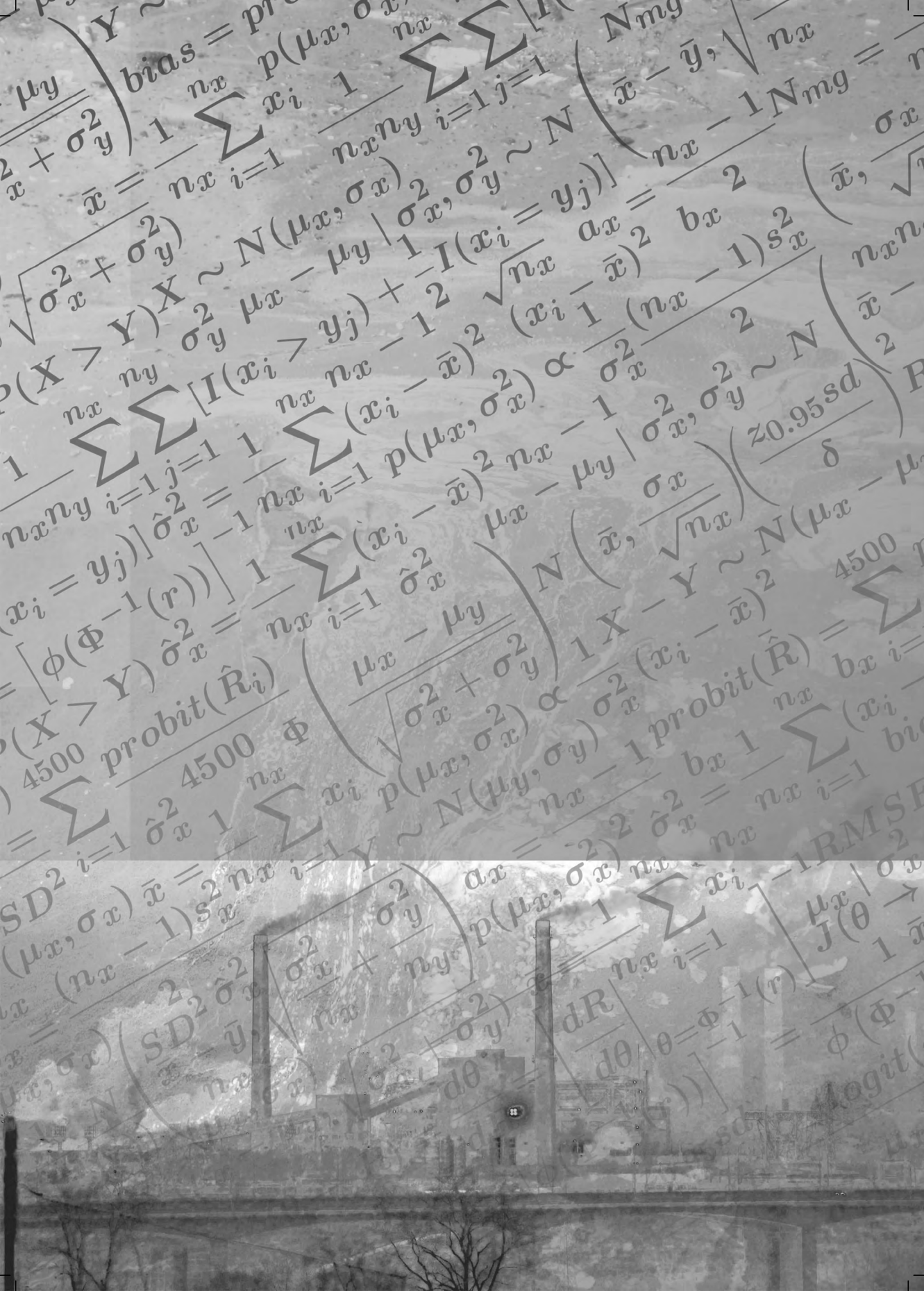
For the second objective, I addressed the problem of small sample size of the critical effect concentration (CEC) in the estimation of $R = P(ExpC > CEC)$, where $ExpC$ is the exposure concentration. First I assumed normality and investigated various parametric and non-parametric estimators. I found that, compared to the non-parametric estimators, the parametric estimators enable us to better estimate and bound the risk when sample sizes and/or small risks are small. Moreover, the Bayesian estimator out-

performed the maximum likelihood estimators in terms of coverage and interval lengths. Second, I relaxed the normality assumption for the tails of the exposure and effect distributions. I developed a mixture model to estimate the risk, $R = P(ExpC > CEC)$, with the assumption of a normal distribution for the bulk data and generalised Pareto distributions for the tails. A sensitivity analysis showed significant influence of the tail heaviness on the risk probability, R , especially for low risks.

In conclusion, to really be able to focus the research into the risks of ENPs to the most needed areas, probabilistic methods as used and developed in this thesis need to be implemented on a larger scale. With these methods, it is possible to identify the greatest sources of uncertainty. Based on such identification, research can be focused on those areas that need it most, thereby making large leaps in reducing the uncertainty that is currently hampering risk assessment of ENPs.

CONTENTS

Abstract	v
1 General Introduction	1
2 Integrated probabilistic risk assessment for nanoparticles: the case of nano-silica in food	27
3 Combining exposure and effect modelling into an integrated probabilistic environmental risk assessment for nanoparticles	53
4 Parametric estimation of $P(X > Y)$ for normal distributions in the context of probabilistic environmental risk assessment	79
5 Normal-GPD mixtures for modelling exposure and effect concentrations in environmental risk assessment of nanoparticles	109
6 General Discussion	151
References	172
Acronym List	190
Summary	193
Samenvatting (Dutch Summary)	196
About the Author	200
List of Publications	201
Dankwoord (Acknowledgements)	202
Education Statement	204



1

GENERAL INTRODUCTION

Rianne Jacobs

Nanotechnology offers possibilities to increase the quality of life. At the same time the past has shown us that ‘unknowns’ about potential human, environmental and societal risks are raised. ... Nanotechnology has enormous technological and economic potential. To take full economic and societal advantage of these opportunities, it is crucial to understand and effectively manage potential environmental risks of [manufactured nanoparticles and nanomaterials]. Any misconception on environmental risks may seriously hamper application of nanotechnology.

(NanoNextNL, 2016)

ENGINEERED nanoparticles (ENPs) are used everywhere (Boysen, 2015). Silver nanoparticles are used in fabrics to kill bacteria, making clothing odour-resistant or in paints, inhibiting bacteria and mildew growth. Zinc oxide and titanium dioxide nanoparticles are used in sunscreens to effectively block ultraviolet rays with minimal white residue. Silica nanoparticles are used to strengthen fishing rods without increasing their weight. And what about the potential of nanoparticles to deliver drugs to tumors? Or gold nanoparticles to breakdown organic pollutants in air and iron nanoparticles to clean up polluted ground water? Or silicon nanoparticles to increase battery power and reduce recharge time? (For these and other applications see Boysen (2015).) Nanotechnology surely offers an unending list of “possibilities to increase the quality of life” with “enormous technological and economic potential” (NanoNextNL, 2016). Production and usage of ENPs, however, do not come without risk.

With this thesis, I will contribute to the methodology needed to “understand and effectively manage potential environmental [and human] risk of [manufactured nanoparticles and nanomaterials]”. I will approach nanoparticle risk assessment with probabilistic methods as already used within the environmental and food health chemical risk assessment areas. Two aspects within probabilistic chemical risk assessment that are of importance are variability (natural variation existing between species, people and environments) and uncertainty (variation existing due to lack of data or knowledge). The quantification of both these aspects will form a central part of this thesis.

In this General Introduction, I provide the necessary background and context needed

to understand this thesis, discuss current challenges related to risk assessment of nanoparticles and provide an outline for the rest of this thesis.

1.1. CHEMICAL RISK ASSESSMENT

CURRENT EU chemical risk assessment is generally based on the technical guidance documents of Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) of the European Chemical Agency (ECHA) for environmental and human health risk assessment. In addition, the European Food Safety Authority (EFSA) and World Health Organization (WHO) provide guidelines for risks in food safety. Human health and environmental risk assessment consists of three parts, namely, (i) exposure assessment, (ii) hazard assessment and (iii) risk characterisation. In the risk characterisation, the exposure and hazard are compared with one another.

In Sections 1.1.1, 1.1.2 and 1.1.3, I introduce the three parts of risk assessment. Section 1.1.4 details the difference between deterministic and probabilistic risk assessment and motivates the use of the latter. A literature review of recent probabilistic environmental risk assessment methodology is provided in Section 1.2.

1.1.1. EXPOSURE ASSESSMENT

Exposure assessment of a chemical is quantified by the exposure concentration (ExpC). The ExpC is the concentration of the chemical to which a person or animal species is exposed.

For human health exposure, there are three major exposure routes, namely inhalation, oral and dermal. Considering the scope of this thesis, only the oral route of exposure will be considered. Oral exposure is expressed as an external dose in $mg\,kg_{bw}^{-1}\,day^{-1}$ (ECHA, 2012a). For a single product, it can be calculated as

$$\frac{Q_{prod}C}{BW}$$

where Q_{prod} is the amount (in grams) of product consumed in one day, C is the concentration (in mg/g) of the substance in the product and BW (in kg) is the body weight of

the individual (ECHA, 2012a). The exposure can be aggregated over multiple products (e.g. foods) to obtain the ExpC of a particular substance.

Environmental exposure assessment is usually performed on two spatial scales: local and regional (ECHA, 2012b). For both scales, the exposure assessment results in ExpCs, referred to as predicted environmental concentrations (PECs), for various environmental compartments such as air, water, soil and sediment (ECHA, 2012b). The local PEC is mainly impacted by chemical releases from point sources and is calculated on the basis of daily release rates (EC/JRC, 2003). The regional PEC takes into account the further distribution and fate of chemicals after their release and also provides a background concentration for use in the calculation of a local PEC (EC/JRC, 2003). Multimedia fate models, such as SimpleBox (Brandes et al., 1996), are used to obtain steady-state concentrations in the various environmental compartments (ECHA, 2012b) for use in regional PECs.

1.1.2. HAZARD ASSESSMENT

Hazard assessment of a chemical is quantified by the critical effect concentration (CEC). The CEC is the concentration of a chemical above which a person or species should not be exposed. Depending on the goal of the risk assessment, the CEC can represent the highest concentration without an adverse effect (i.e. a no-effect concentration) or it can represent a concentration at which a specified adverse effect is expected or cannot be excluded. The CEC is based on a limit concentration obtained from the dose-response curve of an animal toxicity study. The limit concentration is divided by assessment factors, to account for the uncertainty surrounding the limit concentration.

In human health hazard assessment, we generally distinguish between acute and chronic risks. Until the early 1990's, the WHO considered only chronic CECs in the form of the acceptable daily intake (ADI) (Solecki et al., 2005). After addressing the issue of acute risks, the Joint FAO/WHO Meeting on Pesticide Residues came up with the acute reference dose (ARfD) (JMPR, 1999; Solecki et al., 2005). The ADI and ARfD are obtained by dividing a chronic (for ADI) or acute (for ARfD) limit concentration by two assessment factors (FAO/WHO, 2009). The limit concentration can typically be the no observed adverse effect level (NOAEL), benchmark dose (BMD), 25% tumor chronic dose rate (T25)

or 50% lethal concentration (LC50) and is obtained from an animal dose-response model (ECHA, 2012d). The two assessment factors are the interspecies and the intraspecies assessment factor. The interspecies assessment factor captures the differences in sensitivity to a substance between animals and humans. The default assumption here is that humans might be more sensitive to a substance than animals (ECHA, 2011). The intraspecies assessment factor captures the differences in sensitivity among humans which can be caused by biological factors such as gender, age, health status, genomics and nutritional status (ECHA, 2011). Historically, a value of 10 is used for each of the two assessment factors, resulting in a “margin of safety” of $10 \times 10 = 100$ (Lehman and Fitzhugh, 1954). More assessment factors can be included to account for other uncertainties such as in the experimental exposure duration, in the dose-response relationship or in the quality of the whole database (ECHA, 2011). In the REACH guidance documents, the CEC for human health risks is referred to as the chronic or acute derived no-effect level (DNEL) (ECHA, 2011).

In environmental hazard assessment, a CEC is obtained for all relevant species in an environmental compartment. The limit concentrations from which the CECs are derived are usually obtained from single species laboratory toxicity tests and reported as the concentrations at which $x\%$ (e.g. 10%) mortality or inhibition of a function (e.g. growth) was observed (ECHA, 2008b). The most frequently used limit concentrations include the 50% effect concentration (EC50), LC50, no observed effect concentration (NOEC), 10% effect concentration (EC10) and 10% lethal concentration (LC10). These limit concentrations can be obtained by various methods (ECHA, 2008b), including:

- hypothesis testing - comparing a control group with a given concentration to find significant differences
- concentration-response modelling - using regression methods to fit a concentration-response curve
- biological methods - including toxicokinetic models.

To obtain the CECs for a specific environmental compartment, we need the same limit concentration for all relevant species. If these are not available, assessment factors can be used to extrapolate, for example, from an acute to a chronic concentration or from a LC10 to a NOEC.

According to REACH guidelines, a single no-effect concentration should be derived for a given environmental compartment in the form of the predicted no-effect concentration (PNEC) (ECHA, 2011). The PNEC is the “concentration of a chemical in any compartment below which unacceptable effects on the [ecosystem of that compartment] and its organisms will most likely not occur during long term or short term exposure” (ECHA, 2011). The PNEC is derived from the CECs of the species in the given compartment by extrapolating from the laboratory tests to natural conditions using either the species sensitivity distribution (SSD) method or the assessment factor method (ECHA, 2011).

The SSD method requires CEC values for a number of species from different taxonomic groups and assume that these values follow a specific distribution (ECHA, 2008b). The SSD method aims at calculating a concentration which is assumed to protect a certain percentage of the species of the ecosystem against toxic effects (ECHA, 2011). To obtain the PNEC via this method, the following formula is used:

$$PNEC = \frac{5\%SSD}{AF}$$

where 5%SSD denotes the 5th percentile of the SSD (see the Appendix 1.A for more detail) and *AF* is the assessment factor (between 5 and 1) reflecting further uncertainties (ECHA, 2008b).

Often, however, the data requirements for the above-mentioned method are not met and the assessment factor method is used. In the assessment factor method, the most sensitive trophic group of organisms is identified from which the species with the lowest CEC is identified. This CEC, denoted $Min\{CEC\}$, is used to determine the PNEC as follows:

$$PNEC = \frac{Min\{CEC\}}{AF}$$

where *AF* is the assessment factor, the size of which depends on the type and amount of toxicity information available (ECHA, 2011). The *AF* extrapolates from single-species laboratory data to multi-species ecosystems (ECHA, 2011).

For further details on obtaining the PNEC for the various environmental compartments, the interested reader is referred to the REACH guideline documents (ECHA, 2011).

1.1.3. RISK CHARACTERISATION

The third part of the risk assessment process, after the exposure and hazard assessment have been done, is the risk characterisation. During this step, the $ExpC$ and CEC are compared with one another. This is usually done via a concentration ratio (CR):

$$CR = \frac{ExpC}{CEC}. \quad (1.1)$$

In general, a risk is controlled when the CR is below one for all exposure scenarios and all endpoints (i.e. growth, reproduction, mortality etc.) (ECHA, 2012e).

For human health, the CR can be obtained by replacing the CEC in Equation 1.1 with the $ARfD$, ADI or $DNEL$ obtained from the hazard assessment. A slightly different approach is to use the margin of exposure (MoE) which is the “ratio between a defined point on the dose-response curve for the adverse effect and the human intake” (EFSA, 2005). The EFSA adopted this approach for genotoxic and carcinogenic compounds (EFSA, 2005). The MoE is given by

$$MoE = \frac{\text{limit concentration}}{\text{human exposure}}.$$

In addition to the MoE being the reciprocal of the CR , it also omits the inter- and intraspecies assessment factors. The inter- and intraspecies differences are accounted for by requiring a MoE of at least 100 (Dekkers et al., 2011; van der Voet et al., 2009). In contrast to the general interpretation of the CR , the MoE makes no assumptions about a safe intake (EFSA, 2005) or controlled risk. Rather, it is used to consider possible safety concerns of substances present in food.

In environmental risk assessment, the CR can be used as given in Equation 1.1. More specifically, the $ExpC$ and the CEC can be replaced with the PEC and the $PNEC$, respectively. In this case, we define the risk characterisation ratio (RCR) according to REACH (ECHA, 2012e)

$$RCR = \frac{PEC}{PNEC}.$$

The RCR is calculated for all environmental compartments (e.g. water, sediment, soil etc) on both the local scale and regional scale (ECHA, 2012e).

1.1.4. FROM DETERMINISTIC TO PROBABILISTIC RISK ASSESSMENT

According to REACH, the risk assessment process follows a tiered approach (ECHA, 2012c). The tiered approach has four levels which are described in Table 1.1. Levels 0, 1 and 2 constitute different levels in a deterministic risk assessment and Level 3 constitutes a probabilistic risk assessment.

Table 1.1: Description of the four levels of the tiered approach to risk assessment (ECHA, 2012c)

Level	Description
Level 0	a point estimate is derived using agreed conservative assumptions and default values
Level 1	a qualitative uncertainty analysis can be used to refine the estimate of exposure and estimate an indicative range of unquantifiable uncertainties
Level 2	a deterministic approach uses different combinations of assumptions to make a range of point estimates, which ... provide more quantitative information about the sensitivity of the risk characterisation ratio (RCR) to specific parameterisation.
Level 3	a probability distribution is derived which provides statistical information about the likelihood that the RCR will be exceeded

As is clear from Table 1.1, probabilistic risk assessment (Level 3) is the most advanced option in the tiered approach and is, therefore, only used if all the preceding levels fail to control the risk satisfactorily. Despite this tiered approach, EU risk assessment is largely deterministic in nature (Verdonck et al., 2006). Deterministic risk assessment, however, has been criticised over the years and a quick scan of the literature points out the limitations and shortcomings of this method. In a deterministic risk assessment, exposure, hazard and risk characterisation endpoints (such as the PEC, PNEC, CR and MoE) are “considered as single, crisp values” while in reality they contain both uncertainty and variability (Verdonck et al., 2005). Uncertainty and variability are dealt with by using conservative or worst-case scenarios. This leads to three reasons why deterministic risk assessment is not desirable.

First, a risk assessment becomes unrealistic due to the additive (or multiplicative) effect of the conservative scenarios used in intermediate steps. This may result in an over-conservative risk assessment leading to unnecessarily stringent regulation on the use of nanotechnology. In his book, *The Flaw of Averages: Why We Underestimate Risk in the*

Face of Uncertainty (2012), Savage illustrates that using averages always underestimates the risk. This is because the average, being one single value, cannot account for the uncertainty that exists in the data. This relates directly to the risk assessment context where conservative values are used causing an unrealistic overestimation of the risk.

Second, the transparency of the risk assessment is compromised because it is impossible to explicitly quantify how conservative the risk assessment is when using conservative scenarios. It is not clear whether the risk endpoints (CR, PEC, PNEC etc.) represent a “reasonable worst-case” or an unrealistic, “over-protective” situation (Jager et al., 2001a). To obtain a more realistic and defensible risk assessment, worst-case thinking needs to be excluded from the risk assessment itself (Jager et al., 2001a). There is, therefore, a need for a proper and transparent treatment of sources of uncertainty and variability during the exposure, hazard and risk assessment process (Verdonck et al., 2006).

Third, in a deterministic risk assessment, it is not possible to differentiate between uncertainty and variability. Uncertainty and variability are important aspects in risk assessment and fundamentally different. Uncertainty is the, in principle, reducible variation that exists due to lack of data and knowledge. Variability, on the other hand, is the natural inherent variation that is present in all natural processes and living organisms and is not reducible. It is important to always be clear in the distinction between variability and uncertainty. Although this distinction is alluded to in deterministic uncertainty analysis (ECHA, 2012c, p. 24), it is only by using probabilistic methods that these can be given the necessary attention (see Table 1.2). In probabilistic risk assessment, the variability in ExpCs is quantified by an exposure concentration distribution (ECD) and the variability in CECs is quantified by a species sensitivity distribution (SSD).

In a proper risk assessment, the effect of uncertainty on the risk assessment needs to be studied. This is only possible if we quantify uncertainty separately from variability and clearly trace the path to its sources. A deterministic risk assessment does not allow for such a separation in a transparent way. Deterministic risk assessment, therefore, hampers focused research on areas of high uncertainty because these cannot be identified.

There is, therefore, a need to move towards a probabilistic way of dealing with the uncertainties present in the risk assessment process. In a probabilistic risk assessment, all

Table 1.2: Description of the uncertainty analysis of the top three levels of the tiered approach to risk assessment (ECHA, 2012c)

Level	Description
Level 1	In qualitative uncertainty analysis, uncertainty sources and their qualitative characterisation are identified. For each source, it is determined whether it concerns variability or uncertainty. Furthermore, the direction (whether hazard, exposure, or risk estimate is over- or underestimated) and magnitude (low, medium, or high) are determined.
Level 2	In deterministic uncertainty analysis, uncertainty sources are identified. For each source, an average and a worst-case scenario are defined. For each scenario, the risk is estimated in terms of the RCR. In this way, a range of possible RCR values is obtained. In order to distinguish between variability and uncertainty, three scenarios for each source can be defined: average (no uncertainty or variability), typical (only variability) and worst-case (variability and uncertainty).
Level 3	Probabilistic uncertainty analysis includes the probabilistic estimation of hazard, exposure, and risk as well as a sensitivity analysis. In hazard assessment, the SSD quantifies variability and confidence bands represent the uncertainty. The width of the confidence bands indicates the contribution of the sources of uncertainty. In exposure assessment, the ECD quantifies variability and confidence bands represent the uncertainty. In both hazard and exposure assessment, methods such as Monte Carlo simulation, bootstrapping, Bayesian analysis, and fuzzy methods are used in the construction of the SSD and ECD. When both hazard and exposure are treated probabilistically (i.e. with distributions describing the variability), then the RCR is no longer a single deterministic value, but a distribution from which the probability that the RCR exceeds one can be calculated. This probability is known as the expected risk and is then the only value that needs to be communicated.

the available information can be included and the necessary conservatism can be postponed to the last moment (Jager et al., 2001b).

It is our opinion that probabilistic risk assessment is preferable to deterministic quotients to address the complex problems of chemical risk assessment and risk management. (Jager et al., 2001b)

Guidelines for uncertainty analysis in environmental risk assessment are provided by REACH (ECHA, 2012c). For human health risk assessment, the WHO provides guidelines to evaluate uncertainty in hazard assessment (WHO, 2014).

1.2. PROBABILISTIC RISK ASSESSMENT IN LITERATURE

IN the previous section, I gave some background on the established method for both environmental and human health risk assessment. I now focus only on environmental risk assessment and provide an overview of existing probabilistic methods in environmental chemical risk assessment in recent literature.

A quick scan of the literature shows a wide range of methods used in risk assessment in order to account for or deal with uncertainty and variability. These methods can be grouped into three main groups:

- Possibility theory and fuzzy logic
- Probability distributions
- Bayesian methods

Possibility theory and fuzzy logic are used for uncertainty and variability that cannot be quantified. Such uncertainty and variability are expressed in “imprecise linguistic terms” such as severe, high, many, few, very true and mostly (Chen and Pham, 2001) and are referred to as “non-random” uncertainties (Zhang et al., 2009). In fuzzy logic, uncertainty is modelled as the degree of membership of an outcome in a set (Mckone and Eshpande, 2005). Such a degree of membership is described by a membership function which can take on a value from 0 (non-member) to 1 (member) (van der Werf and Zimmer, 1998). Some methods in this group include possibility distributions, belief functions, probability boxes and fuzzy functions.

The most obvious way to deal with quantifiable uncertainty and variability is to use probability distributions as is done by REACH (ECHA, 2012c). A probability distribution is a description of the possible values of a random variable and the probability of occurrence of these values (Upton and Cook, 2011). The CR (or ExpC, CEC) can take on different values because of uncertainty or variability. It is, therefore, possible to describe these possible values with a probability distribution. Probability distributions can be empirical or parametric. While empirical distributions can only be implemented via (re-)sampling methods such as Monte Carlo simulation and bootstrap, parametric distributions can also be implemented analytically if the distribution and the model in which the distribution is used is not too complex.

Bayesian methods extend the use of probability distributions by including (apart from the current data) external sources of information about the uncertainty of the parameters of the probability distribution. Such external information is incorporated via a prior distribution. The prior distribution is then updated with the current data to produce the posterior distribution of the unknown parameter. In this way, not only the data, but also other information about the uncertainty can be used to estimate a parameter.

A comprehensive literature review on probabilistic risk assessment was done by Lester et al. (2007). Although the authors focused on probabilistic risk assessment applications in management of contaminated land, they do provide a comprehensive list of methods used in the probabilistic risk assessment of exposure and hazard assessment, risk characterisation and risk management. These methods include Monte Carlo, 2D Monte Carlo (also in combination with fuzzy methods and statistical methods such as ANOVA), Bayesian methods using Markov Chain Monte Carlo (MCMC) to separate uncertainty and variability, fuzzy methods and probability bounds (Lester et al., 2007). Lester et al. (2007) covers the literature up to 2005. I will, therefore, focus on the literature from 2006 onward.

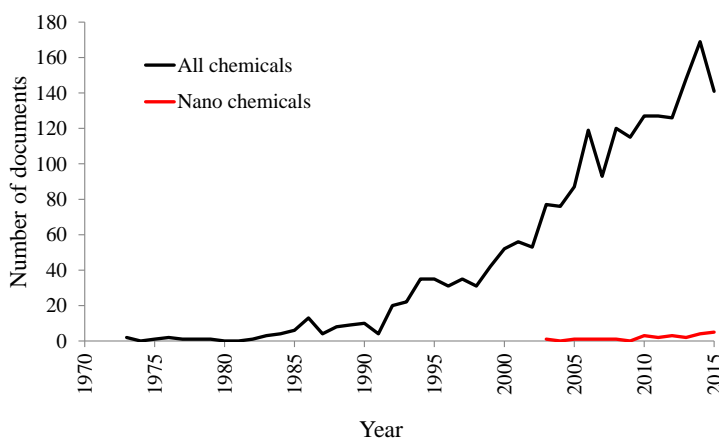


Figure 1.1: The number of documents on probabilistic environmental risk assessment reported (8 December 2015) by SCOPUS for all (black line) and nano (red line) chemicals.

A literature search (8 December 2015) was done using Scopus, an abstract and citation database covering more than 5000 international publishers. Figure 1.1 shows the number of documents found in the Scopus search on the words *probabilistic risk assessment*

*environment**. Focussing on environmental risk assessment of chemicals, I narrowed the search down by searching for *probabilistic AND risk AND assessment AND environment* AND NOT health AND NOT diet AND NOT seismic AND NOT nuclear AND NOT carcinog* AND NOT flood AND NOT collision AND NOT "oil spill" AND NOT economic** in the title, abstract and keywords from 2006 onwards. This resulted in 584 documents. Next, only the peer-reviewed articles were selected. The remaining 379 articles were screened on their titles to include only those on environmental chemical risk assessment (201 documents remaining). Based on the abstracts, 18 papers were identified that present development of methodology for probabilistic environmental risk assessment. Of these 18 papers, 7 concerned nanoparticles and will be discussed in Section 1.4. Table 1.3 provides an overview of the remaining 11 papers. In the remainder of this section, I summarise these papers, focusing on the methods used and the treatment of variability and uncertainty.

Baudrit and Dubois (2006) make a distinction between variability and uncertainty. They assume that variability can be quantified by an imprecise probability function. The imprecision or uncertainty is due to lack of knowledge. This uncertainty can be captured by one of three methods, namely possibility distributions, belief functions or probability boxes. The uncertainty method used will depend on the type of information. For example, probability boxes are appropriate to capture information about mean values. Belief functions model quantile information, while possibility measures are well-suited to represent a family of distributions whose mode is known.

Cervone et al. (2008) use machine learning methods for the risk assessment of atmospheric emissions. K-means clustering is used to group contaminant clouds based on a number of cloud attributes. Then a symbolic machine learning algorithm is used to find common patterns in the meteorological data that can be associated to each group. A new set of meteorological conditions, which displays one of the already identified patterns, can then easily be associated with a contaminant cloud group and its associated risk map. The authors do not mention anything about uncertainty or variability.

Similarly to Baudrit and Dubois (2006), Darbra et al. (2008) clearly distinguish between uncertainty (uncertainty due to incompleteness) and variability (uncertainty due to randomness). They suggest the use of probabilistic methods for variability and fuzzy-logic methods for uncertainty which is often expressed in more linguistic (more, little, tolera-

Table 1.3: The 11 papers on probabilistic methods for environmental risk assessment of conventional chemicals

Reference	Method	Variability / Uncertainty	Risk assessment part	Regulatory authority
Baudrit and Dubois (2006)	probability distributions, possibility distributions, belief functions, probability boxes	Variability: quantified by probability distributions Uncertainty: quantified by possibility distributions, belief functions or probability boxes depending on type of information	Not specifically applied, but relevant for risk assessment	*
Cervone et al. (2008)	K-means clustering, symbolic machine learning algorithm for pattern recognition in input parameters	*	classification and pattern recognition of contaminated clouds	*
Darbra et al. (2008)	probability distributions, Monte Carlo simulation, fuzzy-logic set theory	Variability: quantified by probability distributions and implemented by Monte Carlo simulation Uncertainty: quantified by membership functions in fuzzy-logic sets	Examples given in various risk assessment areas	*
Ho (2008)	probability distributions, Monte Carlo simulation	Uncertainty and variability are used interchangeably and quantified with probability distributions obtained from data, literature or expert judgement	gaseous and liquid-phase radon transport in landfills	*
Zhang et al. (2009)	Fuzzy-stochastic partial differential equations	Variability: quantified by probability distributions Uncertainty: quantified by membership functions in fuzzy-logic sets	subsurface solute transport with case study on trichloroethylene concentrations	*
Hayashi and Kashiwagi (2011)	Hierarchical Bayesian model, Monte Carlo simulation	Variability and uncertainty are referred to, but no clear distinction is made	risk characterisation of nine toxic substances in surface waters	EUFRAM (European Commission)
Schad and Schulzz (2011)	probability distributions, Monte Carlo simulation	Only variability is considered which is quantified by probability distributions and implemented by Monte Carlo simulation	risk quantification for plant protection products	European Commission

* — nothing mentioned

Continued on next page

Table 1.3 – *Continued from previous page*

Reference	Method	Variability / Uncertainty	Risk assessment part	Regulatory authority
Sala et al. (2012)	extrapolation of SSDs	Variability and uncertainty are used interchangeably and can be accounted for by means of application factors	SSD for aquatic risk assessment of insecticides.	European Commission
Ciffroy et al. (2013)	probability distributions, censored data, outlier detection	Only variability is addressed, although no clear distinction is made between variability and uncertainty	aquatic risks from metals	European Water Framework Directive
Franco et al. (2013)	probability distributions, Monte Carlo simulation	Variability and uncertainty are distinguished, but applied distributions are assumed to only quantify variability	fate and elimination of chemicals in sewage treatment plants	European Commission
Gottschalk and Nowack (2012)	empirical distribution functions, Monte Carlo simulation	Variability and uncertainty (used together, not distinguished) are quantified by uniform and triangular distributions	SSD and risk calculations for triclosan in the aquatic environment	*

* — nothing mentioned

ble etc.) terms than variability. An example of such a methodology is provided by Zhang et al. (2009).

Zhang et al. (2012) develop a fuzzy-stochastic partial differential equation model. This model is a hybrid model where parameters with random uncertainties (i.e. variability) are described by a probability distribution, while parameters with non-random uncertainties (i.e. uncertainty) are described by fuzzy functions.

Ho (2008) introduced a probabilistic framework for a radon-transport model. Uncertainty and variability (referred to interchangeably) of the model parameters are quantified by probability distributions obtained from experimental data, literature or expert judgement. The distributions are implemented via Monte Carlo simulation. A sensitivity analysis, using a stepwise linear-rank regression, is also performed.

Hayashi and Kashiwagi (2011) present a Bayesian risk assessment method. A hierarchical Bayesian model was used to estimate the ECD. With such a model it is possible to estimate within-site and between-site variations separately. Standard Bayesian estimation was used to estimate the parameters of the SSD. The expected potentially affected fraction was obtained as $P(ECD > SSD)$ and its uncertainty quantified by using the posterior distributions of the estimated parameters of the ECD and SSD as inputs for a Monte Carlo simulation.

Ciffroy et al. (2013) develop a probabilistic approach to account for data variability in predicting aquatic risks from metals. Probability density functions (PDFs) were fitted to variables used to obtain PECs and PNECs. Data reconstruction and Kaplan-Meier approaches were used to deal with the problem of non-detects. Moreover, Grubb's test was used to identify outliers. After construction of the probability density functions (pdfs), the empirical distributions for the PEC and PNEC were obtained via Monte Carlo simulation. The probability to exceed safety thresholds, $P(PEC/PNEC > 1)$, was calculated after fitting a distribution to the PEC and PNEC. The gamma and the log-normal distributions were considered.

Sala et al. (2012) developed a method to construct SSDs when there is insufficient reliable data. A "typical" SSD is constructed for a group of chemicals. An estimated SSD can then

be obtained by extrapolating the typical SSD based on an indicator species. The authors suggest the use of application factors to account for uncertainties.

Franco et al. (2013) developed a probabilistic add-on for the sewage treatment plant model SimpleTreat. Distributions applied to the input parameters were assumed to be mainly a quantification of variability. The authors mention that these distributions should only quantify variability, but that in practice it is often difficult to distinguish between uncertainty and variability. The distributions were implemented via Monte Carlo simulation.

Gottschalk and Nowack (2012) developed a probabilistic SSD method which accounts for the variability and uncertainty in species sensitivity concentrations. The variability and uncertainty within one species is quantified by triangular and uniform distributions from which an empirical distribution is obtained for each species. These single SSDs are then combined into one generic empirical SSD of a particular environmental compartment. The method is implemented in R Software making use of Monte Carlo simulation.

1.3. NANOPARTICLE RISK ASSESSMENT

HAVING provided the general context of chemical risk assessment, I will now focus on risk assessment of nanoparticles. According to the European Union, a nanoparticle is a particle with at least one of its dimensions between 1 and 100 nm (EC, 2011). While nanoparticles occur naturally, it is generally assumed that risk assessment for nanoparticles concerns engineered nanoparticles (ENPs). This assumption is the basis for all references to nanoparticles in this thesis.

Consider again the quote with which I started this General Introduction, “...it is crucial to understand and effectively manage potential ... risks of [manufactured nanoparticles and nanomaterials]”. With this in mind, government agencies, research institutes and universities are increasingly investing in research on the risks of ENPs. The question arises whether existing methods of chemical risk assessment are applicable to ENPs. Researchers voice their doubt as to whether current risk assessment regulation and legislation are applicable to ENPs.

Handy et al. (2008) claim that current methods can be used, but with modifications to

methodology. Although nanomaterials are in principle covered by existing legislation frameworks, doubts exist as to the actual applicability when it comes to ENPs (Hansen and Baun, 2012). Bowman and Hodge (2007) point out a “distinct lack of nano-specific regulation”. The same authors pointed out earlier (Hodge and Bowman, 2004) that nanotechnology “is sufficiently unlike other technologies [so as] to warrant separate consideration” with respect to regulatory frameworks (Bowman and Hodge, 2007).

In the next section, I will provide a short summary of the recent literature on methodology for risk assessment of ENPs. I will then discuss the challenges in nanoparticle risk assessment which will lead to the objectives of this thesis.

1.4. PROBABILISTIC RISK ASSESSMENT OF ENPs IN LITERATURE

THE motivation to move from deterministic to probabilistic risk assessment discussed in Section 1.1.4 holds even more for the risk assessment of nanoparticles than for conventional chemicals. This is because of the large amount of uncertainty present in the risk assessment of ENPs. This uncertainty is due to scarcity of data and lack of knowledge on the various aspects of nanoparticle risk assessment such as hazard identification, physicochemical properties, life cycle, toxicokinetics and environmental fate. Although the literature provides some evidence of probabilistic methods for the risk assessment of ENPs, the need for methods, including probabilistic methods, to address the problem of uncertainty in risk assessment of ENPs is still very topical (Koelmans et al., 2015). From the previous literature search, 7 papers concerned development of probabilistic risk assessment methodology for ENPs. I discuss these papers together with some papers that applied or extended the methodology.

Gottschalk et al. (2010a) developed a probabilistic material flow (PMF) model to assess the environmental exposure to nanoparticles. PMF considers the whole life cycle of ENPs in the environment, incorporating uncertainty and variability of all model inputs by means of probability distributions. These probability distributions, taken either from previous studies or based on assumptions, were used in the Monte Carlo simulation of the input parameters. MCMC was incorporated to illustrate how to improve and

calibrate future simulations. As more data becomes available, the prior distributions of parameters can be combined with the data in MCMC to produce more reliable results. The output of the PMF model constitutes probability distributions for the PECs in the various environmental compartments. Variability and uncertainty are defined as two different things. In the model, however, they are treated similarly and simultaneously, making it impossible to determine which part of the variation in the obtained PECs is due to variability and which part is due to uncertainty.

From the distribution of the PECs, percentiles can be obtained (Gottschalk et al., 2010b). These percentiles can then be divided by a PNEC value from literature to obtain mode and median values for the risk quotient, $RQ = \frac{PEC}{PNEC}$. In this way, the risk quotient can be obtained for the different environmental compartments for each of the environmentally relevant ENPs such as nanoTiO₂, nanoAg and carbon nanotubes (Gottschalk et al., 2010b). This mode or median risk quotient is then a result of the uncertainty and variability distributions of the input parameters in the exposure assessment. Gottschalk et al. (2011) and Sun et al. (2015) studied, besides uncertainty, the effect of spatial and temporal variability on the PECs. This is an initial attempt at studying the separate effect of variability and uncertainty on some part of the risk assessment.

Gottschalk et al. (2013a) developed a risk quantification method based on the PMF model for probabilistic exposure assessment (Gottschalk et al., 2010a) and the probabilistic SSD method (Gottschalk and Nowack, 2012) for probabilistic hazard assessment. Here again, uncertainty and variability are treated simultaneously, making separation between them in the risk quantification impossible. Recently, this risk quantification method was extended (Coll et al., 2015) to include more toxicity data and extra uncertainty on the assessment factors used in the probabilistic SSD method.

Money et al. (2012) developed the Forecasting the Impacts of Nanomaterials in the Environment model for nanoparticle risk forecasting. This model is a Bayesian network model. Bayesian networks allow the identification of the origins of uncertainty within a set of variables. The baseline model was developed based on expert elicitation, due to lack of data. The model can be updated and adapted as experimental and field data become available (Money et al., 2014). The model was applied to aquatic exposure to nanoAg. The authors do not mention variability.

Chio et al. (2012) fit a Weibull threshold model to LC_{10} cumulative distribution function (cdf) data to obtain a PNEC for Cu/Ag nanoparticle suspensions. The PEC was obtained from non-nano Cu/Ag data and partition coefficients. Using Monte Carlo simulation, a distribution for the risk quotient, $RQ = \frac{PEC}{PNEC}$, is obtained. This distribution then quantifies the uncertainty of the risk quotient. The authors do not mention variability.

Finally, Jacobs et al. (2015a) discuss the estimation of the CR, comparing empirical and parametric estimators in the case of small sample sizes and low risks. Explanation and discussion of the methods are further provided in Chapter 4 of this thesis.

1.5. CHALLENGES FOR NANOPARTICLE RISK ASSESSMENT

As mentioned in Section 1.3, risk assessment of ENPs is in principle covered by current regulatory frameworks. What then are the current challenges surrounding risk assessment of ENPs? The answer to this question is fairly simple, the amount of uncertainty. The main difference between risk assessment of conventional chemicals and that of ENPs is the amount of uncertainty. I discuss some of the possible sources of this increased uncertainty.

One of the unique properties of nanoparticles is that their physical and chemical properties depend on their size, in contrast to conventional chemicals whose physical and chemical properties do not change irrespective of size (Strambeanu et al., 2015). For example, consider surface area. For a nano-sized particle, its surface area to volume ratio is much greater than for conventional chemicals (Shegokar, 2015). This brings us to the point of the measurement metric for nanoparticles. Conventional chemicals are usually measured in mass concentration. It has been suggested, however, that for nanoparticles, mass may not be the best metric, but rather particle number or surface area (Maynard and Kuempel, 2005; Pasupuleti et al., 2012). This is of special importance to toxicity. Choosing the correct biologically effective dose metric is essential in determining the correct dose-response relation in toxicity testing (Shegokar, 2015).

Another point of concern when it comes to nanoparticle risk assessment, is the limitation of analytical techniques (Shegokar, 2015). This complicates the measuring and characterisation of nanoparticles. If nanoparticles cannot be correctly measured and

characterised, risk assessors do not know what they are measuring or if they are measuring anything at all. This leads to a lot of uncertainty.

Other sources of uncertainty include testing considerations (equipment, methodology, risk assessment procedures), assessing effects from nanoparticle exposure, characterisation of nanoparticles (inherent properties of nanoparticles and how they behave) and assessing exposure (human and environmental exposures and routes) (Grieger et al., 2009). Moreover, due to the novelty of nanotechnology and the limitations of equipment, risk assessors are often confronted with very little data on which to perform a risk assessment. Although there has been some attempts at addressing these challenges (see Section 1.4), an in-depth study into the methodology needed for a transparent risk assessment of ENPs is not yet available.

This leads to the main research question of this thesis: How can statistics assist risk assessors in dealing with the large uncertainty and small sample sizes in the risk assessment of nanoparticles?

1.6. OBJECTIVES

THE research in this thesis aims to address the research question by developing and evaluating probabilistic methods for use in risk assessment of ENPs. Before providing the objectives of this thesis, I present the probabilistic framework used in this thesis.

The variability (i.e. natural variation) in exposure and critical effect concentrations can be expressed by distributions. These variability distributions are referred to as the exposure concentration distribution (ECD) and the species sensitivity distribution (SSD), respectively. The variability quantified by the ECD and the SSD is propagated into the risk characterisation, resulting in a variability distribution of the CR. The risk probability is defined as $R = P(ExpC > CEC) = P\left(\frac{ExpC}{CEC} > 1\right) = P(CR > 1)$. Uncertainty and small sample sizes complicate the derivation of the variability distributions of the ExpCs, CECs and the CR as well as the estimation of the risk probability.

In the light of this probabilistic framework, I define the objectives of this thesis:

1. To perform a transparent risk assessment of nanoparticles in the face of large uncertainty in such a way that it can guide future research to reduce the uncertainty.
2. To evaluate empirical and parametric methods to estimate the risk probability in the case of small sample sizes.

1.7. OUTLINE OF THE THESIS

THE objectives of this thesis are addressed in the remaining chapters which are summarised below.

In **Chapter 2**, I adapt an existing integrated probabilistic risk assessment method developed for human health effects due to chemicals (van der Voet and Slob, 2007; van der Voet et al., 2009) and apply it to the case of nanosilica in food. In this method, statistical distributions and bootstrap methods are used to quantify uncertainty and variability in the risk assessment in a two-dimensional Monte Carlo algorithm. This method contributes to a better understandable risk assessment of nanosilica in food. Moreover, it allows the tracing of uncertainty sources that have a large effect on the risk assessment.

In **Chapter 3**, I develop a similar integrated probabilistic risk assessment method for environmental risk assessment of nanoparticles. Again statistical distributions and bootstrap methods are used to quantify uncertainty and variability in the risk assessment in a two-dimensional Monte Carlo algorithm. I illustrate the use of the method using a simplified aquatic risk assessment of nano TiO_2 .

In **Chapter 4**, I address the problem of small critical effect concentration sample size in the estimation of $R = P(ExpC > CEC)$ by comparing parametric estimation methods among one another and with empirical methods which use resampling. This is done under the assumption of normality for both the ECD and the SSD. I investigate the maximum gain possible when making strong parametric assumptions as opposed to making no parametric assumptions at all. Parametric estimators enable us to estimate and bound the risk for smaller sample sizes and small risks where empirical methods fail.

The normal distribution is often criticised as having light tails. In **Chapter 5**, I allow for heavier tails by using the generalised pareto distribution to estimate the tails of the ECD and the SSD. I develop a mixture model to estimate the risk, $R = P(ExpC > CEC)$, with the assumption of a normal distribution for the bulk data and generalised pareto distributions for the tails. I also study the effect of the tail index, which controls tail heaviness, on the risk value.

In the General Discussion (**Chapter 6**), I summarise and discuss the methods and conclusions of Chapters 2-5, including some detail on possible additional methods. Furthermore, I discuss the interpretation of probabilistic risk assessment in a regulatory context and provide some remaining challenges for future research.

Appendix

1.A. SPECIES SENSITIVITY DISTRIBUTION

THE SSD is an extrapolation method used to extrapolate from laboratory toxicity to realistic environmental toxicity or from single species laboratory toxicity to multi species environmental toxicity by allowing for natural variation among species (Aldenberg et al., 2002; ECHA, 2011). The SSD is the distribution of toxicity values (OECD, 1992; Traas et al., 2002). These values can, in principle, be taken to be any toxicity test endpoint (NOEC, LC_{50} , EC_{50} etc.) (Traas et al., 2002). The distribution fitted to these values are commonly taken to be the triangular distribution (Stephen et al., 1985), log-logistic distribution (Aldenberg and Slob, 1993; Kooijman, 1987; van Straalen and Den-neman, 1989), and the log-normal distribution (Aldenberg and Jaworska, 2000; Wagner and Løkke, 1991). A SSD is usually represented as a cdf in which case the y-axis is interpreted as the fraction of species affected. Figure 1.A.1 shows an example of an SSD of cadmium concentrations (Aldenberg and Jaworska, 2000). The graph depicts the cdf with a 95% confidence band around the distribution. The dots denote the data points which were used in constructing the SSD.

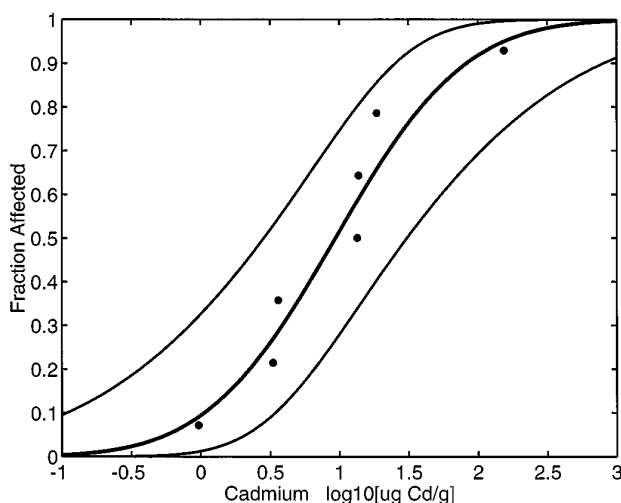
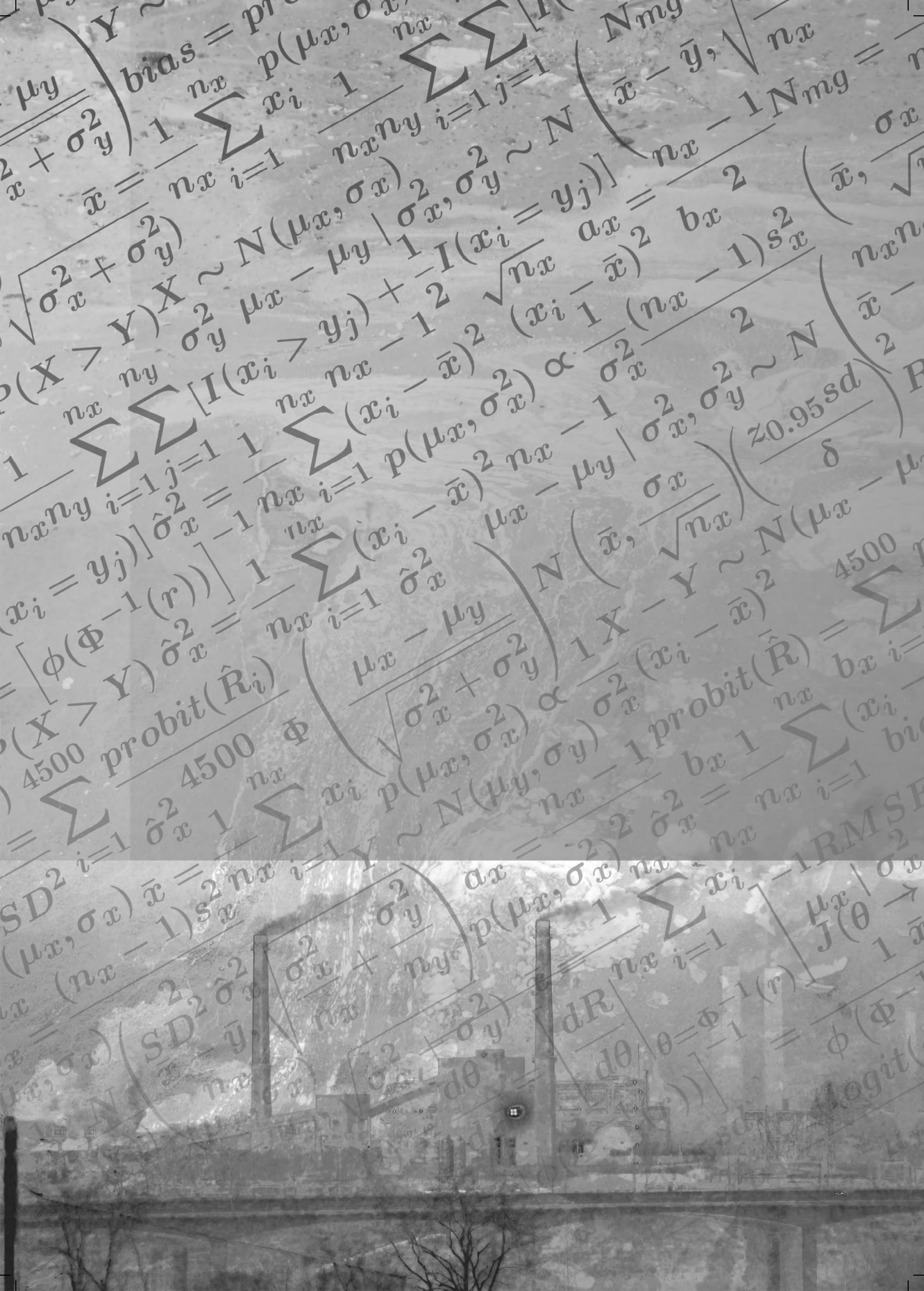


Figure 1.A.1: An SSD of cadmium concentrations (Aldenberg and Jaworska, 2000)

SSDs can be used both in a deterministic and probabilistic risk assessment. In a deterministic risk assessment, a single value (e.g. the 5th percentile) of the SSD is used to construct the PNEC, which in turn is used together with the PEC in deterministic risk characterisation (see Section 1.1.2). Alternatively, the SSD can be used in its entirety together with an exposure distribution in probabilistic risk characterisation (OECD, 1992).

SSDs can only be used under certain assumptions and data requirements. It is assumed that the distribution of species sensitivities follow a theoretical distribution function from which the group of species tested in the laboratory is a random sample (ECHA, 2008b).



2

INTEGRATED PROBABILISTIC RISK ASSESSMENT FOR NANOPARTICLES: THE CASE OF NANOSILICA IN FOOD

Rianne Jacobs, Hilko van der Voet, Cajo J.F. ter Braak

Based on: Jacobs R, van der Voet H, ter Braak CJF (2015) Integrated probabilistic risk assessment for nanoparticles: the case of nanosilica in food. *Journal of Nanoparticle Research* 17:251. doi: 10.1007/s11051-015-2911-y.

ABSTRACT

INSIGHT into risks of nanotechnology and the use of nanoparticles is an essential condition for the societal acceptance and safe use of nanotechnology. One of the problems with which the risk assessment of nanoparticles is faced is the lack of data, resulting in uncertainty in the risk assessment. We attempt to quantify some of this uncertainty by expanding a previous deterministic study on nanosilica (5-200nm) in food into a fully integrated probabilistic risk assessment. We use the integrated probabilistic risk assessment method in which statistical distributions and bootstrap methods are used to quantify uncertainty and variability in the risk assessment. Due to the large amount of uncertainty present, this probabilistic method, that separates variability from uncertainty, contributed to a better understandable risk assessment. We found that quantifying the uncertainties did not increase the perceived risk relative to the outcome of the deterministic study. We pinpointed particular aspects of the hazard characterisation that contributed most to the total uncertainty in the risk assessment, suggesting that further research would benefit most from obtaining more reliable data on those aspects.

2.1. INTRODUCTION

INSIGHT into risks of nanotechnology and the use of nanoparticles is an essential condition for the societal acceptance and safe use of nanotechnology. Many frameworks have been developed for risk assessment of nanomaterials, as were summarised by Grieger et al. (2012). These include frameworks for risk governance, risk screening, adaptable risk assessment tools and risk assessment and management frameworks. The majority of the frameworks were developed for environmental and occupational risk assessment.

One of the problems with which the risk assessment of nanoparticles is faced is the lack of data. This leads to uncertainties in the characteristics of nanomaterials, effects and exposure assessment and testing considerations (Grieger et al., 2009). Specifically within food safety, information is needed on the correct dose metric to use, the toxicokinetics of nanoparticles, the food products containing nanoparticles (Bouwmeester et al., 2009), the state of nanoparticles when manufactured and when used, and the potential for exposure (Cockburn et al., 2012). In practice, this information is hard to come by causing large uncertainty in the risk assessment (Morgan, 2005). Besides the lack of data, other sources of uncertainty can include disagreement among literature sources, linguistic imprecision or uncertainty about simplifications or models used (Morgan and Henrion, 1990). Traditionally this problem of uncertainty is solved deterministically by using worst-case scenarios. For example, a worst-case scenario could make use of a highly sensitive individual, a very high concentration of the specific chemical in the product or a very high consumption of the nano-containing foods. This method, however, compromises the transparency of the risk assessment (Jager et al., 2001b) thereby leaving risk assessors groping in the dark as to the various factors contributing to the risk assessment result. Also the deterministic method does not differentiate between variability and uncertainty. There is neither quantification of uncertainty nor of variability among individual persons regarding exposure and sensitivity. An alternative is to incorporate variabilities and uncertainties into the risk assessment quantitatively using probabilistic methods. In a probabilistic risk assessment, variabilities and uncertainties can be quantified separately and the effect of all separate sources can be seen on the risk assessment.

Within the REACH framework (REACH is a legislation of the European Chemical Agency for the Registration, Evaluation, Authorisation and Restriction of Chemicals) of tiered

risk assessment (ECHA, 2012c), probabilistic risk assessment is a possible next step when deterministic risk assessment methods indicate a potential risk. Integrated Probabilistic Risk Assessment (IPRA) was developed as a framework in which to conduct such a probabilistic risk assessment (Bosgra et al., 2009; van der Voet and Slob, 2007). It was further developed and expanded to allow for more than one adverse effect (van der Voet et al., 2009). Although IPRA was developed for the risk assessment of conventional chemicals (i.e. non-nano chemicals), it is also a potential method for the risk assessment of chemicals in nano-form.

Research on the use of probabilistic methods in the risk assessment of nanoparticles in food is scarce. A Scopus search (January 28, 2015) on the keywords “probabilistic”, “risk”, “food” and “nano”, gave 0 results. Two less restrictive searches were also performed: “probabilistic”, “risk”, and “food” (543 results) and “risk”, “food” and “nano” (146 results). Within this apparent research void, we present a case study to illustrate the use of probabilistic risk assessment in the area of nano and food.

In this chapter, we expand the deterministic study of Dekkers et al. (2011) on nanosilica in food into a fully integrated probabilistic risk assessment. In doing so, we will illustrate two points: how variability and uncertainty in a risk assessment are quantified and how to determine which sources of uncertainty have the biggest influence on the risk assessment results.

In Section 2.2, we will discuss the data and methods used in our probabilistic risk assessment. Section 2.3 provides the results. In Section 2.4, the results and limitations are discussed, followed by a short conclusion in Section 2.5.

Our example is nanosilica-containing products added to food. Food additives are generally assumed safe for human consumption. E551 is a food additive known as silicon dioxide or synthetic amorphous silica (OECD, 2004). It is mainly used as an anti-caking agent in powders or powder-like products such as soup powders, seasoning mix powders and pancake mix. The characterisation and physicochemical properties are extensively outlined in a JRC report (Rasmussen et al., 2013) and summarised by van der Zande et al. (2014). According to the Federation of European Specialty Ingredient Industries, E551 does not contain nanoparticles even though they are used in the production process (ELC, 2011). Dekkers et al. (2011), however, found silica in nano-form in food products

that contain E551. They, therefore, performed a risk assessment of nanosilica as found in E551, describing all steps of the risk assessment process.

2.2. DATA AND METHOD

2

THE risk assessment paradigm consists of three main parts: exposure assessment, hazard assessment (including hazard identification and hazard characterisation) and risk characterisation (FAO/WHO, 1995). The ratio of an estimate of tolerable exposure to an exposure estimate is termed the Margin of Exposure. This paradigm will be used when discussing the various aspects of the deterministic and probabilistic risk assessment. To provide the background from which we develop the probabilistic risk assessment, we first discuss the deterministic risk assessment done by Dekkers et al. (2011). Next, the data used for the risk assessment are described, and then the method, which makes use of the IPRA method for calculating the margin of exposure (MoE) in a probabilistic way (van der Voet and Slob, 2007). Two main aspects include the use of distributions instead of worst-case values and the separation of variability from uncertainty.

2.2.1. BACKGROUND

A deterministic risk assessment of nanosilica in food was done by Dekkers et al. (2011). By considering labels of various brands of different powder products, 27 products were identified that contained E551. These products were measured on their silica content. In 12 of these products, the amount of nanosilica (ranging from 0% to 33% of the total silica content) was also measured. In processed products, such as coffee with coffee creamer, this percentage was higher. The percentage nanosilica of total silica content in coffee with coffee creamer was 43% compared with 19% in raw coffee creamer. It was, therefore, suggested that processing increases the amount of silica in nano-form. Based on this hypothesis, a worst-case assumption of 50% nanosilica was used for the remaining 15 products for which the amount of nanosilica was not measured. The consumption of the 27 products was based on worst-case estimates (maximal consumption) made by

expert judgement. Combining the concentration with the consumption information, a worst-case exposure of $1.8 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ nanosilica was obtained.

For the hazard characterisation, Dekkers et al. (2011) used a published toxicity study on mice (So et al., 2008). This mouse study was a 10-week oral toxicity study with one control group and one treatment group. The treatment group was fed $1500 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ nanosilica. This study showed potential liver toxicity, which was seen in increased alanine aminotransferase (ALT) levels and fatty liver patterns after Hematoxylin and Eosin (H&E) staining (So et al., 2008). A Lowest Observed Adverse Effect Level (LOAEL) of $1500 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ nanosilica was derived because an adverse effect was observed at the tested dose level.

In the risk characterisation, Dekkers et al. (2011) estimated the ratio of the LOAEL to the estimated exposure, which we term the MoE, alternatively referred to as the margin of safety (MoS). A MoE less than one occurs when exposure is greater than LOAEL. The use of one as a threshold for safety is not appropriate, however, because the MoE compares human exposure with animal toxicity. Humans may or may not be more sensitive to a substance than animals. Traditionally, for conventional chemicals, an assessment factor of 10 is applied to animal toxicity to accommodate this difference (Lehman and Fitzhugh, 1954). Besides animal to human extrapolation, we also have to deal with variability in the human population itself. A sick, young or old human being will possibly be more sensitive to a substance than the average-aged healthy human being. This variation is usually also represented by a factor of 10 (Lehman and Fitzhugh, 1954). The two assessment factors result in a combined assessment factor of $10 \cdot 10 = 100$. It is, therefore, common within the risk assessment community to compare the MoE for conventional chemicals with the tolerance value of 100: a MoE greater than 100 is deemed to be safe (Lehman and Fitzhugh, 1954). In this context, the MoE (850) obtained by Dekkers et al. (2011) would be high enough for nanosilica to be judged safe. There are, however, some doubts whether this is the case. First, the scientific basis for a safety margin of 100 is unclear (ECETOC, 1995). Second, whether this value is appropriate for use in the context of nanoparticles is uncertain (Dekkers et al., 2011).

There are more uncertainties in the risk assessment of nanoparticles. Another source of uncertainty is the appropriate dose metric to use. The classical dose metric used in chemical risk assessment is mass per unit of body weight. In nanoparticle risk assess-

ment, however, this might be different. It has been suggested that for oral toxicity, particle number (N) per unit of body weight might be a more appropriate dose metric (Pasupuleti et al., 2012) and for inhalation toxicity, surface area might be a more appropriate dose metric (Maynard and Kuempel, 2005). Dekkers et al. (2011) provided the risk characterisation using both these dose metrics. To derive either of these dose metrics, one needs the particle size. This derivation adds uncertainty to the risk assessment because there is uncertainty about the particle size. This uncertainty results in uncertainty in the MoE. Instead of a single MoE value (850), a range of possible MoE values was obtained: 280 through $5600 \text{ m}^2 \text{ kg}_{bw}^{-1} \text{ day}^{-1}$ and 31 through $250000 \text{ N kg}_{bw}^{-1} \text{ day}^{-1}$ (Dekkers et al., 2011). Here we take note that the added uncertainty of the dose metric results in a potential risk (31 is less than 100). Taking into account this source of uncertainty and other possible uncertainties such as the correct assessment factor to use and the lack of toxicity data, Dekkers et al. (2011) concluded that, even in the initial case of 850, the MoE is probably not large enough to allow for all the extrapolation steps and uncertainty. This issue is the basis for the current chapter. We argue that probabilistic methods, in which uncertainty and variability are quantified as far as possible, provide a more transparent risk assessment. It decreases uncertainty about whether a certain MoE is high enough and also provides insight into which sources of uncertainty contribute most to the final risk assessment.

2.2.2. CONCENTRATION DATA

Concentration data was obtained from the Dekkers et al. (2011) study. Of the 27 products measured for total silica content, 25 products had a positive total silica concentration. Of these 25, 11 products were also measured on nanosilica content (see Table 2.1). In the deterministic study, the nanosilica concentration for the 14 products not measured on nanosilica content was taken as 50% of the total silica concentration. A recent study, however, has pointed out that this percentage is variable after consumption of the food, and can even become as high as 100% in the gastro-intestinal tract (Peters et al., 2012). This makes the measured nanosilica concentrations before consumption less relevant. We, therefore, chose not to use the nanosilica measurements, but rather the total silica measurements and model the uncertain percentage of nanosilica with a distribution. Details are explained in Section 2.2.6.

Some of the 25 silica containing products represented the same basic product, e.g. two types of instant soup powder. We identified 13 basic products with 1-5 measured products per basic product (see Table 2.1). The basic products are assumed to represent all the powder food types which contain the food additive E551.

2.2.3. CONSUMPTION DATA

The consumption data was obtained from the Dutch National Food Consumption Survey (DNFCS) of 2007-2010. This survey was conducted by the National Institute for Public Health and the Environment from 2007 to 2010 on the Dutch population aged 7 through 69 years (van Rossum et al., 2011). It contains the consumption of 3819 respondents on two non-consecutive 24-hour dietary recalls. Consumed products are identified by their NEVO code (Nederlandse Voedingsmideletabel, Netherlands Food Table) and classified into one of the EPIC-SOFT food (sub-)groups (Voss et al., 1998). Using these (sub-)groups, we identified products containing one or more of the measured products, i.e. all products that contain or have as ingredient a powdered product.

2.2.4. LINKING CONCENTRATION AND CONSUMPTION DATA

The basic products had to be linked with the consumed products from the DNFCS. This linking was done by identifying products in the DNFCS that contain or are produced from the basic product. The percentage of the basic product in the DNFCS product was determined as follows. By consulting food packaging labels, the amount of powder ingredient to produce the prepared product was determined. For complete meals that contain the prepared product, a rough estimate was made for the percentage of the meal that consists of the prepared product.

Table 2.1 provides a list of the basic products, corresponding measured products (with total silica and nanosilica concentrations), corresponding product type from the DNFCS and food composition percentages. For a detailed table listing specific products from the DNFCS, the reader is referred to Table S1 (in Supplementary Material in Jacobs et al. (2015b)).

Table 2.1: Concentration of total silica and nanosilica in measured food products and linking to Dutch National Food Consumption Survey (DNFCS) products

Basic product	Measured products	Total silica (mg/g)	Nano-silica (mg/g)	Corresponding product types from DNFCS (percent of basic product in the DNFCS product)
Sauce powder	Mix for lasagna sauce	5.4	0.3	Sauce prepared from sauce powder (18%)
	Cheese sauce	6.6		Dishes containing sauce (9%)
Meat seasoning	Minced meat seasoning mix	2.6	0.2	Pure meat seasoning (100%)
				Dishes containing meat seasoning (3%)
Cake mix	Cake with icing	0.6		Pure cake flour (100%)
				Cakes containing flour (55%)
Instant noodles	Instant noodles tandoori	12.9		Noodle dishes (3%)
	Instant noodles chicken	5.8		
Instant soup powder	Instant asparagus soup	0.6	0.2	Soup prepared from instant soup powder (6%)
	Instant beef soup	0.6		
Coffee creamers	Coffee creamer (brand a)	5.1	1.0	Pure creamer (100%)
	Coffee creamer (brand b)	4.9		Drinks containing creamer (3%)
Rubs	Spicy pepper rub	1.1	< 0.1	Pure bread crumbs (100%)
	Sweets sticky rub	6.0	0.4	Dishes containing bread crumbs (10%)
	Steak house rub	4.3	0.2	
	Roasted vegetable rub	4.9	0.6	
	Sea food rub	4.7	0.5	
International seasoning mixes	Burrito seasoning mix	7.1	0.3	Pure spice mix (100%)
	Taco seasoning mix	11.4		Dishes containing spice mix (10%)
	Guacamole seasoning mix	5.8		Dishes containing spice mix and a starch (5%)
	Nasi rames seasoning mix	6.2		
Pancake mix	Pancake mix	2.8	< 0.1	Pure pancake flour (100%)
				Pancakes containing pancake flour (32%)
Cappuccino creamer	Cappuccino foam creamer	4.9		Drinks containing cappuccino creamer (3%)
Soy shake	Soy slim shake	3.4		Drinks prepared from soy shake (8%)
Vitamin C tablets	Vitamin C	1.5		Vitamin C tablets, pills and capsules (100%)
Multivitamin Junior	Multivitamins junior (brand a)	13.7		Multivitamin junior or kid tablets, pills and capsules (100%)
	Multivitamins junior (brand b)	13.7		

2.2.5. DOSE-RESPONSE DATA

In this chapter, we use the toxicity study of van der Zande et al. (2014) as a replacement for the So et al. (2008) study used in the deterministic assessment discussed in Section 2.2.1. Van der Zande et al. (2014) performed a 28-day and a 84-day oral toxicity study in rats. The 84-day oral toxicity study showed increased liver fibrosis at high doses of nano silica. This was investigated by considering 10 slides of liver cells per rat. The occurrence and severity of fibrosis was noted. In the 84-day oral toxicity study, 15 rats were divided into one control and two treatment groups. In treatment 1, 5 rats were fed Synthetic Amorphous Silica (SAS) at an intended dose of $2500 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$. In treatment 2, 5 rats were fed pyrogenic NM-202 (the OECD representative nanostructured

silica for applications related to food (van der Zande et al., 2014)) at an intended dose of $1000 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$. After preparing the food, the realised concentration of silica in the nano-sized range (5-200 nm) was measured. The concentrations were $0 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ for the control, $819 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ for the SAS treatment and $810 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$ for NM-202. We assumed the two treatments to be two doses of the same substance because no or minor differences were found between SAS and NM-202 in powdered form, after in vitro digestion and in intestinal solubility (van der Zande et al., 2014). In this way, we have three dose groups which we can use in a dose-response model. This is described in Section 2.2.6. We considered chronic risk assessment because liver toxicity is a form of chronic toxicity. To this end, we introduced a subchronic-to-chronic extrapolation factor for the toxicity study which is discussed in Section 2.2.6.

2.2.6. METHOD

We redid the deterministic study based on new toxicity data and recent research on nanosilica fractions and performed a probabilistic risk assessment using IPRA.

In the deterministic study, we replaced the toxicity data of So et al. (2008) with that of van der Zande et al. (2014). Moreover, we replaced the 50% nanosilica percentage by 100% as a new worst case estimate.

IPRA uses a two-dimensional Monte Carlo scheme to quantify uncertainty and variability distributions separately in the risk assessment as illustrated in Figure 2.1. Details on the various aspects of the model are discussed in the next two sections.

First, we describe the quantification of variability in IPRA. Second, we describe how sources of uncertainty for the nanosilica case are quantified.

QUANTIFYING VARIABILITY IN IPRA

Following the basic three-component structure of risk assessment, we discuss exposure assessment, hazard assessment and risk characterisation as part of the IPRA method (variability loop of Figure 2.1).

Exposure is probabilistically expressed as the Individual-day Exposure (IDEXP), which is the nanosilica intake by an individual on one day. The IDEXP is calculated using the

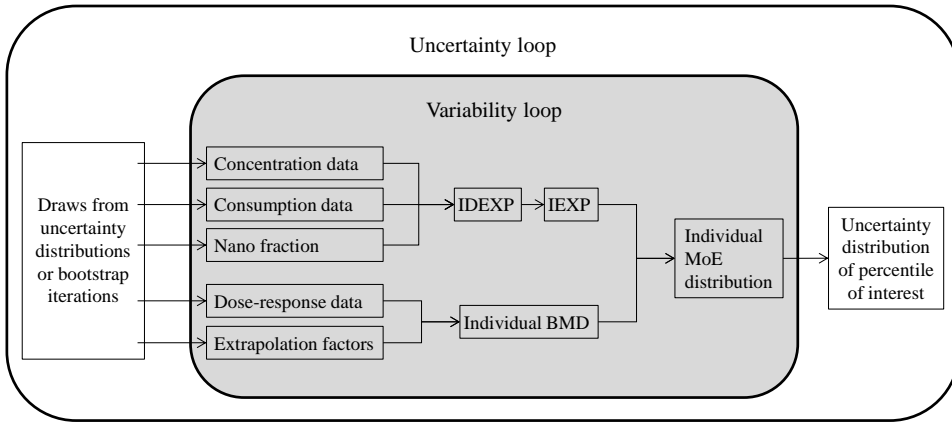


Figure 2.1: A schematic diagram of uncertainty and variability loops in the two-dimensional Monte Carlo scheme used in IPRA (IDEXP = Individual-day Exposure; IEXP = Individual Exposure; BMD = Benchmark Dose; MoE = Margin of Exposure).

formula

$$IDEXP = \sum_{k=1}^p CONS_k \cdot CONC_k \quad (2.1)$$

with $CONC_k = F \cdot C_k$, where $CONS_k$ is the consumption of product k (in $g \cdot kg_{bw}^{-1}$), C_k is the concentration of nanosilica in product k (in $mg \cdot kg^{-1}$) and F is an optional factor that can be added to allow for changes in concentration and/or specific sources of uncertainty. IDEXP is then the individual exposure to nanosilica in $\mu g \cdot kg_{bw}^{-1}$. The IDEXP distribution is obtained by calculating the IDEXP for each of the person-days for which consumption data are available. This distribution represents variability in individual human intake at the person-day level. Because we consider chronic toxicity, we need long term exposure. For this, the person-day level intake (IDEXP) distribution needs to be converted into a distribution of individual long term exposures (Individual Exposure (IEXP)). This conversion is done using the NCI model (Tooze et al., 2006), also known as the LogisticNormal-Normal (LNN) shrinkage model, as detailed in MCRA (2013).

The factor F , from now on referred to as the nano fraction, is used to convert silica concentration to nanosilica concentration. Due to the chronic nature of the risk assessment, any variability in the nano fraction will be averaged out over the long term. This means that we are only interested in a nominal nano fraction value and its uncertainty. We

chose 50% as a nominal value for the percentage of nanosilica in silica, i.e. $F = 0.5$. This nominal value was multiplied with C_k in Equation 2.1 to produce a value for the nanosilica concentration in the basic product as found in the consumed product. The (large) uncertainty associated with F is discussed in the next section.

The above describes the method when using the dose metric, $mg\ kg_{bw}^{-1}\ day^{-1}$. We, however, also consider the dose metric $N\ kg_{bw}^{-1}\ day^{-1}$. Consumed nanosilica particles vary in diameter from 5 nm through 200 nm (Dekkers et al., 2011; Peters et al., 2012). We assume that with each consumption moment a representative sample of particle sizes are consumed. Moreover, because we consider chronic exposure, consumption moments are modelled over time, averaging out any particle size effect. This means that we are only interested in the median particle size. Considering various studies, we assumed a median particle size of 100 nm (Peters et al., 2012; van der Zande et al., 2014). This particle size is used to calculate the number of particles per mass unit (see Appendix 2.A for details). For each of the Monte Carlo iterations, the number of particles per mass unit is multiplied with the concentration, C_k of that iteration to produce an exposure value of nanosilica in $10^{12}\ N\ kg_{bw}^{-1}\ day^{-1}$. For convenience, we divided N by 10^{12} , because N is very large.

Hazard is expressed as the Individual Benchmark Dose (IBMD), which is the dose at which an individual human experiences a predefined response to a substance (higher IBMD means lower hazard). Starting from a BMD_{animal} obtained from a dose-response modelling of data from an animal study, the IBMD is calculated using the formula

$$IBMD = \frac{BMD_{animal}}{EF_{chronic} \cdot EF_{inter} \cdot IEF_{intra}}$$

where $EF_{chronic}$ is the best estimate for an extrapolation factor for subchronic-to-chronic extrapolation, EF_{inter} is an extrapolation factor for the interspecies conversion (from the average animal to the average human) and IEF_{intra} is an individual extrapolation factor for the intraspecies variation (deviation from the average human).

To obtain the BMD_{animal} , we modelled the dose-response data using PROAST software (Slob, 2002; Slob and Cotton, 2013). The effect which we used to quantify the toxicity of nanosilica was the number of positive fibrosis slides (out of 10) per rat. This effect variable has a Binomial(10, p) distribution for each rat, with p the probability of obtaining

Table 2.2: Dose-response models with calculated log-likelihood, AIC and BMD. The null and full model are added for comparison.

Model	Number of parameters	Log-likelihood	AIC	BMD
NULL	2	-86.99	177.98	NA
Full	4	-80.45	168.9	NA
Linearised two-stage	4	-81.03	170.06	1300
Log-logistic	4	-81.02	170.04	630
Weibull	4	-81.03	170.06	1330
Log-probit	4	-81.01	170.02	2650
Gamma	4	-81.02	170.04	1500
Logistic	3	-81.05	168.1	1160
Probit	3	-81.06	168.12	1450
NA – Not Applicable				

a positive fibrosis slide for that rat. We modelled this probability using seven different models: Logistic, probit, log-logistic, log-probit, Weibull, gamma and linearised two-stage (Barlow et al., 2009). All seven models passed the goodness of fit test at a 5% level of significance. For each model, PROAST calculated a BMD as the ED_{50} , which is the only relevant statistic from a dose-response curve on quantal data when the variation around the ED_{50} does not represent true inter-personal differences in response but only differences between laboratory animals and other measurement errors (Slob, 2002; Slob et al., 2014). We also calculated the Akaike Information Criterion (AIC) for each model to determine the best model. From Table 2.2, we see that the logistic model is the best fitting model (i.e. lowest AIC), although the differences are small. From this model (Figure 2.2), we obtained the nominal BMD_{animal} of $1160 \text{ mg kg}_{bw}^{-1} \text{ day}^{-1}$. Model uncertainty is discussed in the next section.

The $EF_{chronic}$ accounts for the extrapolation from a subchronic study to a chronic risk. Bokkers and Slob (2005) studied such extrapolation based on the no observed adverse effect level (NOAEL) and the benchmark approach. Not all studies which derive a NOAEL, however, are suitable for benchmark modelling. A more recent study, therefore, only used the NOAEL approach (Batke et al., 2011). We, therefore, used the data collected by Batke et al. (2011) to obtain a median $EF_{chronic}$ of 1.475.

The EF_{inter} accounts for the extrapolation from the average animal to the average human. Interspecies differences can be quantified in different ways, such as extrapola-

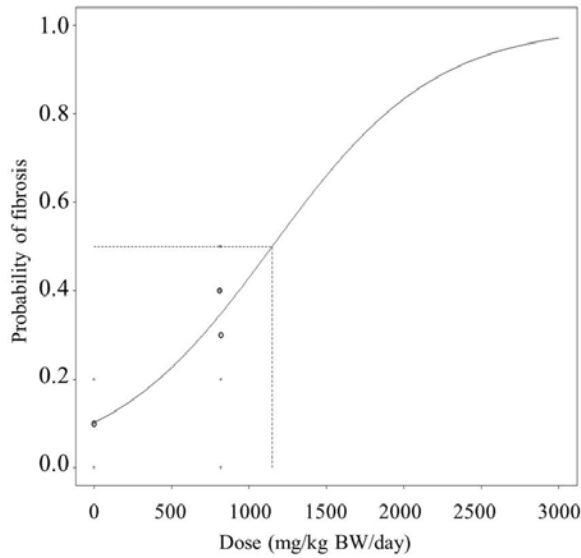


Figure 2.2: Scatterplot of the dose-response data and fitted logistic regression model for dosages of 0, 810 and 819 $\text{mg kg}_{bw}^{-1} \text{ day}^{-1}$ of nanosilica. The response is the probability of having a positive (as defined by van der Zande et al. (2014)) fibrosis slide out of 10 liver cell slides. The vertical dashed line indicates the BMD. The circles indicate the mean response of each dose group.

tion based on body weight, caloric demand or surface area (Vermeire et al., 1999). We used caloric demand because it is a “biologically plausible and data-based extrapolation method applicable to a wide range of chemical substances” (Schneider et al., 2004) and is preferred above body weight scaling (Vermeire et al., 1999). In allometric body weight scaling based on caloric demand, the ratio of animal to human dose is equal to the ratio of animal body weight to human body weight raised to the power 0.75 (Vermeire et al., 1999). The EF_{inter} based on the test species used, in our case the rat, is obtained as

$$\begin{aligned}
 EF_{inter} &= \frac{dose\ rate_{rat}}{dose\ rate_{human}} = \frac{dose_{rat}/bw_{rat}}{dose_{human}/bw_{human}} = \frac{dose_{rat}}{dose_{human}} \frac{bw_{human}}{bw_{rat}} \\
 &= \left(\frac{bw_{rat}}{bw_{human}} \right)^{0.75} \frac{bw_{human}}{bw_{rat}} = \left(\frac{bw_{human}}{bw_{rat}} \right)^{0.25} \\
 &= \left(\frac{70}{0.25} \right)^{0.25} \approx 4.
 \end{aligned}$$

We used the average body weight values as given by Vermeire et al. (1999).

The IEF_{intra} accounts for variability that exists within the human population. We obtained the distribution describing this variability using the method of van der Voet et al. (2009) and the assumption that the 95th percentile sensitive person is 2-10 times more sensitive than the average person (the range describing uncertainty). The variability distribution is obtained as a log-normal distribution with geometric mean equal to one and a geometric standard deviation of 1.91 by simultaneously accounting for the variability and the uncertainty (see next section for further details).

To convert the dose metric of the IBMD, the number of particles per mass unit as obtained previously was multiplied by the BMD of 1160 to obtain a BMD in $10^{12} \text{ Nkg}_{bw}^{-1} \text{ day}^{-1}$.

Finally, in the risk characterisation part, we obtain the distribution of the Individual Margin of Exposure (IMoE) by combining independent draws of IBMD and IEXP

$$IMoE = \frac{IBMD}{IEXP}.$$

A person is at risk when his/her exposure is greater than his/her critical effect dose. Hence, an individual is at risk when $IMoE < 1$.

QUANTIFYING UNCERTAINTY IN IPRA

The method explained above is a probabilistic risk assessment that accounts for variability that is present in the human population (the variability, inner, loop of Fig. 1). The next step is to account for uncertainty (outer loop of Figure 2.1). Again, we will consider exposure and hazard separately.

For exposure, we need to consider the consumption data, concentration data and nano fraction. Sampling uncertainty in consumption data was quantified by bootstrapping the data at the level of individual persons (500 iterations). We quantified the uncertainty in the concentration data as far as possible by bootstrapping the repeated measurements per basic product. Due to the small number of measurements, however, this probably underestimates the uncertainty. This limitation is further discussed in Section 2.4.1.

Uncertainty in the nano fraction, F , was quantified by a statistical distribution. As mentioned in Section 2.2.2, according to recent research the nano fraction, F , can be up to 1 (100%). This uncertainty about the nano fraction is modelled by a logistic-normal dis-

tribution such that its 50th percentile (p50) is equal to 0.5 and its 95th percentile (p95) is equal to 0.8. The logistic-normal distribution was chosen because resulting values of F are fractions bounded by $0 < F < 1$. The probability density function of a logistic-normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\text{logit}(x) - \mu)^2}{2\sigma^2}\right) \frac{1}{x(1-x)}$$

where $0 < x < 1$ and $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ (Aitchison and Shen, 1980). This distribution is denoted by $F \sim \text{Logistic-Normal}(\mu, \sigma)$. Considering various distributional shapes (see Figure S1 in Supplementary Material in Jacobs et al. (2015b)), we chose p95 = 0.8. This allows for at least 1% chance of values greater than 0.85 and at least 1% chance of values less than 0.15.

In each uncertainty iteration, an IEXP distribution was calculated using inputs from the bootstrapped data and uncertainty distributions. In this way, uncertainty in the IEXP distribution was quantified.

For hazard, we need to consider uncertainty in the BMD_{animal} , $EF_{chronic}$, EF_{inter} and IEF_{intra} . We first consider the BMD_{animal} . The BMD_{animal} is subject to two main sources of uncertainty: limitations of the dose-response data and model uncertainty. These uncertainties were quantified by generating 100 datasets from each of the seven dose-response models (see previous section) by the parametric bootstrap method (Efron and Tibshirani, 1993, p53; Moerbeek et al., 2004). From these, the BMDs were calculated and all 700 BMD values pooled into one set. This set now includes uncertainties from the limitations of the dose-response data and model uncertainty.

The uncertainty in the $EF_{chronic}$ was quantified by a log-normal distribution with expected value 1.80 and standard deviation 1.52 (Batke et al., 2011).

From the previous section, we calculated the nominal EF_{inter} as four. This value, however, is uncertain due to uncertainty about potential toxicokinetic and toxicodynamic differences between rats and humans. This uncertainty was accounted for by deriving an uncertainty distribution around EF_{inter} . This derivation was done assuming the 99th percentile (p99) of the uncertainty distribution to equal 10 (Slob and Pieters, 1998). With

four as the geometric mean of the log-normal distribution, a geometric standard deviation (1.48) for the uncertainty distribution was obtained.

The IEF_{intra} was assumed to contain both variability and uncertainty. The variability distribution was already given in the previous section. The uncertainty in the geometric standard deviation was quantified by a χ^2 -distribution. By setting the 2.5th percentile (p2.5) of the uncertainty distribution equal to 2 and the 97.5th percentile (p97.5) equal to 10, we are able to solve for the number of degrees of freedom for the χ^2 -distribution (van der Voet et al., 2009).

By drawing an uncertainty value for each of the BMD_{animal} , $EF_{chronic}$, EF_{inter} and IEF_{intra} , we calculated the IBMD. Repeating this process for 500 uncertainty iterations, we quantified the uncertainty in the IBMD distribution. In the risk characterisation, independent draws from the uncertainty distributions of the IEXP and the IBMD were combined into the IMoE. In this way, uncertainty in the IMoE distribution was quantified from which the uncertainty distribution of percentiles was obtained.

A simple graphical representation of both variability and uncertainty of the IMoE can be given in the form of a so-called IMoE bar graph. In an IMoE bar, a box represents the variability distribution of the IMoE between specified percentiles. These can be more or less extreme percentiles (denoted by px for the xth percentile), e.g. p0.1 and p99.9, p1 and p99, or p5 and p95, depending on the level of protection required. Whiskers are used to represent one-sided uncertainty of these percentiles. We chose uncertainty limits such, that the left whisker represents the lower 5% uncertainty bound of the lower percentile and the right whisker represents the upper 95% uncertainty bound of the upper percentile.

In addition to considering the effect of uncertainty on the final risk assessment, it is important to identify the extent to which sources of uncertainty contribute to the total uncertainty present in a certain percentile of interest. In IPRA, a probabilistic sensitivity analysis was implemented to this effect, following the method of van der Voet and Slob (2007). In short, seven sources of uncertainty were considered, namely, consumption data, concentration data, nano fraction, benchmark dose (BMD), subchronic-to-chronic factor, interspecies factor and intraspecies factor. These seven sources of uncertainty result in a full 2^7 factorial design where sampling from the uncertainty distribution for each

source is turned on and off. For each of the uncertainty iterations, $2^7 = 128$ values are obtained for a given percentile. In this way, 128 distributions are obtained each of which is summarised by its variance. An additive model is then fitted to the 128 variances. When this model explains most of the variance, which is usually the case, the coefficients of the main effects can indicate the contribution to the total variance. The intercept term estimates uncertainty when all specified seven sources of uncertainty are turned off and represents uncertainty due to Monte Carlo calculations. Results are illustrated by means of a bar graph, showing the percentage contribution of each source of uncertainty.

2.3. RESULTS

FIGURE 2.3 illustrates the IEXP and the IBMD distributions. The IEXP distribution is plotted as an exceedance curve indicating the percentage of the population that exceeds the exposure value on the x-axis. The IEXP curve starts at 1, indicating that all individuals have some intake of nanosilica on the long run. This is due to the nature of a chronic risk assessment. The amount of overlap of the curves, which indicates the amount of risk present and is related to the expected risk concept of van Straalen (2002), appears to be small.

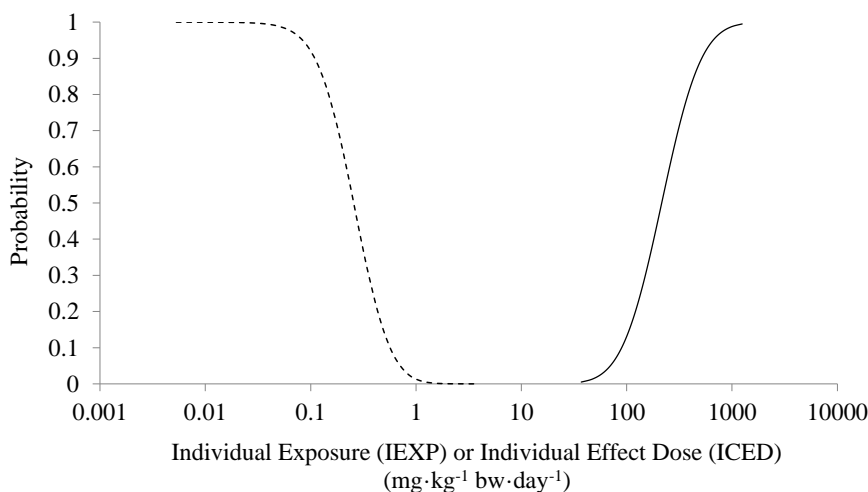


Figure 2.3: The IEXP exceedance distribution (dashed curve) and IBMD cumulative distribution (solid curve). The amount of overlap of the curves indicates the amount of risk present.

Figure 2.4 gives the results of the deterministic and probabilistic risk assessments. Figure 2.4a-d provide the deterministic MoE* values. Note that the deterministic MoE values were divided by 100 to make them comparable with the IPRA results ($\text{MoE}^* = \text{MoE}/100$). Figure 2.4a ($\text{MoE}^* = 8.5$) and Figure 2.4b ($\text{MoE}^* = (0.31 - 2500)$) are calculated according to Dekkers et al. (2011). Figure 2.4c ($\text{MoE}^* = 2.94$) and Figure 2.4d ($\text{MoE}^* = (0.11 - 8736.36)$) are calculated using the more recent data as detailed in Section 2.2.6. As discussed in Section 2.2, the MoE* when using $\text{mg kg}_{bw}^{-1} \text{day}^{-1}$ seems to be on the safe side. The lower limits of the MoE* when using $\text{N kg}_{bw}^{-1} \text{day}^{-1}$, however, give some cause for concern, because they are less than one.

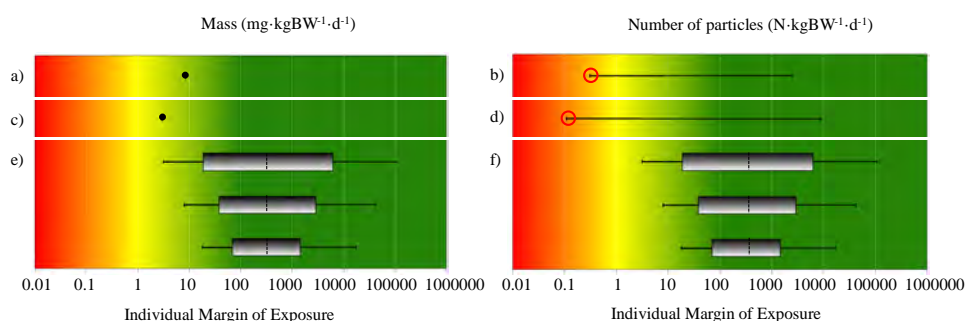


Figure 2.4: Deterministic estimates $\text{MoE}^* = \text{MoE}/100$ in $\text{mg kg}_{bw}^{-1} \text{day}^{-1}$ (a,c) and in $\text{N kg}_{bw}^{-1} \text{day}^{-1}$ (b,d) according to Dekkers et al. (2011) (a,b) and calculated with more recent data (c,d) and IMoE bars illustrating the variability and uncertainty distributions of the IMoE (e,f). The three boxes in each plot represent IMoE distributions between p0.1 - p99.9, p1 - p99 and p5 - p95, respectively, in $\text{mg kg}_{bw}^{-1} \text{day}^{-1}$ (e) and in $\text{N kg}_{bw}^{-1} \text{day}^{-1}$ (f). In each box, the left whisker represents the lower 5% uncertainty bound of the lower percentile. The right whisker represents the upper 95% uncertainty bound of the upper percentile. The dashed line indicates the median of the IMoE variability distribution. Background colouring visualises the transition from high (red) to low risk (green).

Figure 2.4e provides the probabilistic equivalent of Figure 2.4c and Figure 2.4f that of Figure 2.4d. IMoE bars are shown below one another for p0.1 - p99.9, p1 - p99 and p5 - p95. Changing to a probabilistic approach, including quantified uncertainties, does not lead to a higher perceived risk, but rather tends to confirm that the results from the deterministic study are worst-case. Moreover, considering the uncertainty in particle size in a probabilistic way, has removed the concern of possible risk in the deterministic case. Comparing Figure 2.4f to Figure 2.4e, we see that the different dose metrics give comparable results, in contrast to the deterministic study. This difference between the deterministic study and our probabilistic study is due to the fact that in the deter-

ministic study all extremes in particle size were included in the calculations while in the probabilistic study, consumers are expected to be confronted with the whole distribution of particle sizes, and therefore the particle size effect cancels out in a chronic risk assessment.

For any choice of percentile of the IMoE distribution as the main indicator of a possible risk, it is important to determine which sources of uncertainty influence its value. As an example we will consider p1, which is the left side of the middle box in the IMoE bars in Figure 2.4e and Figure 2.4f. Figure 2.5 illustrates the contribution of each source of uncertainty to the total uncertainty of the p1 of the variability distribution of the IMoE. The highest contributions to uncertainty of p1 are the BMD, subchronic-to-chronic extrapolation factor and intraspecies factor, together accounting for more than 80% of the total uncertainty. A similar pattern was obtained for p0.1 and p5 (not shown).

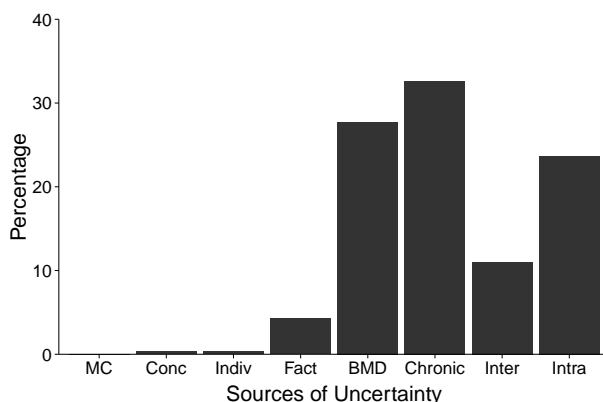


Figure 2.5: Bar graph illustrating the relative contribution of each source of uncertainty to the total uncertainty of the 1st percentile of the IMoE distribution. The sources of uncertainty are MC: Monte Carlo uncertainty, Conc: concentration uncertainty, Indiv: consumption uncertainty, Fact: Nano fraction, F , uncertainty, BMD: BMD_{animal} uncertainty, Chronic: $EF_{chronic}$ uncertainty, Inter: EF_{inter} uncertainty and Intra: IEF_{intra} uncertainty. The variance explained by the additive model is 99.38%.

2.4. DISCUSSION

ONE of the main features of IPRA is its ability to quantify uncertainty and variability separately. This is important because uncertainty is in principle reducible, where-

as variability is not. Uncertainty is variation existing due to lack of data or information. In this chapter, the quantified uncertainty gave better insight about whether the MoE obtained from the deterministic study was large enough, given the large uncertainties. We conclude that this is the case, because our results indicate that even with the uncertainty accounted for, we still end up on the safe side. We cannot claim, however, that nanosilica in food poses no risk, because certain uncertainties were not quantified. This is discussed in the next subsection.

One of the features of IPRA is the ability to study the contribution of each source of uncertainty to the total uncertainty. Knowledge of the most influential sources of uncertainty can focus research to those areas in order to decrease this uncertainty. We found that uncertainty in the BMD, the subchronic-to-chronic extrapolation factor and the intraspecies factor contributed most to the total uncertainty in the risk assessment. Better dose-response data and more research on dose-response model choice, subchronic-to-chronic extrapolation and on the variability of human sensitivity would contribute much to reducing the total uncertainty.

To decrease the uncertainty in the identified sources, more research is needed to obtain more data and better understanding of the data. More research in the form of experiments can be expensive. An additional method for obtaining data is expert elicitation (Linkov et al., 2009). In such a method, experts are asked to give their opinion on a certain variable. These opinions can be used to obtain a first impression about the uncertainty of that variable. This prior knowledge can then be used to improve on the uncertainty distributions which are fitted to the variables in a probabilistic risk assessment, possibly in a Bayesian context. One such an approach was illustrated in the development of a multi-criteria decision model based on expert judgement (Flari et al., 2011).

2.4.1. LIMITATIONS

Although we replaced many worst-case assumptions where possible with statistical distributions, we still made worst-case assumptions in some instances. These include the decisions on which consumed products contain the measured products and how much of the measured product they contain. The description of products in the DNFCs and the description of measured products are not detailed enough to be able to make perfect

matches. We, therefore, assumed that all DNFCs products mentioned in Table S1 (see Supplementary Material in Jacobs et al. (2015b)) were prepared using the basic product mentioned in the first column. This assumption, of course, is a worst-case assumption because DNFCs products could well have been alternatively prepared or ready-made. On the other hand, it is possible that there are other food products that also contain E551. In that case, we could have underestimated exposure. For some food composition percentages, we assumed a worst-case scenario. For instance, dishes containing sauce were assumed to consist 50% of sauce (e.g. minced meat and white sauce).

The worst-case scenarios could have been replaced by distributions describing the uncertainty. This could be done by including more factors in Equation 2.1 to quantify the various uncertainties. These factors would then represent uncertainty about whether we included all products that contain E551, about the food composition percentages and about the increase of nanosilica in processed products.

Another concern could be the choice of 100 nm for the median particle size. This choice was based on studies (Peters et al., 2012; van der Zande et al., 2014) in which particle sizes were measured. Due to the limitations of the instrumentation and the difficulty in measuring single particles in a large agglomeration such as found in food products, a perfect estimate for the median particle size was not possible. Further research might increase our knowledge.

The contribution of the uncertainty of the nano fraction is directly related to the uncertainty distribution chosen. The wider this distribution, the more uncertainty we provide. Our choice was explained in Section 2.2.6, but other choices are possible. Our results, therefore, should be interpreted with this in mind. Choosing a different value for p50 and p95 could have resulted in a different situation, especially in the sensitivity analysis as illustrated in Figure 2.5.

It should be noted that not all sources of uncertainty were quantified. The uncertainty in silica concentration was not well quantified because of the very limited number of measurements (only 1 to 5 measurements per food product). Another unquantified uncertainty is whether the SAS and NM-202 used in the toxicity study are comparable to the nanosilica found in food for human consumption. A recent study questioned the use of SAS for risk assessment of nanosilica in food (Dekkers et al., 2013). Further unquantified

fied uncertainties could be related to limited toxicity data (only one study with 2 dose groups), technical limitations in measuring nanosilica concentrations in toxicity studies and exposure assessment, and possible interaction with other (nano)substances which could increase or decrease the toxic effect of nanosilica. More research is, therefore, necessary to find and reduce other uncertainties that were not quantified. We suggest that a probabilistic risk assessment be part of a larger risk assessment framework in which possible other aspects of risk or unquantifiable uncertainties are dealt with qualitatively.

Finally, there are limitations to the application of current risk assessment methods to nanoparticles. A case-study on nanosilver questioned whether the REACH requirement of demonstrating 'safe use' of a substance is currently possible for nanomaterials (Pronk et al., 2009). To support a full risk assessment for regulatory needs, further research is necessary in generating high quality data and developing methodologies (Peters, 2011).

2.5. CONCLUSION

IN this chapter, we expanded the deterministic risk assessment of Dekkers et al. (2011) into a fully integrated probabilistic risk assessment. The overall result of the probabilistic analysis of the risk of nanosilica in food products shows, with 95% confidence, that at least 99% of the Dutch population would experience no risk. This conclusion is similar for the two metrics chosen, in contrast to the results of Dekkers et al. (2011) who indicated a much lower possible MoE on the particles scale. We do need to caution that this risk assessment is not 100% comprehensive and should only be seen as an illustrative exercise. The results are not intended as an authoritative risk assessment on nanosilica.

Approaching this risk assessment from a probabilistic method, we obtained a more transparent picture of risks and a better insight into the extent that uncertainty plays in the risk assessment process. We, therefore, conclude that in cases where deterministic methods show possible risks in a lower-tier assessment and different sources of variability and uncertainty are known to influence the results, probabilistic methods of risk assessment are preferable.

Appendix

2

2.A. DOSE METRIC CALCULATIONS

IN this section, we detail the calculations for obtaining a different dose metric. To convert $mg\ kg_{bw}^{-1}\ day^{-1}$ to number of particles, we use the particle diameter and the density of nanosilica, assuming that particles are solid spheres.

1. Calculate the volume of one particle (in nm^3):

$$v_{particle} = \frac{4}{3}\pi r^3$$

with r the radius of a particle in nm.

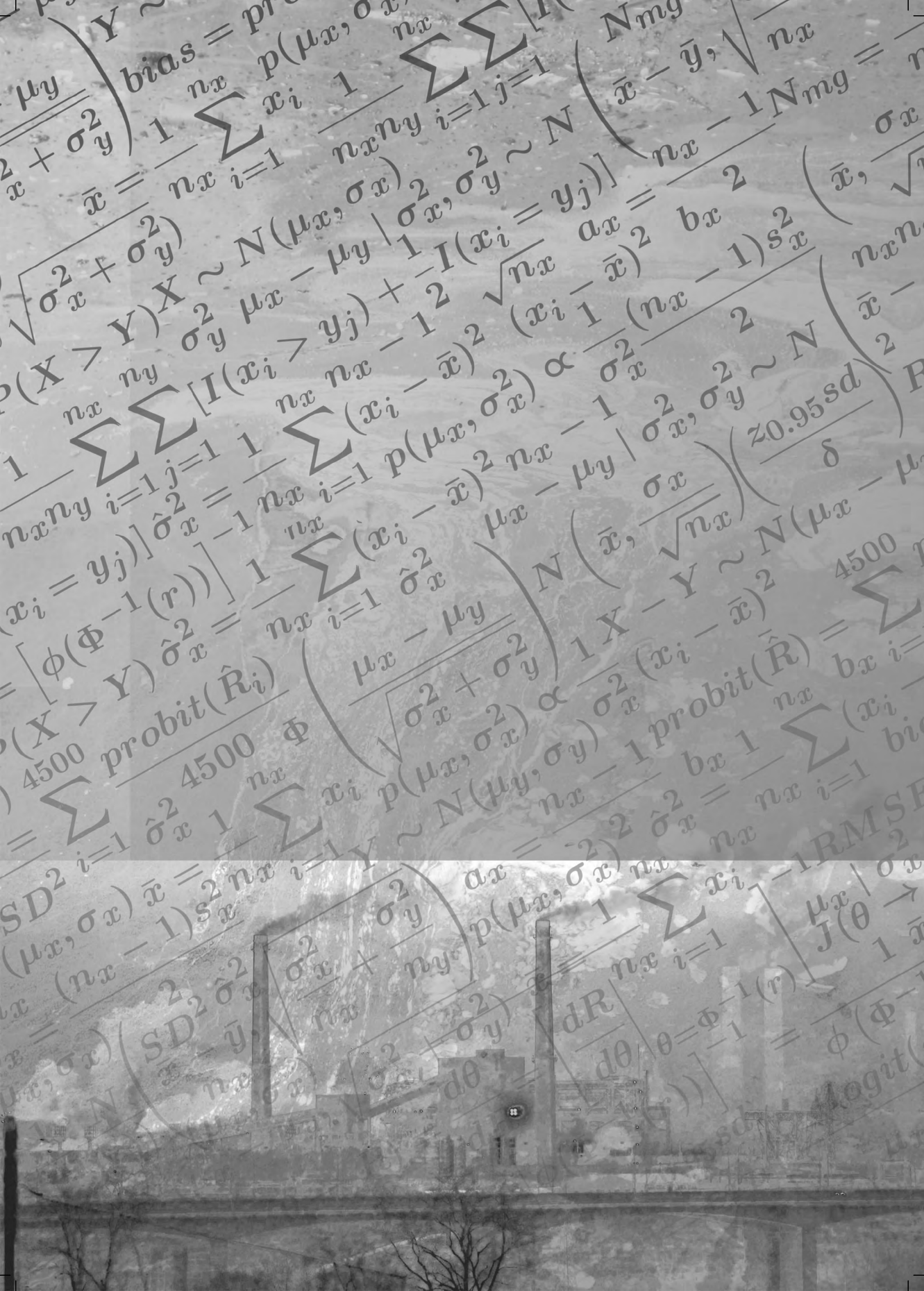
2. Calculate the mass of one particle (in mg):

$$m_{particle} = v_{particle} \cdot d_{nanosilica}$$

with $d_{nanosilica}$ the density of nanosilica in mg/nm^3 .

3. Calculate the total number of particles per mg:

$$N_{mg} = \frac{1}{m_{particle}}$$



3

COMBINING EXPOSURE AND EFFECT MODELLING INTO AN INTEGRATED PROBABILISTIC ENVIRONMENTAL RISK ASSESSMENT FOR NANOPARTICLES

**Rianne Jacobs, Johannes A.J. Meesters, Cajo J.F. ter Braak, Dik van de
Meent, Hilko van der Voet**

Based on: Jacobs R, Meesters JAJ, ter Braak CJF, van de Meent D, van der Voet H. (2016) Combining exposure and effect modelling into an integrated probabilistic environmental risk assessment for nanoparticles. *Environmental Toxicology and Chemistry*. Accepted Author Manuscript. doi: 10.1002/etc.3476.

ABSTRACT

THERE is a growing need for good environmental risk assessment of engineered nanoparticles (ENPs). Environmental risk assessment of ENPs has been hampered by lack of data and knowledge about ENPs, their environmental fate and their toxicity. This leads to uncertainty in the risk assessment. To effectively deal with uncertainty in the risk assessment, probabilistic methods are advantageous. In this chapter, we develop a method to model both the variability and uncertainty in environmental risk assessment of ENPs. This method is based on the concentration ratio (CR), the ratio of the exposure concentration to the critical effect concentration, both considered to be random. In our method, variability and uncertainty are modelled separately, so as to allow the user to see which part of the total variation in the CR is due to uncertainty and which part is due to variability. We illustrate the use of our method using a simplified aquatic risk assessment of nanoTiO₂. Our method allows a more transparent risk assessment and can also direct further environmental and toxicological research to the areas in which it is most needed.

3.1. INTRODUCTION

THERE is a growing need for good environmental risk assessment of ENPs. ENPs may pose a potential risk in the environment because the increased production and use of ENPs and derived products result in relevant release of ENPs into the environment (European Commission, 2013; Royal Commission on Environmental Pollution, 2008; Wiesner et al., 2006).

Environmental risk assessment of ENPs has been hampered by large uncertainty. This uncertainty may be due to lack of data and knowledge about ENPs, their environmental fate, their toxicity (Klaine et al., 2012) and about how to apply standard methods (Grieger et al., 2009). Part of the uncertainty may also be due to artifactual results (Petersen et al., 2014). In traditional risk assessment procedures, the problem of uncertainty is commonly addressed by making use of conservative or worst-case scenarios.

Using conservative scenarios to deal with uncertainty, however, is not desirable for three reasons. First, a conservative scenario is by definition unrealistic in order to be on the safe side. This may result in an over-conservative risk assessment leading to unnecessarily stringent regulation on the use of nanotechnology. Second, the transparency of the risk assessment is compromised in that it is nearly impossible to explicitly quantify how conservative the risk assessment is. Third, in a deterministic conservative risk assessment, it is not possible to differentiate between uncertainty and variability. Uncertainty is the, in principle, reducible variation that exists due to lack of data and knowledge (ECHA, 2012c). On the other hand, variability is the natural inherent variation that is present in all natural processes and living organisms and is, therefore, not reducible (ECHA, 2012c). To improve a risk assessment, the effect of uncertainty on the risk assessment needs to be studied and if necessary, reduced. This is only possible if we separately quantify uncertainty and clearly trace the path to its sources. A deterministic risk assessment does not allow for such a separation in a transparent way. This hampers focused research on areas of high uncertainty because these cannot be identified. Probabilistic methods are a way forward to deal effectively with uncertainty in the risk assessment.

A literature search on the words *probabilistic risk assessment* in Scopus, the world's largest abstract and citation database covering more than 21 000 peer-reviewed journals, produced more than 10 000 results. Figure 3.1 illustrates the massive increase in the number

of publications in the last 15 years. Adding the word *nano** to the search, however, only gives a meagre 60 results (20 February 2016). The difference is evident in Figure 3.1 and underlines the need for more research into probabilistic methods for the risk assessment of ENPs. This call is echoed by Koelmans et al. (2015) who call for probabilistic modelling when dealing with uncertainty.

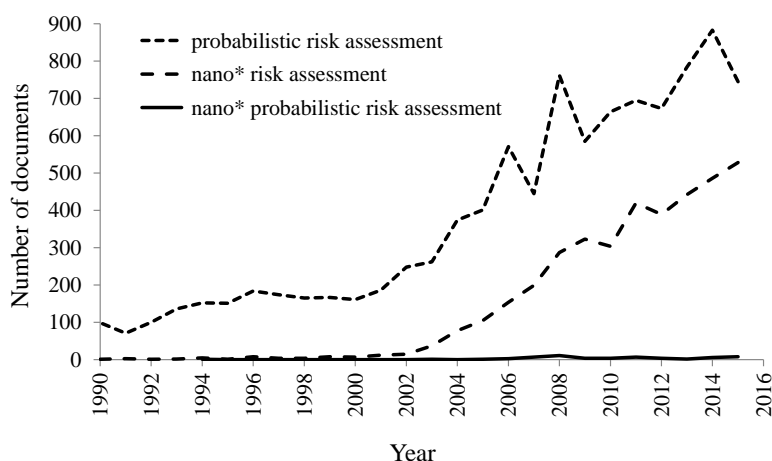


Figure 3.1: A line graph illustrating the number of publications from 1990 to 2015 on probabilistic risk assessment, nano risk assessment and probabilistic nano risk assessment.

Probabilistic methods for the risk assessment of ENPs include Monte Carlo analysis and Bayesian networks (Chio et al., 2012; Coll et al., 2015; Gottschalk et al., 2013a; Money et al., 2012). Although these methods quantify the variation in the various components of the risk assessment, this variation is referred to as uncertainty only in the mentioned publications. Some of this variation, however, is also due to variability.

In this chapter, we use Integrated Probabilistic Risk Assessment (IPRA) to model both the variability and uncertainty in environmental risk assessment of ENPs. The IPRA method was developed for the risk assessment of human health effects due to chemicals (van der Voet and Slob, 2007; van der Voet et al., 2009) and has found many applications (Bokkers et al., 2009; Bos et al., 2009; Bosgra et al., 2009; Müller et al., 2009; Muri et al., 2009; Slob et al., 2014). It has also been applied to nanosilica in food (Jacobs et al., 2015b). To the best of our knowledge, this is the first time that IPRA is used for environmental risk assessment of ENPs. We illustrate the use of the method using a simplified aquatic risk

assessment of nano TiO_2 . The case study is presented as an illustration of our proposed method, and it should be noted that a full risk assessment of nano TiO_2 is outside the scope of this chapter.

Risk assessment consists of exposure assessment, hazard assessment and risk characterization. For environmental exposure assessment, we use the multimedia fate model, SimpleBox4Nano (SB4N) (Meesters et al., 2014), to predict exposure concentrations (ExpCs) of nano TiO_2 in the aquatic compartment. By extending the model with uncertainty and variability distributions, we can quantify the variability of predicted ExpCs in a cumulative distribution function (cdf) with confidence bands that quantify the uncertainty.

For environmental hazard assessment, we start from the probabilistic species sensitivity distribution (pSSD) model of Gottschalk and Nowack (2012) and Coll et al. (2015) and adjust it to separately quantify variability and uncertainty. Similarly to the exposure assessment, our method allows the variability in critical effect concentrations (CECs) to be quantified in a cdf with confidence bands that quantify the uncertainty.

Finally, the exposure and hazard assessment are combined into the CR. Besides being designed for the separate quantification of variability and uncertainty, our method also allows us to study the contribution of the different uncertainty sources to the total uncertainty in the CR.

In Section 3.2, we provide the background of the SB4N and pSSD models and describe the IPRA method. In Section 3.3, we provide the results of applying our method to nano TiO_2 in water. In Section 3.4 we discuss the results and discuss our method and its limitations.

3.2. METHOD

IN this section, we provide the background of the SB4N and pSSD models. We then describe the IPRA method and its implementation.

3.2.1. BACKGROUND

EXPOSURE ASSESSMENT

SB4N is a multimedia fate model that simulates the environmental fate of ENPs (Meesters et al., 2014). It is a modification of the original SimpleBox model (Brandes et al., 1996; den Hollander et al., 2004; van de Meent, 1993) used for chemical exposure assessment in Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). SB4N models the fate of ENPs in four compartments, namely, atmosphere (including rain), surface water, sediment and soil (including soil pore water). Within each compartment, ENPs can occur in different physical-chemical forms: (i) freely dispersed (free), (ii) hetero-aggregated with natural colloidal particles, or (iii) attached to larger natural coarse particles that are prone to gravitational forces in aqueous media (Meesters et al., 2014). Using a mass balance modelling system (Mackay, 2003; van de Meent et al., 2011), SB4N obtains the masses (in kg) of ENPs in each of the four compartments and for each of the three forms. These can be converted to concentrations by dividing the mass by the total volume of the corresponding compartment.

SB4N performs a deterministic exposure risk assessment. The model takes single-value inputs and returns single-value masses. In Section 3.2.2, we place SB4N in a two-dimensional Monte Carlo (2DMC) structure in order to feed the model with variability and uncertainty data and obtain the variability and uncertainty distributions of the ExpCs of ENPs. For easier implementation in IPRA, we coded SB4N (which is an Excel model) in R software (R Core Team, 2015) in an object-oriented way.

HAZARD ASSESSMENT

Different species have different sensitivities. Sensitivities are quantified in the form of, what we call, limit concentrations, such as no observed effect concentration (NOEC), 10% lethal concentration (LC10), 50% lethal concentration (LC50), 10% effect concentration (EC10), 50% effect concentration (EC50). A statistical distribution describing the differing sensitivities among a group of species is called a species sensitivity distribution (SSD). Gottschalk and Nowack (2012) developed the pSSD method which, in addition to quantifying the variability in species' sensitivity also includes the variation within a species due to different experimental conditions. This method was further extended to

include further uncertainty about the data points and the assessment factors used (Coll et al., 2015; Gottschalk et al., 2013a).

The pSSD method was developed on data from literature. The data was collected (Coll et al., 2015) according to selection criteria in accordance with REACH guidance (ECHA, 2008b). First, only effects on survival, growth, reproduction and changes in significant metabolic processes (e.g. photosynthesis (Coll et al., 2015)) were included. Second, only toxicity studies on living organisms were included (i.e. no tissue or in vitro experiments). Third, if chronic and acute limit concentrations were available, the chronic one was chosen. Fourth, only one limit concentration per study was used. Finally, all different limit concentrations from tests which used different particle type, particle sizes or media were included. For the specific case of nanoTiO₂ in the aquatic compartment, there were 73 limit concentrations for 31 species from 5 taxonomic groups (Table S5 in Supporting Information in Jacobs et al. (2016)) (Coll et al., 2015).

In order to incorporate all the different limit concentrations into one SSD, the limit concentrations are transformed to species sensitivity values by making use of two assessment factors (AFs) (Coll et al., 2015; Gottschalk et al., 2013a). In this chapter, we refer to these species sensitivity values as chronic critical effect concentrations (CECs). The first AF transforms the limit concentration to a critical effect concentration. An AF of 1 was used for NOEC and highest observed no effect concentration (HONEC), AF of 2 for LC10, 20% lethal concentration (LC20), EC10, 20% effect concentration (EC20), lowest observed effect concentration (LOEC), and lowest effective dose (LED) and AF of 10 for 25% lethal concentration (LC25), LC50, 25% effect concentration (EC25) and EC50 values. The second AF transforms from short-term to long-term effects. An AF of 1 was used for long-term experiments and an AF of 10 was used for short-term experiments. The exposure time needed to classify an experiment as long- or short-term varies according to the taxonomic group (Coll et al., 2015).

After the data transformation step, the SSD is constructed in two steps. In the first step, a single empirical SSD for each species is constructed using a Monte Carlo routine (Gottschalk et al., 2013a). In the second step, all the single SSDs are combined into one empirical SSD.

Risk assessors are often interested in a predicted no-effect concentration (PNEC) which

generally is the 5th percentile of the SSD, also referred to as the 5% hazard concentration (HC5). Coll et al. (2015) extended the pSSD method to include uncertainty on the assessment factors and extra uncertainty on the endpoints, which, in a Monte Carlo simulation, provides an uncertainty distribution for the PNEC.

The pSSD method quantifies uncertainty and variability. The constructed SSD, however, contains both the variability of species sensitivity as well as the uncertainty due to experimental differences within a single species. In addition, the uncertainty distribution of the PNEC contains the uncertainty of the assessment factors and only partially the uncertainty of the limit concentrations. This is due to the fact that the experimental uncertainty was modelled together with variability in the constructed SSD, thereby, combining variability and uncertainty in one SSD. It is, therefore, neither possible to study the effect of uncertainty on the effect assessment, nor to study the contribution of the different sources of uncertainty to the hazard assessment and ultimately the CR. In Section 3.2.2, we adjust the pSSD method to allow for the separate quantification of variability and uncertainty.

3.2.2. INTEGRATED PROBABILISTIC RISK ASSESSMENT

IPRA uses a 2DMC scheme to quantify uncertainty and variability distributions separately in the risk assessment as illustrated in Figure 3.2. Details on the exposure, hazard and risk aspects of the model are discussed in Sections 3.2.2, 3.2.2 and 3.2.2, respectively. IPRA is available in the Monte Carlo Risk Assessment (MCRA) software (van der Voet et al., 2015) in the context of human health, but for our environmental risk assessment application, it was coded in R Software (R Core Team, 2015).

QUANTIFYING VARIABILITY AND UNCERTAINTY IN EXPOSURE

In order to allow defining variability and uncertainty in exposure concentrations, it is important to define the unit at risk of the risk assessment. SB4N is designed to predict exposure concentrations on the regional scale, where regions are defined as spatial units of 200km × 200km. Variability in exposure is, therefore, defined as the natural occurring variation in ExpCs between regions.

SB4N has many input variables, which may be variable between regions, uncertain, or

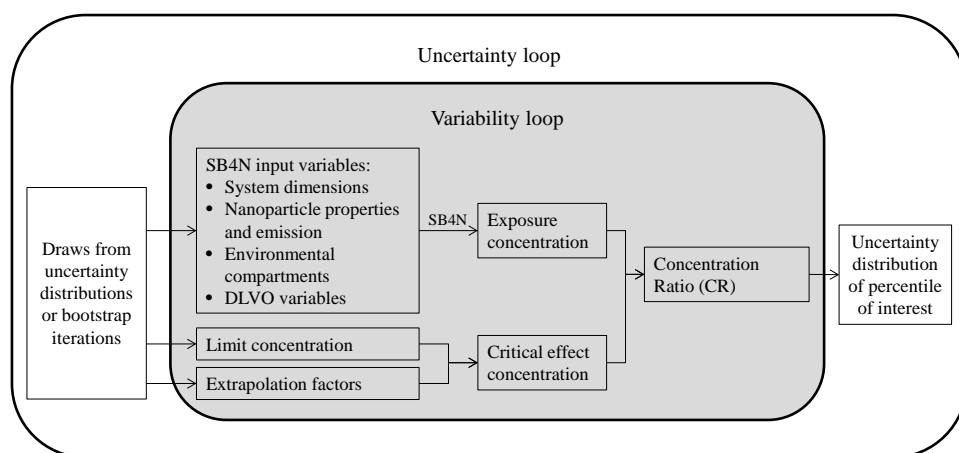


Figure 3.2: A schematic diagram of uncertainty and variability loops in the two-dimensional Monte Carlo scheme used in IPRA. SB4N (SimpleBox4Nano) is a multimedia fate model that simulates the environmental fate of ENPs producing exposure concentrations from input variables. DLVO: Derjaguin Landau Verwey Overbeek theory (Derjaguin and Landau, 1941; Verwey and Overbeek, 1948) to calculate attachment efficiencies.

both. In order to keep the number of variables manageable, we made a selection of the most important variables on which to apply the 2DMC algorithm. (Meesters et al., 2016) conducted a sensitivity analysis to determine which variables play a large role in determining the nanoparticle masses in the various compartments. In this chapter, we only considered aquatic risk assessment and are, therefore, only interested in the nanoparticle masses in the aquatic compartment. From the sensitivity analysis (Meesters et al., 2016), we selected those variables that had a large influence on the nanoparticle masses in the aquatic compartment.

For each of the selected variables, we obtained ranges of possible values from literature (see Tables S1-S4, column 7 in Supporting Information in Jacobs et al. (2016)). We assumed that the main source of variation for each variable was due to either variability or uncertainty. Moreover, the available information in the literature was not sufficient to determine which part of the range of values was due to variability and which part was due to uncertainty. We, therefore, assumed that the reported variation was due completely to either uncertainty or variability. The selected variables are indicated by a “V” (for a variable input) or “U” (for an uncertain input) in Tables S1-S4, column 4. The

remaining variables (indicated by a “C” in Tables S1-S4, column 4 in Supporting Information in Jacobs et al. (2016)) were given the default value.

First, we will look in more detail at variability. In terms of the final output of the exposure assessment, namely predicted environmental concentration, variability is the naturally occurring variation in environmental concentration between regions. The distributions should, therefore, quantify the natural variation of that variable between regions.

3

As an example, consider the variable `waterdepth`, which is one of the system dimension variables of SB4N. It is obvious that water depth is variable when looking at a collection of water bodies. Considering the regional scale of SB4N, however, the water depth variable does not represent the depth of an individual water body, but rather the average water depth of all water bodies in a $200\text{km} \times 200\text{km}$ region. The variability distribution for `waterdepth` should, therefore, quantify the variability in average water depth between regions. More concrete, this means capturing the variability in average water depth between $200\text{km} \times 200\text{km}$ regions in the Netherlands, Italy and Norway for example. This variability is quantified by providing `waterdepth` with a variability distribution, namely a log-normal distribution with a mean of 3 and a standard deviation of 0.237 (Table S1 in Supporting Information in Jacobs et al. (2016))

Similarly, variability distributions were applied to all the variables from the pre-selection considered to vary due to variability. These variables are indicated by a “V” in Tables S1-S4 (in Supporting Information in Jacobs et al. (2016)). The choice of distributions were based on experimental or expert knowledge from literature (see references in Tables S1-S4 in Supporting Information in Jacobs et al. (2016)). To keep the method simple, we assumed that the variability distributions were fully known, i.e. the distribution parameters are assumed to be known and not subject to uncertainty.

For one variable, namely `invkdebye`, we obtained an empirical variability distribution. `Invkdebye` is the debye length used to calculate attachment efficiencies between engineered and natural (nano)particles with the Derjaguin Landau Verwey Overbeek (DLVO) theory (Derjaguin and Landau, 1941; Verwey and Overbeek, 1948). Experimental debye length values were obtained by Hammes et al. (2013). Because of the large number of values (808), it was possible to quantify the variability via an empirical distribution.

Next we consider uncertainty. A variable that is subject to uncertainty only, is in theory considered to be a constant.

As an example, consider the variable `diameterenp`, which represents the mean nanoTiO₂ particle size. We assume a similar nanoTiO₂ production between regions and, therefore, similar particle size distributions between regions. In a perfect world where everything is known, the mean particle size would be a known constant. In practice, however, we are uncertain about what this average particle size actually is. This uncertainty is quantified by providing `diameterenp` with an uncertainty distribution, namely a log-normal distribution with 2.5th percentile equal to 1 and 97.5th percentile equal to 100 (Table S2 in Supporting Information in Jacobs et al. (2016))

All variables of the pre-selection considered to be uncertain are given an uncertainty distribution. These distributions are based on experimental or expert knowledge from literature (see references in Tables S1-S4 in Supporting Information (Jacobs et al., 2016)). As was the case for variability, there is one variable, namely `prodvol`, for which we obtained an empirical uncertainty distribution. `Prodvol` is the production volume of nanoTiO₂. We obtained production volumes from a Monte Carlo simulation study (Sun et al., 2014) from which we created the empirical distribution.

The variability and uncertainty distributions applied to the variables are given in Tables S1-S4 (in Supporting Information in Jacobs et al. (2016)). These distributions are the inputs for the 2DMC algorithm (Frey, 1992) in which we generated 200 draws from the joint distribution of the uncertain variables and, given these draws, thus for each row, 1000 draws from the joint distribution of the variables that cause variability in the exposure (see Supporting Information in Jacobs et al. (2016) for detailed algorithm). The values obtained for each combination of uncertainty and variability draws are used as input to SB4N, resulting in an exposure concentration which thus represents a draw from the exposure distribution for a particular draw of the uncertain variables.

The output of the algorithm is thus a 200×1000 matrix with exposure concentrations, where each row represents the variability distribution of the exposure given a particular draw from the joint uncertainty distribution. If each row of the 200×1000 exposure concentration matrix is sorted from small to large, then the value in the 10k-th column of each row is an estimate of the k-th percentile of the exposure distribution for that par-

ticular row. Consequently each column then represents the uncertainty distribution of 1000 equally spaced percentiles of the exposure distribution.

QUANTIFYING VARIABILITY AND UNCERTAINTY IN HAZARD

Building on the pSSD method (Coll et al., 2015; Gottschalk et al., 2013a; Gottschalk and Nowack, 2012), we develop a 2DMC method to separately quantify uncertainty and variability. Again, we need to define the unit at risk, which for effect distributions is commonly taken to be the species.

For the SSD, we need chronic critical effect concentrations (CECs). These, however, are often not directly available for the species we want to include in the SSD. In this case they are calculated as follows

$$CEC_{chronic} = \frac{CONC}{AF_{time} \cdot AF_{no-effect}} \quad (3.1)$$

with CONC referring to the limit concentration (e.g. LC10 or EC20) obtained from a toxicological study, AF_{time} the assessment factor to extrapolate from acute to chronic studies and $AF_{no-effect}$ the assessment factor to extrapolate from the limit concentration to the CEC.

Variability in the $CEC_{chronic}$ refers to the natural variation in critical effect concentrations between species. This variability is quantified by defining a distribution over $CEC_{chronic}$ for different species. In practice, such a distribution is often taken as the log-normal distribution (Aldenberg et al., 2002; Jacobs et al., 2015a; Verdonck et al., 2003; Wagner and Løkke, 1991) assuming the species critical effect concentrations are normally distributed on the log-scale. Also in our method we use log-normal distributions.

The three variables used to calculate $CEC_{chronic}$ (the limit concentration and two assessment factors in Equation 3.1) can all be subject to uncertainty.

Uncertainty in the limit concentration is due to differences between toxicity studies within a species. To quantify this uncertainty, we divide the limit concentrations into groups per species. The uncertainty distribution for each species is taken to be a log-normal distribution with the 2.5th (97.5th) percentile equal to the minimum (maximum) concentration in that group divided (multiplied) by uncertainty factor 2. This uncertainty factor

is based on the assumption that, due to uncertainty, the limit concentration can be a factor 2 lower or higher than the measured experimental limit concentration value(s). This factor of 2 is similar to the 50% coefficient of variation used in the pSSD method (Gottschalk et al., 2013a).

Table 3.1: NanoTiO₂ effect data for *Danio rerio* (Coll et al., 2015)

Limit concentration type	Limit Concentration ($\mu\text{g/l}$)	Exposure time (h)	AF_{time}	$AF_{no-effect}$
HONEC	500	4320	1	1
LC50	124500	96	10	10
LC50	156000	24	10	10
LC50	300000	24	10	10
HONEC	500000	96	10	1

Within one species, however, one can have different experimental duration types (two types: short or long) and limit concentration types (three types: (i) NOEC, HONEC, (ii) LC10, LC20, EC10, EC20, (iii) LC25, LC50, EC25, EC50). An example of such a species is *Danio rerio* as shown in Table 3.1. We cannot combine these into one uncertainty distribution, because each group needs to have different assessment factors applied to it. For these species, we sample, in each uncertainty run, one group with probability equal to the number of concentration values divided by the total number of concentration values for that species. For the example in Table 3.1, we would sample one of the groups with probabilities 0.2, 0.6 and 0.2 for the three groups. The log-normal species uncertainty distribution is then assumed for that group as explained above.

The uncertainty distribution for the assessment factors are centred around the nominal values as explained in Section 3.2.1 and given in Table S5 columns 6 and 7 (in Supporting Information in Jacobs et al. (2016)). As in the case of the limit concentrations, we again use an uncertainty factor of 2 below and above each assessment factor value. This is similar to the 50% deviation used in the pSSD method (Coll et al., 2015). The obtained lower and upper bounds are again equated to the 2.5th and 97.5th percentiles of a log-normal distribution.

In each uncertainty run, one limit concentration is drawn from each of the 31 species uncertainty distributions. Each of these limit concentrations is then divided by a value

drawn from the corresponding uncertainty distribution of each assessment factor. The resulting 31 $CEC_{chronic}$ values are used to estimate the mean and standard deviation for the log-normal distribution of the variability, i.e. the SSD. The interested reader is referred to the Supporting Information in Jacobs et al. (2016) for a detailed algorithm.

Similarly to the exposure assessment, the output of the algorithm is a 200×1000 matrix with critical effect concentrations, where each row represents the variability distribution of the critical effect, i.e. the SSD, for a particular draw from the joint uncertainty distribution. If each row of the 200×1000 critical effect concentration matrix is sorted from small to large, then the value in the 10k-th column of each row approximately is an estimate of the k-th percentile of the critical effect distribution for that particular row. Consequently each column then represents the uncertainty distribution of 1000 equally spaced percentiles of the critical effect distribution or SSD.

INTEGRATED PROBABILISTIC RISK ASSESSMENT

In this section, we discuss the integration of the exposure and hazard assessments into the risk characterization. For this, we use the concentration ratio (CR), given by

$$CR = \frac{ExpC}{CEC_{chronic}}$$

A CR less than 1 indicates that the exposure concentration is lower than the chronic critical effect concentration of the species and, therefore, indicates a safe situation. A CR greater than 1, however, indicates a possibly unsafe situation.

Combining the units of the exposure and the effect models we obtain the unit at risk as a species in a $200km \times 200km$ region. The variability distribution, therefore, describes variation between random species in random regions.

The matrix of CR values is obtained by dividing the (unsorted) exposure matrix by the (unsorted) critical effect matrix element-wise. Each row represents the variability distribution of CR given a particular draw from the joint uncertainty distribution. If each row of the 200×1000 CR matrix is sorted from small to large, then the value in the 10k-th column of each row is an estimate of the k-th percentile of the CR distribution for that particular row. Consequently each column then represents the uncertainty distribution of 1000 equally spaced percentiles of the CR distribution.

A simple graphical representation of both variability and uncertainty of the CR can be given in the form of a so-called CR bar graph (similar to the IPRA bar graphs in Jacobs et al. (2015b) and van der Voet et al. (2009)). In a CR bar graph, a box represents the variability distribution of the CR between specified percentiles. These can be particular percentiles (denoted by p_x for the x th percentile), e.g. $p_{0.1}$ and $p_{99.9}$, p_1 and p_{99} , or p_5 and p_{95} , depending on the level of protection required. Whiskers are used to represent the 5% lower and 95% upper uncertainty limits of these percentiles. A dot on the bar indicates the median of the variability distribution.

We also calculate the risk, $R = P(CR > 1)$, together with its uncertainty bounds.

To study the extent to which sources of uncertainty contribute to the total uncertainty present in a certain percentile of interest, we implement a probabilistic uncertainty analysis (van der Voet and Slob, 2007). We grouped all sources of uncertainty into just two groups, namely exposure related and effect related uncertainties. This results in a 2^2 factorial design where sampling from the uncertainty distributions for each group is turned on and off. For a given percentile, $2^2 = 4$ values are obtained in each uncertainty run, resulting in 4 distributions which are summarised by their variance. An additive model is then fitted to the 4 variances. When this model explains most of the variance, which is usually the case, the coefficients of the main effects can indicate the contribution to the total variation (Jacobs et al., 2015b). The intercept term represents the additional uncertainty due to Monte Carlo sampling when the two input group uncertainty sources are turned off. Without any uncertainty in the inputs, there is still variation in output due to the random Monte Carlo sampling of variability. Results are illustrated by means of a bar graph.

3.3. RESULTS

IN this section, we describe the results obtained from an application of the method that we propose in this chapter, to an aquatic risk assessment of nanoTiO₂. Hereby, we illustrate what kind of information can be obtained from our method and how our method can be used to gain insight into the role that variability and uncertainty play in nanoparticle risk assessment.

The variability and uncertainty distributions applied to variables of the SB4N model are provided in Tables S1-S4 (in Supporting Information in Jacobs et al. (2016)).

Figure 3.3 illustrates the total exposure and the critical effect distributions, with uncertainty bands. The exposure distribution is plotted as an exceedance (1-cdf) curve indicating the percentage of regions that exceed the concentration on the x-axis. The amount of overlap of the curves is an indication of the amount of risk and is related to the expected risk concept (van Straalen, 2002) and the area under the joint probability curve (Solomon et al., 2000; van Straalen, 2002).

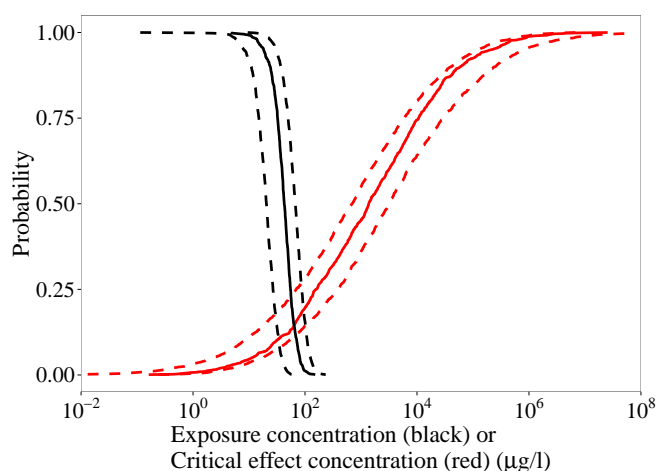


Figure 3.3: The exposure exceedance distribution (black solid line) and critical effect cumulative distribution (red solid line). The dotted lines indicate the 95% uncertainty band.

Figure 3.4 shows the CR bar graph plotted for various forms of nanoparticle exposure. For each bar, a different exposure concentration (as indicated by the labels) was used to calculate the CR. The five bars represent the variability distribution of the CR between the 1st (p1) and the 99th (p99) percentiles. For each bar, the whiskers represent the 5% lower and 95% upper uncertainty limits of these percentiles. The dot on each bar indicates the median of the variability distribution. Except for the free nanoparticle exposure concentrations, all the exposure concentrations caused to some extent a CR greater than 1. The implications of this is further discussed in Section 3.4.

Figure 3.5 illustrates the uncertainty distribution of the risk, $R = P(CR > 1)$, using total

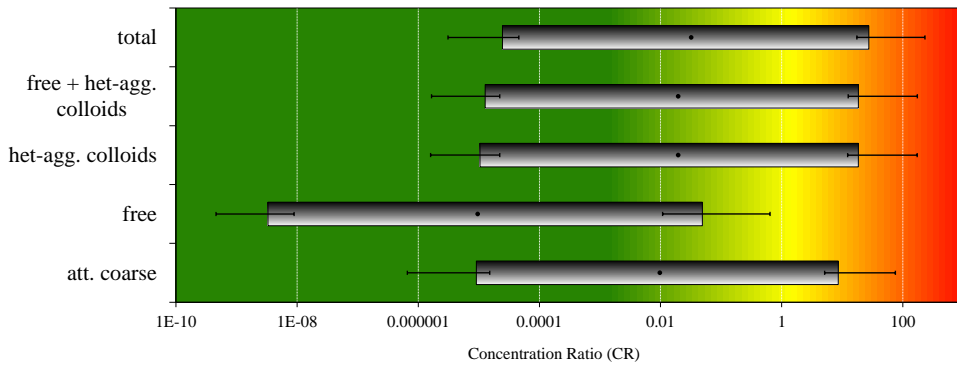


Figure 3.4: Concentration ratio (CR) bars illustrating the variability and uncertainty distributions of the CR. The five boxes represent CR distributions between p1 - p99 for ENPs that are attached to coarse particles, free, hetero-aggregated with colloids, free + hetero-aggregated with colloids, and total exposure. In each box, the whiskers represent the lower 5% uncertainty bound and the upper 95% uncertainty bound of the lower and upper percentiles. Background colouring visualises the transition from low (green) to high risk (red).

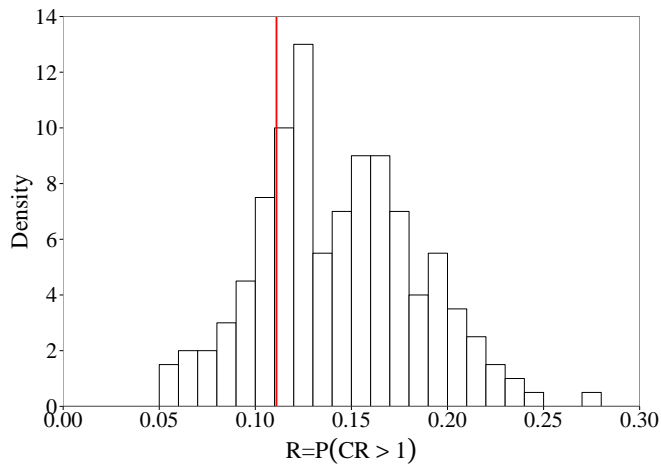


Figure 3.5: Histogram of the uncertainty distribution of the risk, $R = P(CR > 1)$. The red vertical indicates the nominal risk.

exposure. The vertical line indicates the nominal risk value (0.111) which is $R = P(CR > 1)$ calculated using only the variability distribution of CR without any uncertainty. The risk distribution specifies variation between species and between regions. Note that this

can correspond with many different situations, with as extremes: (i) 11.1% of species are at risk in all regions or (ii) all species are at risk in 11.1% of the regions. A discussion on this double interpretation and its drawbacks can be found in Verdonck et al. (2003).

It is important to determine how the uncertainty in the percentiles of the CR is affected by the different uncertainty sources. Figure 3.6 indicates the relative contribution of each source of uncertainty to the total uncertainty in four upper percentiles of the CR distribution, namely p90, p95, p97.5 and p99. We note that the contribution of Monte Carlo uncertainty is negligible for all the percentiles, therefore our choice for 1000 Monte Carlo iterations to describe the variability seems sufficiently high. The uncertainty in the critical effect concentration is the main contributor to the total uncertainty for all percentiles, increasing for the more extreme percentiles. In order to further study the exact source of this uncertainty, one could perform a similar uncertainty analysis on the individual uncertainty sources that contribute to the uncertainty in the exposure and critical effect concentrations.

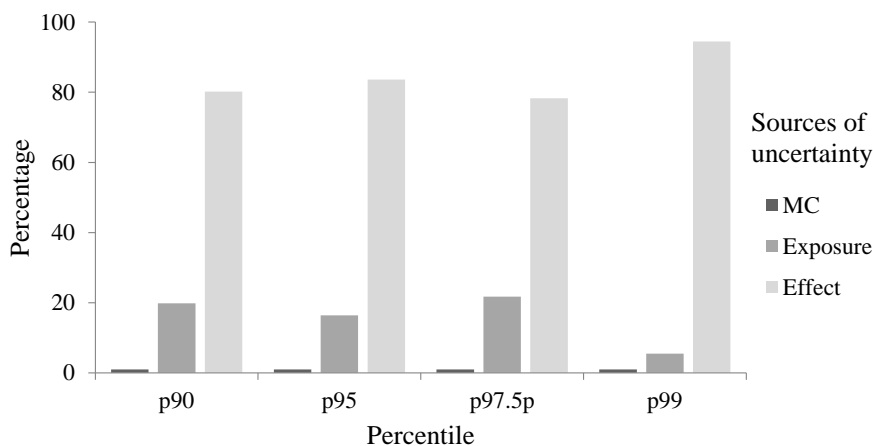


Figure 3.6: A bar graph illustrating the relative contribution of each source of uncertainty to the uncertainty of the four upper percentiles, p90, p95, p97.5 and p99 of the CR distribution. The variance explained by the additive model for the four percentiles are 98.47%, 98.76%, 94.76% and 99.76%, respectively. MC: Monte Carlo, Exposure: exposure concentration, Effect: critical effect concentration

3.4. DISCUSSION

IN this section, we discuss the results, the use of free nanoparticles, ENP hetero-aggregates with natural colloid particles and ENPs attached to natural coarse particles in risk assessment and model uncertainty of the exposure and effect models.

Although Figure 3.4 may lead to believe that nanoTiO₂ poses some risk to the aquatic environment, care should be taken in its interpretation. The CR is calculated using the $CEC_{chronic}$ which is assumed to be a no-effect concentration. This concentration is extrapolated from some limit concentration by an assessment factor, $AF_{no-effect}$. The true no-effect concentration is not known. Keeping this in mind, an CR greater than 1 does not indicate a negative effect with certainty, but rather a potentially unsafe situation. We can no longer exclude a possible risk.

Another point to consider is the quantification of variability in the exposure assessment. All of the distributions used (as given in Tables S1-S4 in Supporting Information in Jacobs et al. (2016)) are motivated by literature as the possible range the variable can take. In the case of variability, however, these distributions do not always necessarily reflect the realistic variability. This is due to the fact that in literature one usually finds the possible range of individual values that a certain variable can take. The variability distribution, however, should quantify the variability of the mean value of a variable in a region ($200km \times 200km$). From the central limit theorem, we know that the standard deviation of the mean is σ/n where σ is the standard deviation of the individual values and n the sample size (Bain and Engelhardt, 1992, pg 238). We would, therefore, expect the true variability distributions to be narrower than the ones we used. In the case of the exposure, this would result in a narrower exposure concentration distribution. This will work through into the CR distribution in Figure 3.4 and may cause the grey bars to be less wide, resulting in a less extreme upper percentile.

Variability in exposure, even when possibly over-estimated, does not, however, seem to be the major source of variability in the CRs. The large contribution of effect variability is clearly illustrated in Figure 3.7. Figure 3.7 illustrates the variability distributions (p1-p99) of the various exposure concentrations and the critical effect concentrations. The distribution of the critical effect is much wider than that of the exposure. The large vari-

ability in the effect is due to the large variation in critical effect concentrations among species. Some species are much more sensitive to nano TiO_2 than others.

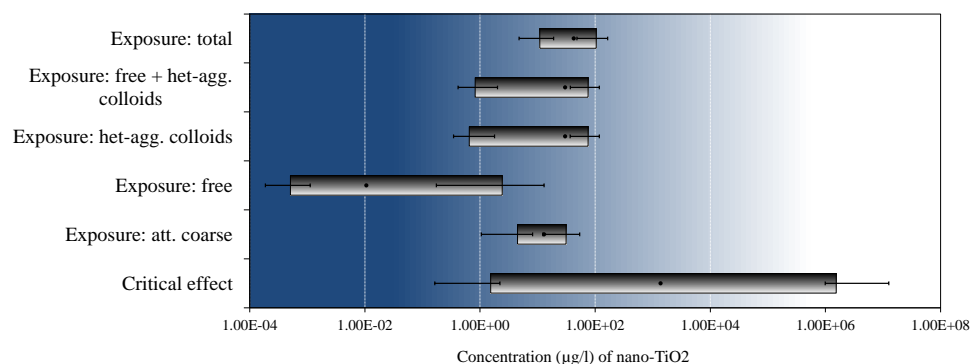


Figure 3.7: Exposure and critical effect concentration bars illustrating the variability and uncertainty distributions of the exposure and critical effect. The first five boxes represent the exposure distributions between p1 - p99 for ENPs that are attached to coarse particles, free, hetero-aggregated with colloids, free + hetero-aggregated with colloids, and total exposure. The last box represents the critical effect distribution between p1 - p99. In each box, the whiskers represents the lower 5% uncertainty bound of the lower percentile and the upper 95% uncertainty bound of the upper percentiles. Background colouring visualises the transition from low (blue) to high nano TiO_2 concentration (white).

Although we might be able to reduce the total variability of the CR distribution by more accurate specification of the variability distributions in the exposure assessment, the large amount of variability in the effect concentrations will prevent any significant reduction.

In this chapter, we applied the IPRA method to the aquatic risk assessment of nano TiO_2 . This method, however, is not limited to the aquatic compartment. SB4N is a generic model, modelling the fate of nanoparticles for the environmental compartments of air, water, soil and sediment (Meesters et al., 2014). Our method can, therefore, be applied to any of these four compartments, provided there is sufficient critical effect data available for that compartment.

3.4.1. EXPOSURE OF FREE, HETERO-AGGREGATED AND ATTACHED ENPs

The ratios between exposure and critical effect concentrations only suggest safe concentrations for the free forms of nano TiO_2 , but not for exposure to hetero-aggregates with natural colloid particles, ENPs attached to natural coarse particles or the sum of all ENP forms (Figure 3.4). That does not directly indicate that aquatic organisms are at risk. This is complex to assess as it is not yet known to what extent the relevant exposure concentration should include ENPs that are attached to natural particles (Koelmans et al., 2015). There are no approaches designed to quantify predicted ExpCs into bioavailable exposure estimates (Gilbertson et al., 2015), because fate and exposure of ENPs are not incorporated in aquatic toxicity tests (Petersen et al., 2015). The current risk assessment frameworks, such as REACH, do not consider the fraction of chemicals or metals that is associated with suspended particles to contribute to environmental exposure, because free metal forms are “far more bioavailable than most complexed metal” forms (ECHA, 2008a). Under REACH, the free (dissolved) concentration of a metal (oxide) is defined as “the fraction of a metal that passes through a filter of 450 nm” (ECHA, 2008a). Following this definition for ENPs would mean that the sum of dissolved/ionic, free pristine nanoparticulate forms and hetero-aggregates (< 450 nm) is considered to be the bioavailable fraction. Moreover, aggregation might even increase the uptake and bioavailability of ENPs. Depending on the feeding mechanism of the organism at stake, (hetero)-aggregated ENPs may have grown to a size so that they no longer pass the filtering apparatus of filter feeders (Handy et al., 2012; Koelmans et al., 2015). On the other hand, an aggregated state probably yields higher effect thresholds, because particle toxicity would be lowered by aggregation or encapsulation of the nanoscale particles (Handy et al., 2012; Koelmans et al., 2015). The CECs that are applied in the CR are based on toxicity testing of free and homo-aggregated ENPs (Coll et al., 2015) and do not account for such possible reduction of the toxicity of the hetero-aggregated form. Hence, the CRs for nano TiO_2 that do not ensure safe concentrations (Figure 3.4) are still conservative estimates and should thus be interpreted with caution. Nonetheless, the results emphasize the relevance of the debate whether aquatic toxicity of ENPs should be tested in their freely dispersed and unaggregated state or in a more environmentally realistic state that include ENPs present as aggregates (Koelmans et al., 2015). The CRs prove that only in-

cluding free ENPs and excluding the hetero-aggregated ENPs may lead to supposed safe but in reality unsafe concentrations.

From the above discussion, we see that the nanoparticle form in current toxicity tests is not compatible with the nanoparticle form to which species are exposed to in the environment. Toxicity testing is performed on a substance (free and homoaggregated ENPs) to which species are hardly exposed (see Figure 3.7). At the same time, we do not know the toxicity of the substance (heteroaggregated ENPs) that species are exposed to in reality. This incompatibility between toxicity and exposure data constitutes extra uncertainty, which, if not resolved, could possibly be modelled.

Moreover, possible cumulative exposure to natural and engineered nanoTiO₂ is not considered in our case study serving as a proof of concept for the IPRA approach. Such natural background concentrations are derived as elemental Ti concentrations in field samples filtered for submicron particles for < 450 nm, which are found to typically range between 0.02 and 2.3 µg/L in rivers (De Klein et al., 2016; Gaillardet et al., 2003). Hence, these measured concentrations are actually the sum of the elemental mass of Ti in dissolved, free and aggregates of nanoparticles able to pass through a filter of < 450 nm (De Klein et al., 2016; Gaillardet et al., 2003). The current measurement techniques are not able to quantify the different forms of Ti in these field samples (Nowack et al., 2015), so that considering the cumulative exposure of natural and engineered particles is only possible for the predicted concentrations that reflect the sum of free and hetero-aggregated nanoTiO₂ (Figure 3.7). Indeed, there is some overlap between the range of typical natural concentrations and of the concentrations calculated for engineered nanoTiO₂, but only at the lower end of the exposure distribution. Moreover, the natural Ti concentrations hardly surpass the critical effect concentrations, so that cumulative exposure of natural and engineered nanoparticles would only contribute to a minor extent in the environmental risk assessment of nanoTiO₂. Nonetheless, the possibility of such cumulative exposure again emphasises the need for consensus on what forms of ENPs should be interpreted as the relevant exposure concentration and their compatibility with effect concentrations determined in the current toxicity testing protocols (Koelmans et al., 2015).

3.4.2. EXPOSURE MODEL UNCERTAINTY

The simulation of environmental fate and concentrations of nano TiO_2 are performed within the context of the chemical safety assessment guidelines of REACH (ECHA, 2009). Within this context, environmental exposure models are considered a means to determine whether manufacture, import or uses of a substance does not lead to concentrations that are not safe for the environment (ECHA, 2009). In a first-tier approach, non-spatial multimedia fate models such as SimpleBox suffice, but further iteration is required if the conservative estimates for exposure levels are not below predicted no-effect concentrations (ECHA, 2009).

SB4N is a screening level model that is designed for exposure assessment of background concentrations on a regional or continental scale (Koelmans et al., 2015; Liu et al., 2014). Simplifications in environmental exposure modelling are inevitable but acceptable if they can be justified scientifically (Praetorius et al., 2013). As such, SB4N is a generic model that is not temporal or spatially explicit, whereas complex chemical reactions between ENPs and environmental matrices are only implicitly included in the calculations of speciation (Meesters et al., 2014). Multimedia fate models that are spatially explicit, however, only yield better estimates if data on spatial variability in emission intensities are available (Hollander et al., 2012). To our knowledge, such data is only available for nanomaterials by extrapolation of the global and U.S. production volume data in proportion to the Gross Domestic Product of the region (Gottschalk et al., 2013b; Mueller and Nowack, 2008; Sun et al., 2014). Including temporal explicitness in exposure estimation also does not yield better estimates. The exposure concentrations are calculated for a steady state, but recalculation of the progress over time in reaching the simulated steady state predicted ExpCs in surface water (see Supporting Information in Jacobs et al. (2016)) shows this only leads to an overestimation of insoluble ENPs that are attached to natural coarse particles (see Supporting Information in Jacobs et al. (2016)).

Furthermore, chemical transformation processes such as functionalization, oxidation, sulfidation, phosphorization and adsorption of natural organic carbon are considered to be too complex to explicitly include in a screening level exposure model such as SB4N (Koelmans et al., 2015; Meesters et al., 2014; Sani-Kast et al., 2015). These complex transformations of ENPs in the environment, however, are not disregarded in SB4N. Rather,

they are indirectly included through their contribution in ENP dissolution rates and the interaction between ENPs with natural particles (see Supporting Information in Jacobs et al. (2016)). Hence, the simplification in chemical speciation modelling and the lack of spatial and temporal detail do not hamper the evaluation of the exposure of aquatic species to nano TiO_2 as long as the SB4N model outcomes are interpreted on a generic screening level, i.e. conservative and first-tier (Koelmans et al., 2015).

3

3.4.3. EFFECT MODEL UNCERTAINTY

One of the problems in hazard assessment is how to deal with more than one critical effect concentration per species. REACH suggests using the geometric mean for each species with equivalent data on the same toxic endpoint (ECHA, 2008b). The geometric mean, however, favours small values, because it shifts the SSD to the left (Gottschalk and Nowack, 2012). This may lead to lower critical effect concentrations and a possibly over conservative risk assessment (Gottschalk and Nowack, 2012). The pSSD method solves this problem by using all the available data to first construct single species SSDs which are then combined into one SSD. The pSSD method, however, does not differentiate whether these single species SSDs quantify uncertainty or variability. In our method, we assume that under identical experimental test conditions and using identical test species, repeated experiments would, in theory, result in the same limit concentration, i.e. no variability. From this assumption, we deduce that the observed differences in limit concentrations for the same species should be attributed to uncertainty. The next question is then, how this uncertainty should be quantified. In our method, we assumed a log-normal distribution.

Another source of model uncertainty is usage of the assessment factors. Gottschalk et al. (2013a) applied the 2 assessment factors (AF_{time} and $AF_{no-effect}$) according to REACH guidelines as explained in Section 3.2.1. To apply AF_{time} , we need to know which studies are acute studies and which ones are chronic. Gottschalk et al. (2013a) provide a detailed description of the choice of AF_{time} for different taxonomic groups. For example, for algae, limit concentrations were considered chronic from an experiment duration of 72 h and more, while for vertebrates, an experiment duration of 21 days was considered chronic. Such choices, although based on recommendations from literature, are ulti-

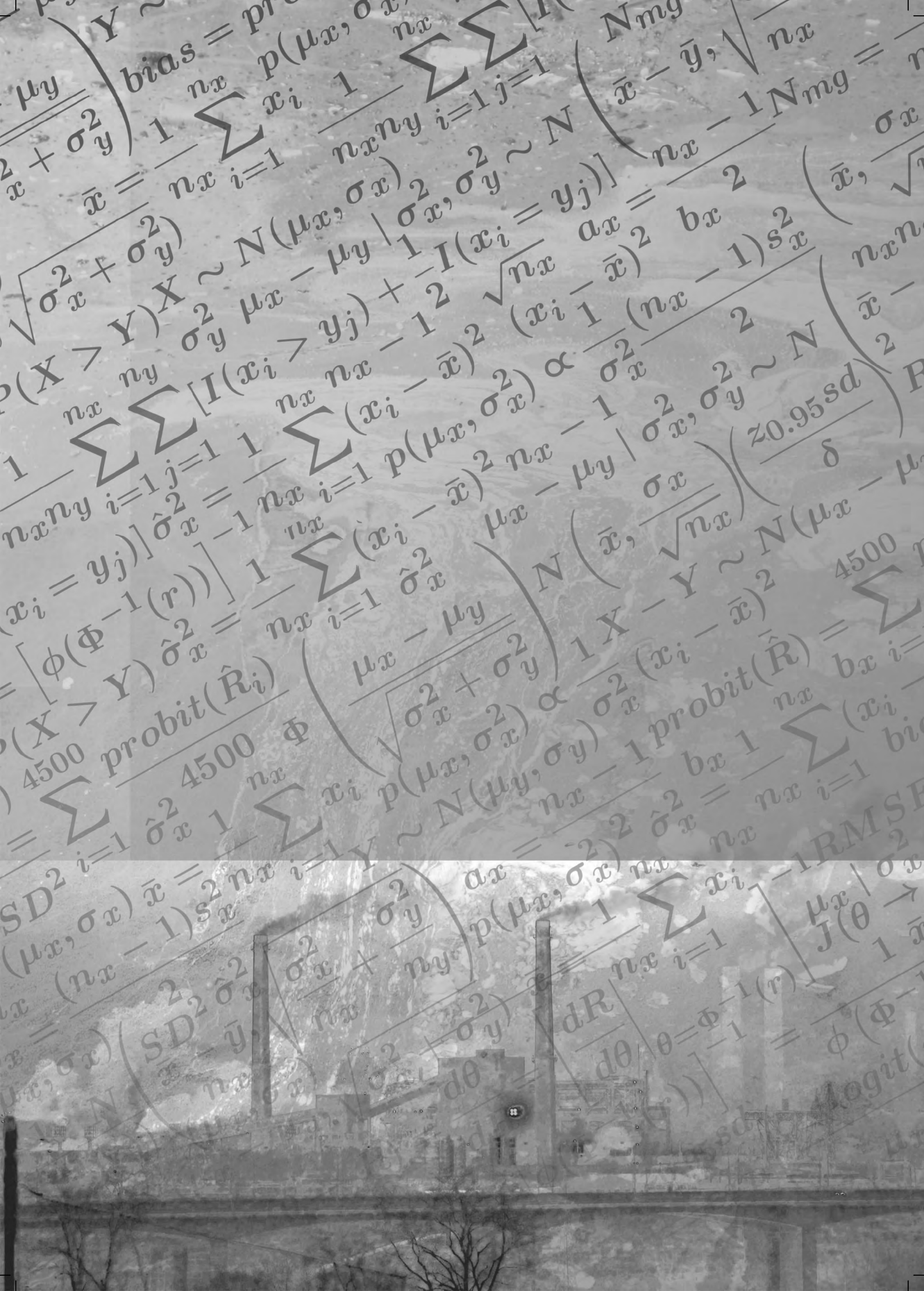
mately subject to varying levels of uncertainty. This also holds for the choice of applying a value of 1, 2 or 10 for $AF_{no-effect}$. This uncertainty was not quantified in our method. It is, however, possible to extend our method to include additional uncertainty sources. These can be added as extra uncertainty factors.

3.5. CONCLUSION

IN this chapter, we developed an integrated probabilistic risk assessment method and applied it to nano TiO_2 in the aquatic environmental compartment. This method allows for separate quantification of the variability and uncertainty in the risk assessment. In this way, we can see which part of the total variation in the CR is due to uncertainty and which part is due to variability. Variability was found to contribute the most. This was mainly due to the large variability in the critical effect concentrations. Furthermore, the uncertainty contribution of the exposure and critical effect to the total uncertainty in the CR was studied. Here we found that the uncertainty in the critical effect is by far the greatest contributor. This conclusion is, of course, dependent on the choice of uncertainty distributions.

We do need to caution that the results do not constitute a fully comprehensive risk assessment. They should, therefore, be interpreted in the context of model development and not as an authoritative aquatic risk assessment of nano TiO_2 . As discussed in Section 3.4.1, there is a need to broaden the scope of nanoparticle forms used in toxicity tests to include hetero-aggregated nanoparticles. This constitutes an important future research area.

We conclude that a probabilistic risk assessment in which variability and uncertainty are quantified separately adds to a more transparent risk assessment. Such a method allows for easy identification of variability and uncertainty sources which in turn can direct further environmental and toxicological research to the areas in which it is most needed.



4

PARAMETRIC ESTIMATION OF $P(X > Y)$ FOR NORMAL DISTRIBUTIONS IN THE CONTEXT OF PROBABILISTIC ENVIRONMENTAL RISK ASSESSMENT

Rianne Jacobs, Andriëtte Bekker, Hilko van der Voet, Cajo J.F. ter Braak

Based on: Jacobs R, Bekker A, van der Voet H, ter Braak CJE (2015) Parametric estimation of $P(X > Y)$ for normal distributions in the context of probabilistic environmental risk assessment. *PeerJ* 3:e1164. doi:10.7717/peerj.1164.

ABSTRACT

ESTIMATING the risk, $R = P(X > Y)$, in probabilistic environmental risk assessment of nanoparticles is a problem when confronted by potentially small risks and small sample sizes of the exposure concentration X and/or the effect concentration Y . This is illustrated in the motivating case study of aquatic risk assessment of nanoAg. A non-parametric estimator based on data alone is not sufficient as it is limited by sample size. In this chapter, we investigate the maximum gain possible when making strong parametric assumptions as opposed to making no parametric assumptions at all. We compare maximum likelihood and Bayesian estimators with the non-parametric estimator and study the influence of sample size and risk on the (interval) estimators via simulation. We found that the parametric estimators enable us to estimate and bound the risk for smaller sample sizes and small risks. Also, the Bayesian estimator outperforms the maximum likelihood estimators in terms of coverage and interval lengths and is, therefore, preferred in our motivating case study.

4.1. INTRODUCTION

LIKE all novel materials, engineered nanoparticles (ENPs) have no history of safe use. A risk assessment is important for the societal acceptance and safe use of ENPs. In order to perform a proper risk assessment, one needs knowledge and data on the properties of nanoparticles. These properties can be different in nanoparticles compared to conventional chemicals in areas such as physicochemical properties, lifecycle, toxicokinetics and environmental fate. This information is hard to come by because of lack of knowledge and technical limitations, resulting in no or only small datasets for effect concentrations of ENPs. In the EU, environmental risk assessment is regulated by the European Chemicals Agency (ECHA) and probabilistic risk assessment is level 3 of their tiered risk assessment approach (ECHA, 2012c).

In the motivating case study of aquatic risk assessment of nanoAg (Gottschalk et al., 2013a), we are confronted with such a small dataset of effect concentrations. Gottschalk et al. (2013a) modeled the exposure of nanoAg from surface water with a probabilistic material flow model (Gottschalk et al., 2010a) to obtain a distribution of exposure concentration values. They collected the effect concentration data from available toxicity studies found in the literature. These effect concentration data consist of toxic endpoints (eq. LC_{50} , EC_{50} , NOEC) for 12 aquatic species. For some of these species there were more than one data point. We averaged these to obtain one value for each species. Histograms and normal density curves of the exposure and effect concentration data are given in Figure 4.1.

In probabilistic risk assessment, the variability of environmental exposure due to natural variation in concentration values over various environments is modelled by an exposure concentration (X) distribution (ECD). Similarly, the variability in effect concentration values due to natural variation among species in their sensitivity to nanoparticles is modelled by a species sensitivity distribution (SSD) or effect concentration (Y) distribution. Probabilistic risk estimation is based on the overlapping of the ECD and the SSD (ECHA, 2012c). The risk, $R = P(X > Y)$, is the area under the curve obtained by multiplying the probability density function (pdf) of the ECD with the cumulative distribution function (cdf) of the SSD. Verdonck et al. (2003) critically discuss this approach to risk assessment.

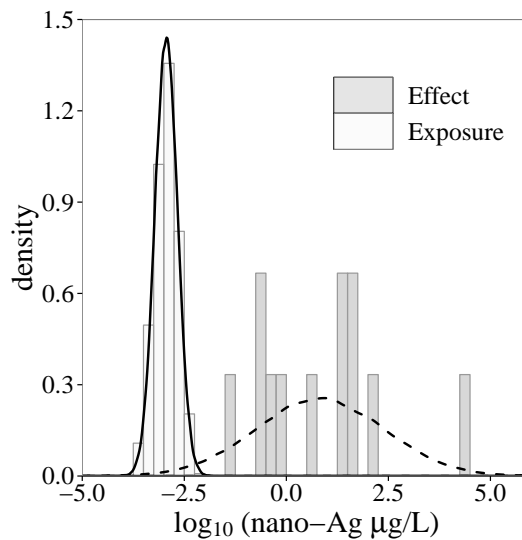


Figure 4.1: Histograms and normal density curves of exposure ($n_x = 1000$) and effect ($n_y = 12$) concentration nanoAg ($\mu\text{g/L}$) from Gottschalk et al. (2013a).

In the ecotoxicological risk assessment literature, $R = P(X > Y)$ as a definition for risk was first developed by Suter et al. (1983). This concept was further developed by van Straalen (2002) as ecological risk (ER). A similar concept, known as expected total risk, was developed by the Water Environment Research Foundation (WERF) (Cardwell et al., 1993; Warren-Hicks et al., 2002). For a visual representation of risk, the exceedance function (1-cdf) of the ECD is plotted against the cdf of the SSD. Such a plot is called a joint probability curve (JPC) (ECOFRAM, 1999; Solomon et al., 2000; Solomon and Takacs, 2002). The area under the JPC is the risk (Solomon and Takacs, 2002). Aldenberg et al. (2002) showed that the area under the JPC, ER and expected total risk are mathematically identical. In this chapter, we refer to the probability, $P(X > Y)$, as the risk, R .

When we consider the case study data, there is no overlap between the effect and the exposure histograms (Figure 4.1). There is no exposure concentration that is greater than an effect concentration, and, therefore, the empirical estimate of $R = P(X > Y)$ is zero for these datasets. To conclude, however, that the true R is zero based on a small sample is rather imprudent. The denial of a probability of zero is referred to as Cromwell's rule by Lindley (1971, pp. 105-106; 2006, pp. 90-91). Several possibilities exist to address the

zero problem empirically. This is further discussed in Section 2.2.4. The zero problem can also be solved by fitting a parametric distribution to the data. When considering the normal density curves in Figure 4.1, we note that there is some overlap between the exposure and effect concentration distributions and, therefore, some non-zero probability of exposure values exceeding effect values.

It is common to assume independent log-normal distributions for the exposure distribution and the SSD (Aldenberg et al., 2002; Verdonck et al., 2003; Wagner and Løkke, 1991). This is the same as assuming normal distributions on the log-transformed exposure and effect concentrations. This normal-normal model was developed in some detail by Aldenberg et al. (2002) and allows an analytic expression for the risk when parameter values are known.

Estimation of $R = P(X > Y)$ is also of interest in other areas such as engineering and medical applications. In engineering, $R = P(X > Y)$ is known as the reliability in stress-strength models. This is a well-known concept and has been studied extensively for the normal-normal model (Barbiero, 2011; Church and Harris, 1970; Downton, 1973; Enis and Geisser, 1971; Govindarajulu, 1967; Nandi and Aich, 1996; Voinov, 1986; Weerahandi and Johnson, 1992) as well as for other distributions (Kundu and Gupta, 2006; Mokhlis, 2005; Nadar et al., 2014). None of these papers gives sufficiently general theory for obtaining trust-worthy interval estimates in the case study. In receiver operating characteristic (ROC) analysis such as used in medical applications, $P(X > Y)$ is known as the area under the ROC curve (Li and Ma, 2011). Although usually used for categorical data, the area under the ROC curve can also be obtained for continuous data in both a non-parametric way and for the normal-normal model (Krzanowski and Hand, 2009).

In this chapter, we will investigate the influence of sample size on the estimation of $R = P(X > Y)$, with special attention to the sample size of effect concentrations. We also investigate the behaviour of the estimators of $R = P(X > Y)$ for small risks. We consider one non-parametric estimator and three parametric estimators, namely, the maximum likelihood estimator (MLE), quasi maximum likelihood estimator (QMLE) and Bayesian estimator with noninformative prior for the normal-normal model. In comparing the parametric estimators with the non-parametric one, we investigate the maximum gain possible when making strong parametric assumptions as opposed to making no parametric assumptions at all. This is done in a simulation study in which we also assess the

accuracy and precision of the estimators and compare them for various combinations of sample sizes and risks.

In Section 2, we derive the estimators and provide the simulation structure. In Section 3, the simulation results will be given and discussed. In Section 4, the application is discussed in the context of the simulation results. Section 5 provides some general discussion, conclusions and recommendations for further study.

4.2. THEORY AND METHODS

4

IN this section, we describe the theory and methodology of our approach. We start by deriving the risk for the normal-normal model. Next we discuss the four estimation methods, provide the simulation structure and discuss the performance measures that we used.

4.2.1. RISK

Let X be the \log_{10} exposure concentration random variable and Y be the \log_{10} effect (or \log_{10} sensitivity) concentration random variable.

In the normal-normal model, the distributions are given by

$$X \sim N(\mu_x, \sigma_x) \text{ and } Y \sim N(\mu_y, \sigma_y).$$

Due to the additive property of the normal distribution we have

$$X - Y \sim N\left(\mu_x - \mu_y, \sqrt{\sigma_x^2 + \sigma_y^2}\right).$$

The risk for the normal-normal model is given by

$$\begin{aligned} R &= P(X > Y) = P(X - Y > 0) \\ &= 1 - \Phi\left(\frac{-(\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right) = \Phi\left(\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right) \end{aligned} \quad (4.1)$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Equation 4.1 is a well-known result (Reiser and Guttman, 1986).

We note location-scale invariance in Equation 4.1. The value of R is determined only by the difference of the expected values and the sum of the variances. The absolute value of the individual parameters is not relevant.

4.2.2. POINT ESTIMATION

In the following sections, we derive the MLE, QMLE, Bayesian estimator and non-parametric estimator for the risk, R . We let $(x_1, x_2, \dots, x_{n_x})$ be a random sample of size n_x of log exposure concentrations and $(y_1, y_2, \dots, y_{n_y})$ an independent random sample of size n_y of log effect concentrations.

MAXIMUM LIKELIHOOD ESTIMATOR

The most straightforward way of estimating R is by means of maximum likelihood estimation. The estimator obtained in this way is denoted as \hat{R}_{MLE} . From the invariance property of MLEs (Bain and Engelhardt, 1992, p. 296), we obtain \hat{R}_{MLE} by substituting the MLEs of μ_x , μ_y , σ_x^2 and σ_y^2 in Equation 4.1. These MLEs are given by

Parameter	Maximum Likelihood Estimator
μ_x	$\bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$
μ_y	$\bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$
σ_x^2	$\hat{\sigma}_x^2 = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$
σ_y^2	$\hat{\sigma}_y^2 = \frac{1}{n_y} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$

Equation 4.1 then becomes

$$\hat{R}_{MLE} = \Phi \left(\frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2}} \right). \quad (4.2)$$

Note that $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are the MLEs of the variance, which are not unbiased.

QUASI MAXIMUM LIKELIHOOD ESTIMATOR

The QMLE is similar to the MLE, differing only in the use of unbiased estimators for σ_x^2 and σ_y^2 instead of the MLEs. We then obtain

$$\hat{R}_{QMLE} = \Phi \left(\frac{\bar{x} - \bar{y}}{\sqrt{s_x^2 + s_y^2}} \right) \quad (4.3)$$

where $s_x^2 = \frac{1}{n_x-1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ and $s_y^2 = \frac{1}{n_y-1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$.

4

BAYESIAN ESTIMATOR

Our third way of estimating R is Bayesian. Whereas maximum likelihood estimation uses the data only, Bayesian estimation combines prior knowledge about the parameter(s) with the data. The prior knowledge is specified by a prior distribution and the information in the data by the likelihood. The prior distribution and the likelihood are then combined into what is called the posterior distribution of the parameter (Gelman et al., 2014). We will derive the joint posterior distribution of the parameters μ_x, μ_y, σ_x^2 and σ_y^2 . This distribution together with Equation 4.1 will provide us with the posterior distribution, $f_R(r)$ of R .

Unfortunately we have often very little prior knowledge. Therefore, we derive \hat{R}_{Bayes} assuming a non-informative prior distribution for the parameters, namely $p(\mu_x, \sigma_x^2) \propto \frac{1}{\sigma_x^2}$ for the joint prior distribution of μ_x, σ_x^2 and $p(\mu_y, \sigma_y^2) \propto \frac{1}{\sigma_y^2}$ for μ_y, σ_y^2 (Gelman et al., 2014, p. 64).

The posterior distributions are then given by (Gelman et al., 2014, p. 65)

- $\mu_x | \sigma_x^2 \sim N \left(\bar{x}, \frac{\sigma_x}{\sqrt{n_x}} \right)$
- $\mu_y | \sigma_y^2 \sim N \left(\bar{y}, \frac{\sigma_y}{\sqrt{n_y}} \right)$
- $\sigma_x^2 \sim \text{Inverse-gamma} \left(\frac{n_x-1}{2}, \frac{(n_x-1)s_x^2}{2} \right)$
- $\sigma_y^2 \sim \text{Inverse-gamma} \left(\frac{n_y-1}{2}, \frac{(n_y-1)s_y^2}{2} \right)$.

From this, we obtain the conditional posterior distribution of $\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}$ (Weerahandi and Johnson, 1992):

$$\begin{aligned} \mu_x - \mu_y \mid \sigma_x^2, \sigma_y^2 &\sim N\left(\bar{x} - \bar{y}, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) \\ \therefore \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} \mid \sigma_x^2, \sigma_y^2 &\sim N\left(\frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \sqrt{\frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2}}\right). \end{aligned}$$

Using the variable transformation method and integrating σ_x^2 and σ_y^2 out of the joint pdf, $f(r, \sigma_x^2, \sigma_y^2)$, we obtain the marginal pdf, $f_R(r)$ (see Appendix Result 4.A.2 for details). This marginal posterior pdf of R (Equation 4.A.4) can be evaluated using numerical integration.

Alternatively, we can use Monte Carlo sampling to approximate the marginal posterior pdf of R . We used the Method of Composition (Lesaffre and Lawson, 2012, pp. 93-94) in which we sample from the known posterior distributions of σ_x^2 and σ_y^2 , then μ_x and μ_y from their known posterior conditional distributions and then apply Equation 4.1 to obtain the corresponding R value. Figure 4.2 shows histograms of samples drawn from the marginal posterior distribution of R (sample size of 10000) together with the marginal pdf computed by numerical integration using Equation 4.A.4 for different sample sizes and R values. It can be seen that the Monte Carlo method gives a very good approximation to the theoretical posterior pdf evaluated by numerical integration.

In this chapter, we obtained the posterior distribution of R by sampling, because it required less computing time than numerical integration in our implementation. The posterior mean is often taken as the Bayesian point estimator but we also investigated the posterior median and mode as point estimators of R .

NON-PARAMETRIC ESTIMATOR

As a benchmark comparison for the parametric estimators, we included a basic non-parametric estimator. This estimator, \hat{R}_{NP} , is calculated from the data without any dis-

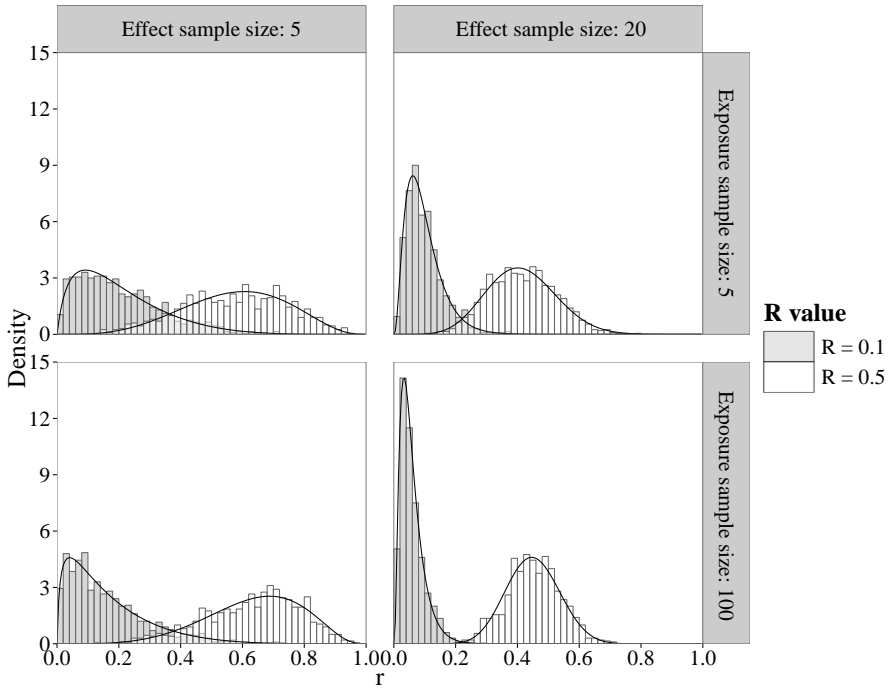


Figure 4.2: Histogram and theoretical posterior pdf (black solid line) of R for sample sizes of 5 and 20 for effect concentrations, sample sizes of 5 and 100 for exposure concentrations and R value of 0.1 and 0.5.

tributational assumptions by

$$\hat{R}_{NP_0} = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} [I(x_i > y_j) + \frac{1}{2} I(x_i = y_j)] \quad (4.4)$$

where $I(\mathcal{S}) = 1$ if \mathcal{S} is true and 0 otherwise (Krzanowski and Hand, 2009, p. 65).

Alternatively, Equation 4.4 can be written as

$$\hat{R}_{NP_0} = \frac{U}{n_x n_y}$$

where U is the Mann-Whitney statistic (Gibbons and Chakraborti, 2011). Equation 4.4 is

also known as the area under the ROC curve and its equivalence to the Mann-Whitney statistic has been shown (Bamber, 1975).

The non-parametric estimator of R is related to estimating the success probability, p , in a binomial experiment. As noted in the Introduction, we encounter the zero problem. One possible solution is making use of Laplace's Law of Succession (Zabell, 1989). This law states that given k successes in n trials of a binomial experiment, the probability of a success on the next trial is $\frac{k+1}{n+2}$. The validity of this expression has a Bayesian basis. On assuming a uniform prior for p , the posterior distribution of p is a $Beta(k+1, n-k+1)$ distribution (Gelman et al., 2014, p. 30), so that the posterior mean is $\frac{k+1}{n+2}$. This expression, denoted $\hat{R}_{NP_{LLS}}$, is then used instead of the estimator in Equation 4.4. Note that the posterior mean is equal to the predictive probability of a success on the next trial.

An alternative solution is to replace the zero with some non-zero value. One option is to estimate the probability of an outcome outside the range of the data as $\frac{1}{2n_x n_y}$. This method is used by Matlab and Genstat to compute quantiles. Another alternative is to use $\frac{1}{n_x n_y + 1}$ which is used by Minitab and SPSS.

4.2.3. INTERVAL ESTIMATION

We propose interval estimators by calculating credible intervals for Bayesian methods and confidence intervals for others. For the Bayesian estimator, we calculated 90% two-sided highest posterior density (HPD) credible intervals (Box and Tiao, 1973, p. 123). These intervals are obtained by finding the interval of the posterior distribution with the highest density, for which we used the 'HPDinterval' function in the 'coda' package in R (Plummer et al., 2006). HPD intervals produce the shortest intervals on a chosen scale, e.g. R or a transformation thereof, but are not transformation invariant. To estimate an upper credible bound of the risk, we also calculated the 95% percentile of the posterior. The upper bound of the 90% two-sided HPD intervals is not necessarily equal to the 95% percentile as the probabilities to the left and right of the two-sided HPD interval can be unequal.

For the non-Bayesian estimators, we calculated 90% Bias corrected and accelerated (BCa) parametric bootstrap confidence intervals using the 'boot' package in R (Canty and Rip-

ley, 2013; Davison and Hinkley, 1997) with 1000 bootstrap samples. For the non-parametric estimator, the BCa interval algorithm did not converge for small R values and also had some difficulty with the small sample sizes. For these cases we calculated percentile confidence intervals. The percentile method obtains a symmetric $100(1 - \alpha)\%$ confidence interval by calculating the $(\frac{\alpha}{2})^{th}$ and $(1 - \frac{\alpha}{2})^{th}$ percentiles of the bootstrap sample. In the BCa method, these percentiles are adjusted to correct for bias and skewness. For symmetric distributions, the percentile and BCa intervals are equal. Both intervals are also transformation invariant (Efron and Tibshirani, 1993, pp. 175, 187). When calculating the confidence intervals for small risks, all bootstrap values may be equal, resulting in a zero interval length. For the BCa and percentile interval, the upper bound of a 90% two-sided interval is equal to the upper bound of a 95% one-sided interval (Carpenter and Bithell, 2000). The upper 95% confidence bound is, therefore, trivially obtained from the 90% two-sided interval.

For the MLE-like estimators, we also calculated confidence intervals based on the non-central t -distribution (Reiser and Guttman, 1986). In this method, the sum of the two variances (s_x^2 and s_y^2) are approximated with a chi-squared distribution.

The upper confidence (credible) bounds are of special interest in the context of managing risks, as they indicate (with some certainty) that the risk will not be higher than the upper bound.

4.2.4. SIMULATION STUDY

In this section, we discuss the design of the Monte Carlo (MC) simulation study following the guidelines provided in Burton et al. (2006).

SIMULATION SETUP

The simulation study is performed in R (R Core Team, 2015). We use the built-in `rnorm` function to sample from a normal distribution using the Mersenne-Twister pseudo-random number generator (Matsumoto and Nishimura, 1998). Starting seeds for the different scenarios were drawn from a discrete uniform distribution to produce independent samples for each sample size scenario. The four estimators are calculated on the same sample, thereby avoiding differences among the estimators due to sampling.

Substance	Sample size		Risks			
	Aquatic	Soil	1	2	3	4
Ag	12	1	0.007	0.397	0	0
CNT	9	2	0	0	0	0
TiO2	18	2	7.2e-13	0.187	0	1.2e-7
ZnO	17	2	0	0.011	0	0
Fullerenes	4		0	0	0	0

Table 4.1: Substances, sample sizes and estimated risks in an environmental risk assessment performed by Gottschalk et al. (2013a). Sample size of effect concentration data are given for aquatic and soil toxicity. Risks are given for four environmental compartments.

To make the simulation as realistic as possible, we chose scenarios that are in line with recent studies of environmental risk assessment. When exposures are measured, it is common to have small sample sizes (Johnson et al., 2011; Westerhoff et al., 2011), whereas any number of exposure values can be obtained when they are modeled (Gottschalk et al., 2013a). For the exposure sample size, therefore, we chose two scenarios: the case of a small number of exposures ($n_x = 5$) and the case of a (relatively) large number of exposures ($n_x = 100$). We chose sample size of effect concentrations and risks loosely suggested by data from Gottschalk et al. (2013a). From this data (Table 4.1), we chose the following scenarios:

- Sample sizes for effect concentrations (n_y): 2, 5, 12, 20, 100
- Risks: a grid of values from $1e-14$ to 0.5.

The sample size of 100 for effect concentrations was added to study the influence of a large sample size. A risk of 0.5 is obtained when $\mu_x = \mu_y$. We, therefore, chose increasing values of $\mu_x - \mu_y$ to obtain the required range of risks. Considering the standard deviations, we note that the standard deviation of the effect concentration data in the case study is 5.6 times larger than that of the exposure concentration data. Based on this, we chose three scenarios: $\sigma_y = \sigma_x$, $\sigma_y = \frac{1}{5}\sigma_x$ and $\sigma_y = 5\sigma_x$.

The number of simulations was determined by running a pilot simulation (1000 simulations) for the MLE. From this pilot, we obtained the median empirical standard deviation ($sd = 0.0719496$) of \hat{R}_{MLE} and the median absolute bias ($\delta = 0.002107476$) in \hat{R}_{MLE} over all scenarios. These were used to calculate the number of simulations, B , according to

Burton et al. (2006)

$$\begin{aligned} B &= \left(\frac{z_{0.95} sd}{\delta} \right)^2 \\ &= \left(\frac{1.96 \cdot 0.0719496}{0.002107476} \right)^2 \\ &= 4477.58 \approx 4500. \end{aligned}$$

For each of the 4500 MC simulations, we calculated \hat{R}_{MLE} , \hat{R}_{QMLE} , \hat{R}_{Bayes} and \hat{R}_{NP} . Due to the skewness of their sampling distributions, especially for small R values, we decided to use a transformation. Due to the nature of the analytical expression for R , (see Equation 4.1), a probit (inverse standard normal cdf) transformation is a natural choice. Some comparisons between the original scale and the probit transformation are further discussed for the Bayesian case in Section 4.3.1.

Simulations were run on a HP Desktop Computer running Microsoft Windows 7 with processor specification Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz, 3401 Mhz, 4 Core(s), 8 Logical Processor(s). Three R-sessions were running the three cases $\sigma_y = \sigma_x$, $\sigma_y = \frac{1}{5}\sigma_x$ and $\sigma_y = 5\sigma_x$ simultaneously. The $\sigma_y = \sigma_x$ case took the longest with the following time (in hh:mm:sec) for each of the four estimators:

- MLE: 03:43:06.18 (bootstrap); 00:10:21.82 (noncentral t)
- QMLE: 03:35:36.44 (bootstrap); 00:10:13.30 (noncentral t)
- Bayes: 03:12:56 (sample size 10000); 00:41:02.9 (sample size 1000)
- Non-parametric: 19:13:17.12.

The bootstrap of the MLE, QMLE and non-parametric estimator was the cause of the longer runtime. The runtime for the Bayesian estimator is directly related to the size of the posterior sample. All further results are given for the large sample case.

PERFORMANCE MEASURES

We calculated various performance measures to evaluate the performance of the four point estimators. We calculated the performance measures on the probit scale, so as to be able to highlight differences among methods for small values of R :

- mean: $probit(\hat{R}) = \sum_{i=1}^{4500} \frac{probit(\hat{R}_i)}{4500}$
- absolute bias: $bias = probit(\hat{R}) - probit(R)$
- empirical (or MC) standard deviation: $SD = \sqrt{\sum_{i=1}^{4500} \frac{(probit(\hat{R}_i) - probit(\hat{R}))^2}{4500}}$
- root mean squared error: $RMSE = \sqrt{bias^2 + SD^2}$.

The quality of the interval estimators will be assessed by calculating the coverage probability for each scenario. Note that the confidence intervals we calculated are approximate and do not claim to deliver the correct coverage. In addition, Bayesian credible intervals also do not claim a coverage frequency. For each of the 4500 simulations, we calculated the confidence (credible) intervals and calculated the proportion of intervals that contained the true R value. We also investigated lengths of confidence (credible) interval over the 4500 simulations for each scenario and R value. Coverages and lengths of confidence intervals for the different non-parametric estimators were similar. We, therefore, only consider the estimator based on Laplace's Law of Succession.

4

4.3. SIMULATION RESULTS

ALL results given and discussed are for the scenario that most resembles the case study ($\sigma_y = 5\sigma_x$ and sample sizes $n_x = 100$ and $n_y = 12$) unless explicitly stated otherwise. Graphs and tables for the other scenarios are given in the Supplementary Information in Jacobs et al. (2015a). All sampling distribution graphs plot the estimated $probit(\hat{R})$ (or \hat{R}) value (y-axis) against the true $probit(R)$ (or R) value (x-axis). A diagonal 1-1 line is drawn to indicate where $probit(\hat{R}) = probit(R)$ (or $\hat{R} = R$). A logarithmic scale is used when R is plotted.

4.3.1. COMPARISON OF BAYESIAN POINT ESTIMATORS

For Bayesian estimation, we considered three point estimators, namely, the posterior mean, median and mode. We summarize the sampling distribution of each by way of three quantiles (0.5 or median, 0.025 and 0.975). In Figure 4.3 these quantiles are plotted as a function of the true value. The left hand panels (A, C) show the median and the right hand panels (B, D) show the 0.025 and 0.975 quantiles.

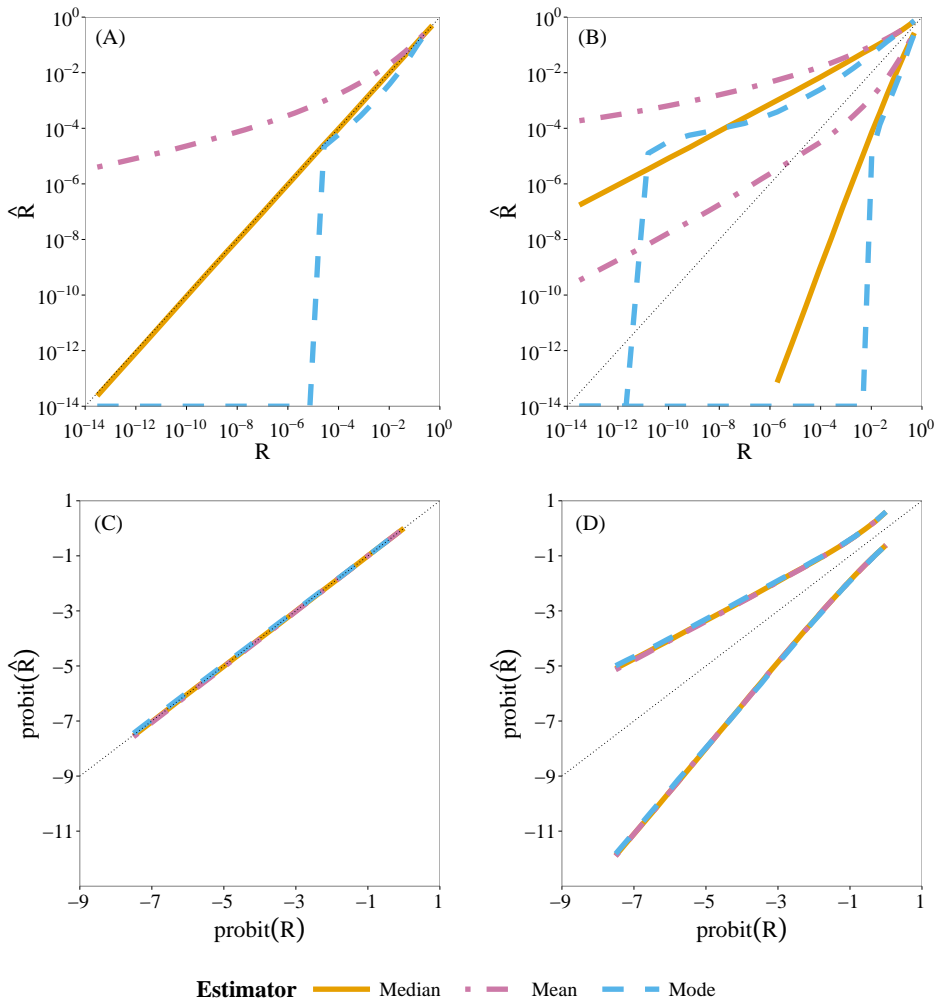


Figure 4.3: The median (A, C) and 0.025 and 0.975 quantiles (B, D) of the sampling distribution of the three Bayesian point estimators (mean, median and mode) calculated on the original scale (A, B) and on the probit scale (C, D). When plotted on \log_{10} -scale, a zero mode becomes $-\infty$. The diagonal dotted line represents the values where $\text{probit}(\hat{R}) = \text{probit}(R)$ (or $\hat{R} = R$).

In the top panels (A, B), the quantiles are calculated from the sampling distribution of the estimators on the original scale (\hat{R}) and plotted on \log_{10} -scale. The lines for the posterior mean are above the 1:1 line, so indicating large positive bias. The lines for the posterior

mode go to $\log_{10}(0) = -\infty$, due to the very skew posterior distributions for smaller risks (on original scale of R , as already illustrated in Figure 4.2), so indicating large negative bias. The lines for the posterior median are in between and closer to the 1:1 line.

In the bottom panels (C, D), the quantiles are calculated similarly but on the probit-transformed \hat{R} and plotted on the probit scale as well. Here we see that quantiles of the posterior mean, median and mode almost coincide, indicating that the skewness problem is solved. Simulations for very small sample size of effect concentrations ($n_y = 2$) showed that the mean has a slight advantage because of narrower intervals between the 0.025 and 0.975 quantiles than that of the median and the mode. This difference, however, is very quickly lost for higher sample sizes ($n_y \geq 5$) as shown in Figure S23 (Supplementary Information in Jacobs et al. (2015a)).

From this study of Bayesian point estimators on different scales, we see the advantage of the use of the probit scale for the Bayesian case. Moreover, for ease and its transformation invariance, we chose the posterior median as the Bayesian point estimator of R . The probit scale stretches out small values of R , making possible differences between methods more clearly visible for small R . On this basis, we decided to perform, for all estimators, all further calculations in the simulation study on the probit scale.

4.3.2. COMPARISON OF THE FOUR POINT AND INTERVAL ESTIMATORS

In this section, we show the simulation results for the four estimators.

We first compare the sampling distributions of the four estimators. Figure 4.4 illustrates the median (A) and 0.025 and 0.975 quantiles (B) of the sampling distributions of \hat{R}_{MLE} , \hat{R}_{QMLE} , \hat{R}_{Bayes} and \hat{R}_{NP} . For \hat{R}_{NP} , we plotted both the standard estimator, \hat{R}_{NP_0} , (Equation 4.4) which goes to minus infinity on the probit scale and the Laplace version, $\hat{R}_{NP_{LLS}}$. These provide the extreme endpoints of the different solutions in solving the zero problem in the non-parametric estimator.

The median of the Bayesian estimator lies closest to the true R (Figure 4.4A). This is especially apparent in scenarios with $n_y \leq 12$ (Figure S24 in Supplementary Information in Jacobs et al. (2015a)). The non-parametric estimators are clearly not able to estimate R for smaller values as they very quickly jump to their lower bound of either $\text{probit}\left(\frac{1}{n_x \times n_y + 2}\right)$

or minus infinity, indicated by horizontal and vertical dash-dot lines respectively. As the sample sizes increase, the three parametric estimators converge (Figures S24 and S25 in Supplementary Information in Jacobs et al. (2015a)). The non-parametric estimators remain the worst estimators for all sample sizes when estimating small R values. In our further study, we consider the Laplace version only.

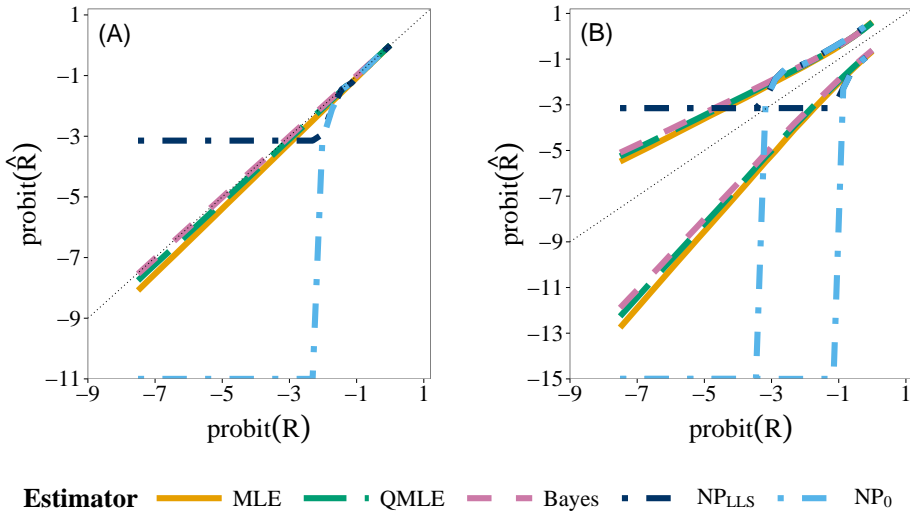


Figure 4.4: The median (A) and 0.025 and 0.975 quantiles (B) of the sampling distribution of the point estimators, \hat{R}_{MLE} , \hat{R}_{QMLE} , \hat{R}_{Bayes} , $\hat{R}_{NP_{LLS}}$ and \hat{R}_{NP_0} calculated on the probit scale. The diagonal dotted line represents the values where $\text{probit}(\hat{R}) = \text{probit}(R)$.

Next, we study the coverage and interval lengths of the two-sided 90% confidence (credible) intervals of the estimators on the probit scale. For each of the 4500 simulations, we calculated interval lengths and then obtained the median interval length for each combination of estimator, sample sizes and risk value combinations. In order to compare the median interval lengths across different risk values in a single graph, we standardised each one by dividing by the true $|\text{probit}(R)|$ value to obtain the relative median interval length. Figure 4.5 plots the relative median interval length (y-axis) against the coverage probabilities (x-axis) for $n_x = 5$ and $n_x = 100$, all investigated sample sizes for effect concentrations, and all R values. Coverage probabilities of less than 0.5 were plotted at 0.5. The vertical line indicates a coverage probability of 90%. A good interval estimator gives

points lying on this line with short interval length. This translates to good coverage and narrow intervals. Points corresponding to $n_y = 12$ are indicated by an open black circle.

We found that the MLE (not shown) and the QMLE had a similar pattern for both the bootstrap and noncentral t intervals, with the QMLE consistently having better coverage. Figure 4.5 shows that the bootstrap intervals are liberal compared to the noncentral t intervals for small sample sizes of effect concentrations. As the sample sizes for effect concentrations increase (bigger dots), the estimators have a better coverage. Very small sample size for effect concentrations ($n_y = 2$) gives the worst coverage. For the Bayesian estimator, the sample size has a lesser influence. For small exposure sample size ($n_x = 5$), the coverage of the Bayesian interval estimator tends to be too high. This problem largely disappears for $n_x = 100$, although there is some under-coverage for the $n_y = 2$ case. The parametric estimators have shorter interval lengths when $n_x = 100$ (right column) as compared to $n_x = 5$. For the non-parametric estimator, sample size has a slightly less systematic influence on the coverage.

Compared to the other estimators, the Bayesian interval estimator best maintains the nominal coverage without having larger median interval length and despite the fact of often having a higher than nominal coverage (Figure 4.5). The non-Bayesian estimators have smaller than nominal coverage for small sample size of effect concentrations with the non-parametric estimator being the worst. For better comparison of interval lengths among the estimators, the reader is referred to Figure S26 in the Supplementary Information in Jacobs et al. (2015a).

In risk assessment, one is often interested in an upper bound on the risk. We studied the coverages and interval lengths of the upper 95% confidence (credible) bounds of the estimators on the probit scale. The interval lengths were quantified as the difference between the upper bound and the true $probit(R)$ value. The median of the 4500 differences was obtained. In order to compare the median differences across different risk values in a single graph, each median difference was divided by the true $|probit(R)|$ value being estimated to obtain the relative median difference. Figure 4.6 plots the relative median difference (y-axis) against the coverage probabilities (x-axis) for $n_x = 5$ and $n_x = 100$, all investigated sample sizes for effect concentrations, and all R values. Coverage probabilities of less than 0.5 were plotted at 0.5. The vertical line specifies a coverage probability of 95%. A good upper bound estimator gives points lying on this vertical line

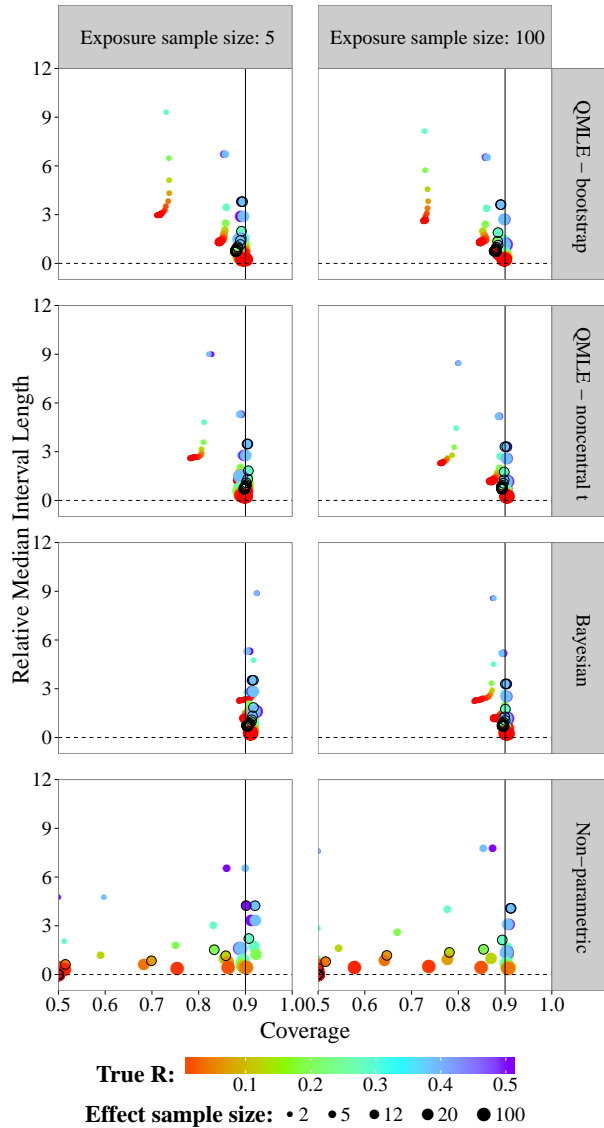


Figure 4.5: Scatterplots of the 90% two-sided coverage probabilities against the relative median interval length calculated on the probit scale. The value of the true R value is illustrated by the colour scale. The size of the dots corresponds to the size of the sample size of effect concentrations. A vertical reference line is drawn at a coverage probability of 90%. The points corresponding to $n_Y = 12$ are indicated by an open black circle.

and being close to the horizontal 0 line. This translates to good coverage and an upper bounds close to the true R . Points corresponding to $n_y = 12$ are indicated by an open black circle.

We see similar patterns as in the case of the two-sided intervals, with the bootstrap intervals being too liberal. The Bayes estimator gives higher than nominal coverage for small exposure sample size (left column). For large exposure sample size, the Bayesian estimator clearly outperforms the other estimators with good coverage for all R values and sample sizes for effect concentrations without having larger median interval difference. The non-parametric estimator has a severe coverage problem.

The results for the performance measures of the different estimators are given in Tables S2, S3, S4 and S5 (see Supplementary Information in Jacobs et al. (2015a)). Due to the lower bound of the non-parametric estimator, the SD, bias and RMSE are not reliable for small R values. Only for a few cases where $R = 0.5$, the non-parametric estimator has slightly lower SD and bias than the parametric estimators. The various graphs have also shown the inability of the non-parametric estimator to estimate small R values.

Among the parametric estimators, the Bayesian estimator as the smallest SD, bias and RMSE on probit scale for all sample sizes and R values. This confirms that the Bayesian estimator is better than the non-Bayesian estimators as also seen for the interval estimator case. The QMLE has smaller SD, bias and RMSE than the MLE for all sample sizes and R values. This also confirms the results of the interval estimators where QMLE has better coverage than MLE.

The Bayesian estimator was in general the best estimator and specifically so for the scenario that is closest to the case study. The Bayesian point estimator was less biased than the MLE and the QMLE in all cases ($\sigma_y = \sigma_x$ (Figures S8 and S9), $\sigma_y = \frac{1}{5}\sigma_x$ (Figures S16 and S17) and $\sigma_y = 5\sigma_x$ (Figures S24 and S25) in Supplementary Information in Jacobs et al. (2015a)), and this was especially apparent for small sample sizes for effect concentration ($n_y \leq 12$) and small R values. The Bayesian interval estimator (90% two-sided) had better coverage, with even higher than nominal coverage for exposure sample size $n_x = 5$. This is also seen for the case $\sigma_y = \sigma_x$ (Figure S10 in Supplementary Information in Jacobs et al. (2015a)). For the case $\sigma_y = \frac{1}{5}\sigma_x$ (Figure S18 in Supplementary Information in Jacobs et al. (2015a)), the higher coverage is only seen for small sample size for

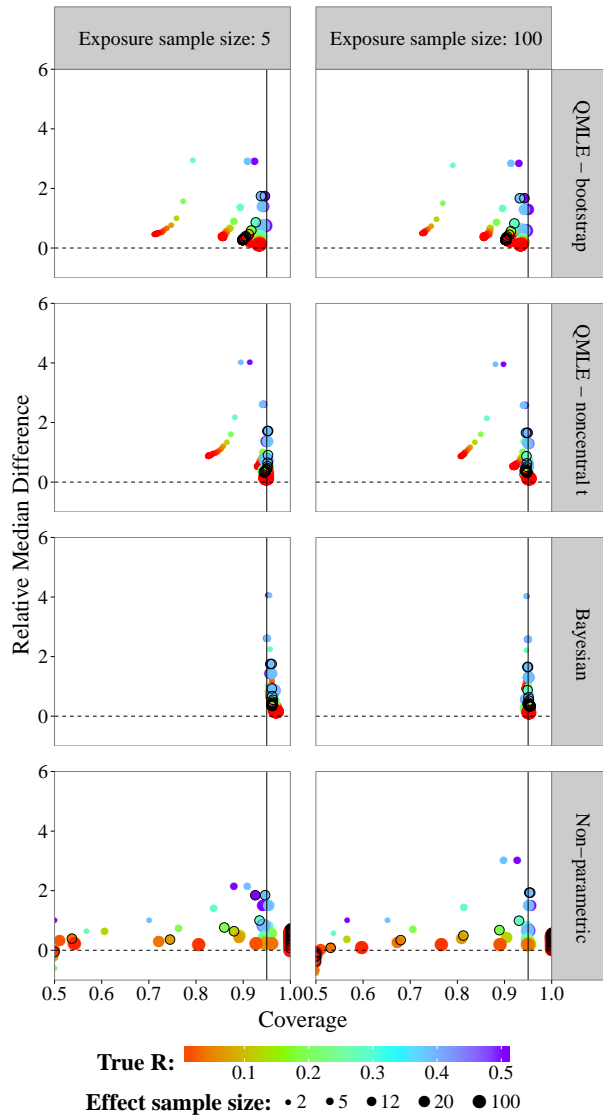


Figure 4.6: Scatterplots of the 95% one-sided coverage probabilities against the relative median difference calculated on the probit scale. The value of the true R value is illustrated by the colour scale. The size of the dots corresponds to the size of the sample size of effect. A vertical reference line is drawn at a coverage probability of 95%. The points corresponding to $n_y = 12$ are indicated by a black circle.

effect concentrations as seen by the small dots. When considering the 95% upper bound of the Bayesian estimator compared to MLE and QMLE, we also see better coverage with similar higher than nominal coverage for exposure sample size $n_x = 5$. This is similar for the $\sigma_y = \sigma_x$ case (Figure S11 in Supplementary Information in Jacobs et al. (2015a)) and slightly more pronounced for the $\sigma_y = \frac{1}{5}\sigma_x$ case (Figure S19 in Supplementary Information in Jacobs et al. (2015a)). Considering the performance measures, the Bayesian estimator performs better (lower values), also for the $\sigma_y = \sigma_x$ case (Tables S2-S4 in Supplementary Information in Jacobs et al. (2015a)) and the $\sigma_y = \frac{1}{5}\sigma_x$ case (Tables S6-S8 in Supplementary Information in Jacobs et al. (2015a)). For the corresponding case study scenario of $\sigma_y = 5\sigma_x$ and $n_x = 100, n_y = 12$, the Bayesian estimator clearly outperformed the MLE and QMLE. It was less biased and maintained the nominal coverage in both the two-sided and one-sided cases. For better comparison of interval lengths among the estimators, the reader is referred to Figure S27 in the Supplementary Information in Jacobs et al. (2015a).

4.4. CASE STUDY

IN this section, we evaluate the case study results on the basis of the simulation study results. In the case study, we have a sample size 1000 of exposure concentrations and a sample size of 12 of effect concentrations. We note that the exposure concentrations come from a simulation model, so that it is possible to generate an arbitrary large sample exposure concentrations. We treat the size of 1000 exposures as being effectively of size 100.

First, we verify that the normal-normal model is not in conflict with the data. Visually, the normal distribution fits the concentration data quite well (Figure 4.1). The small sample size of the effect concentrations gives low power to any formal normality tests, where the large sample size of exposure concentrations gives high power, so that even small deviations from normality are detected. Even so, we cannot reject the null hypothesis of normality (see Table S1 in Supplementary Information in Jacobs et al. (2015a)) for either the effect or the exposure samples at a 5% significance level. In the exposure concentration data, there is some indication for non-normality evident from two of the

normality tests which are only just not significant (p-values of 0.0564 and 0.0538). Even so, we take the normal-normal model as a useful model.

Estimator	\hat{R}	90% 2-sided CI	95% upper CB
MLE	0.0068	0.0003 - 0.0684	0.0684
QMLE (noncentral t)	0.0090	0.0006 - 0.0784	0.0784
QMLE (bootstrap)	0.0090	0.0002 - 0.0571	0.0571
Bayes	0.0108	0.0006 - 0.0776	0.0806
Empirical	0.0001	0.0001 - 0.0001	0.0001

Table 4.2: Estimated risks (\hat{R}), 90% two-sided confidence (credible) intervals (CI) and 95% upper confidence (credible) bounds (CB) for the MLE, QMLE (bootstrap and noncentral t), Bayesian and non-parametric estimator

Next, we consider the estimates of the risk (Table 4.2). The estimates and intervals were calculated on the probit-scale and then transformed back to the original scale so as to be able to evaluate the case study results in the light of the simulation study results. For the MLE we calculated the interval estimators based on the noncentral t -distribution and for the QMLE, the noncentral t and parametric bootstrap .

The non-parametric estimator was calculated using Laplace's Law of Succession. For the sample sizes of this case study, $\hat{R}_{NP_{LLS}}$ then becomes $\frac{0+1}{1000 \cdot 12+2} = \frac{1}{12002} = 0.000083$. We note, however, that this value is very much dependent on the sample size. A larger exposure sample size will decrease the estimate. As seen in the simulation study results, it is impossible to draw any meaningful conclusions for small risks based on the non-parametric estimator.

The three parametric estimates are similar. The bootstrap 90% confidence interval of the QMLE is clearly narrower. From the simulation study, however, the bootstrap intervals showed liberal coverage and are, therefore, less trustworthy. Considering the 95% upper confidence bound, we note that the Bayesian bound is slightly higher than that of the MLE and QMLE and higher as well than the upper bound of the Bayesian 90% credible interval. Investigating these aspects in the simulation results, we found that these differences are to be expected (see Figures S1-S3 in Supplementary Information in Jacobs et al. (2015a)), although the difference between the Bayesian 95% upper bound and the upper bound of its 90% credible interval is not so typical. The distances between the 95% upper bound and \hat{R} as well as the ratio of the Bayesian upper bound to both the

QMLE upper bound and the Bayesian upper bound of the two-sided interval fall within the respective sampling distributions as obtained in the simulations.

Based on the simulation results, we, therefore, conclude that the Bayesian estimate is the most appropriate. The upper bound (0.0806) is most reliable as it has the best coverage (compared to MLE and QMLE). This is clearly seen by the black circles in the Bayesian panel in the right column of Figure 4.6. This case corresponds most closely to the case study data. Based on the model and the data used, we state with 95% confidence, that the risk will not be greater than 0.0806.

4.5. DISCUSSION AND CONCLUSION

IN this chapter we studied the problem of estimating the risk for the case of small sample size for effect concentrations and small R values. The case study data showed discrepancies between the parametric and non-parametric estimators which we investigated via a simulation study. We derived and compared three parametric estimators and one non-parametric estimator for the risk. This was done under the assumption of normality for both the exposure and effect concentration data. We found that, overall, the parametric estimators have better performance than the non-parametric estimator, and the Bayesian estimator outperformed the maximum likelihood-based ones.

The Bayesian estimator in this chapter was based on a non-informative prior on the underlying parameters. This resulted in a prior tendency of R toward 0.5. For small sample sizes, there was not enough data to counter this prior tendency and this resulted in an overestimation of R by the posterior mean estimator calculated on the original R scale (Figures S4, S5, S12, S13, S20 and S21 in Supplementary Information in Jacobs et al. (2015a)). To overcome this problem, it was needed to switch to the posterior median estimator or to switch to the *probit*(R) scale. We used both the probit-scale and the posterior median resulting in an estimator that outperformed its parametric counterparts. More benefit can presumably be obtained from the Bayesian estimator if we can use an informative prior, at least when the prior is not in conflict with the data. In addition, the use of probability matching priors (Datta and Sweeting, 2005) may also improve on the coverage of the credible intervals. Ventura and Racugno (2011) used a strong matching prior for Bayesian estimation of $P(X > Y)$ based on a profile-likelihood approach.

Using the *probit*(R) scale in the simulation study enabled us to more easily compare the estimators for small R values. Despite giving nice statistical properties, the probit scale may not directly address a specific risk assessment question.

Comparing the parametric bootstrap and noncentral t interval estimators for the MLE and the QMLE, we found the noncentral t intervals to have better coverage. The bootstrap intervals, although a good alternative, are liberal in coverage (i.e. resulted in smaller than nominal coverage) for small sample sizes of effect concentrations. This was also found by Tian (2008).

It was clearly seen that the non-parametric estimator was not able to estimate the risk for small sample sizes and small R values. For R values above the lower bound of $\text{probit}\left(\frac{1}{n_x n_y + 2}\right)$, the non-parametric estimator had performance measures similar to that of the MLE. As seen in our case study, however, the non-parametric estimator failed completely. The bootstrap cannot provide any variability of outcome with which to provide an interval for the estimate. In the simulation study, we also found that for small sample sizes, there was often too little variability in the data for the bootstrap to be able to quantify it. Although this translated to 0 coverage in Figures 4.5 and 4.6, it really shows that the non-parametric estimator completely fails in these cases. For small sample sizes and small R values, therefore, we advise to use parametric estimators.

Considering the computation times of the simulation study, we note that, in addition to the Bayesian estimator being the best estimator, it can also require shorter computation time compared to the bootstrap alternatives depending on the posterior sample size. The larger posterior sample size (10000) tends to result in slightly narrower estimates of the posterior distribution than those based on the smaller sample size (1000). Nevertheless, the main results in Figures 4.5 and 4.6 remain basically unchanged. In the case study, the credible interval becomes slightly wider for the larger sample size and the upper bound is slightly lower. Even so, not much is lost by taking the smaller sample size and this drastically reduces the computation time. The maximum likelihood based estimators have shorter computation time when calculating the interval estimators based on the noncentral t -distribution than the Bayesian estimator. The non-parametric estimator is by far the most computationally demanding due to the bootstrapping and the calculation of Equation 4.4.

Assuming normality in the case and simulation study may seem as a strict assumption and going non-parametric is a way to avoid strict assumptions. For many situations in statistics the normal distribution is considered to have too light tails. In our case with very little data, going non-parametric leads to zero tails outside the range of the data. The usual area-under-the-curve-based non-parametric method can then severely underestimate the risk (often resulting in zero risk), whereas the estimate based on Laplace's Law of Succession overestimates the risk for small true risks. To be able to draw any sensible conclusion, one has to use a parametric method. Our comparison of methods shows the advantage of using parametric methods in this case.

We conclude that making parametric assumptions, enabled us to estimate the risk for smaller sample sizes and small risks in the case the data is in fact normally distributed. Further research is needed to investigate the robustness of the parametric methods on non-normal data. We need to investigate whether semi-parametric methods and methods based on the extreme value distribution are able to estimate the tails of distributions sufficiently well from small data sets, so that they outperform the parametric methods used in this chapter.

Appendix

4.A. RESULTS

Result 4.A.1. Transformation method for obtaining the pdf of a function, $R = \Phi(\theta)$, from the pdf of θ .

Let $\theta \sim N(\mu, \sigma)$.

Then the pdf of R is given by $f_R(r) = f_\theta(r) |J(\theta \rightarrow R)|$. We first obtain the Jacobian, $J(\theta \rightarrow R)$:

$$\begin{aligned}
 J(\theta \rightarrow R) &= \frac{d\theta}{dR} \\
 &= \left[\frac{dR}{d\theta} \Big|_{\theta=\Phi^{-1}(r)} \right]^{-1} \\
 &= \left[\Phi'(\theta) \Big|_{\theta=\Phi^{-1}(r)} \right]^{-1} \\
 &= [\phi(\Phi^{-1}(r))]^{-1} \\
 &= \frac{1}{\phi(\Phi^{-1}(r))}
 \end{aligned} \tag{4.A.1}$$

where ϕ denotes the pdf of the standard normal distribution.

We then obtain $f_R(r)$:

$$\begin{aligned}
 f_R(r) &= f_\theta(r) |J(\theta \rightarrow R)| \\
 &= f_\theta(\Phi^{-1}(r)) \frac{1}{\phi(\Phi^{-1}(r))} \\
 &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[-\frac{(\Phi^{-1}(r) - \mu)^2}{2\sigma^2} \right] \frac{1}{\frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(\Phi^{-1}(r))^2}{2} \right]} \\
 &= \frac{1}{\sqrt{\sigma^2}} \exp \left[-\frac{(\Phi^{-1}(r) - \mu)^2}{2\sigma^2} + \frac{(\Phi^{-1}(r))^2}{2} \right].
 \end{aligned} \tag{4.A.2}$$

Result 4.A.2. From the conditional posterior distribution of $\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}$ given by

$$\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} \mid \sigma_x^2, \sigma_y^2 \sim N \left(\frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \sqrt{\frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2}} \right),$$

we obtain the conditional posterior distribution of R (using Result 4.A.1):

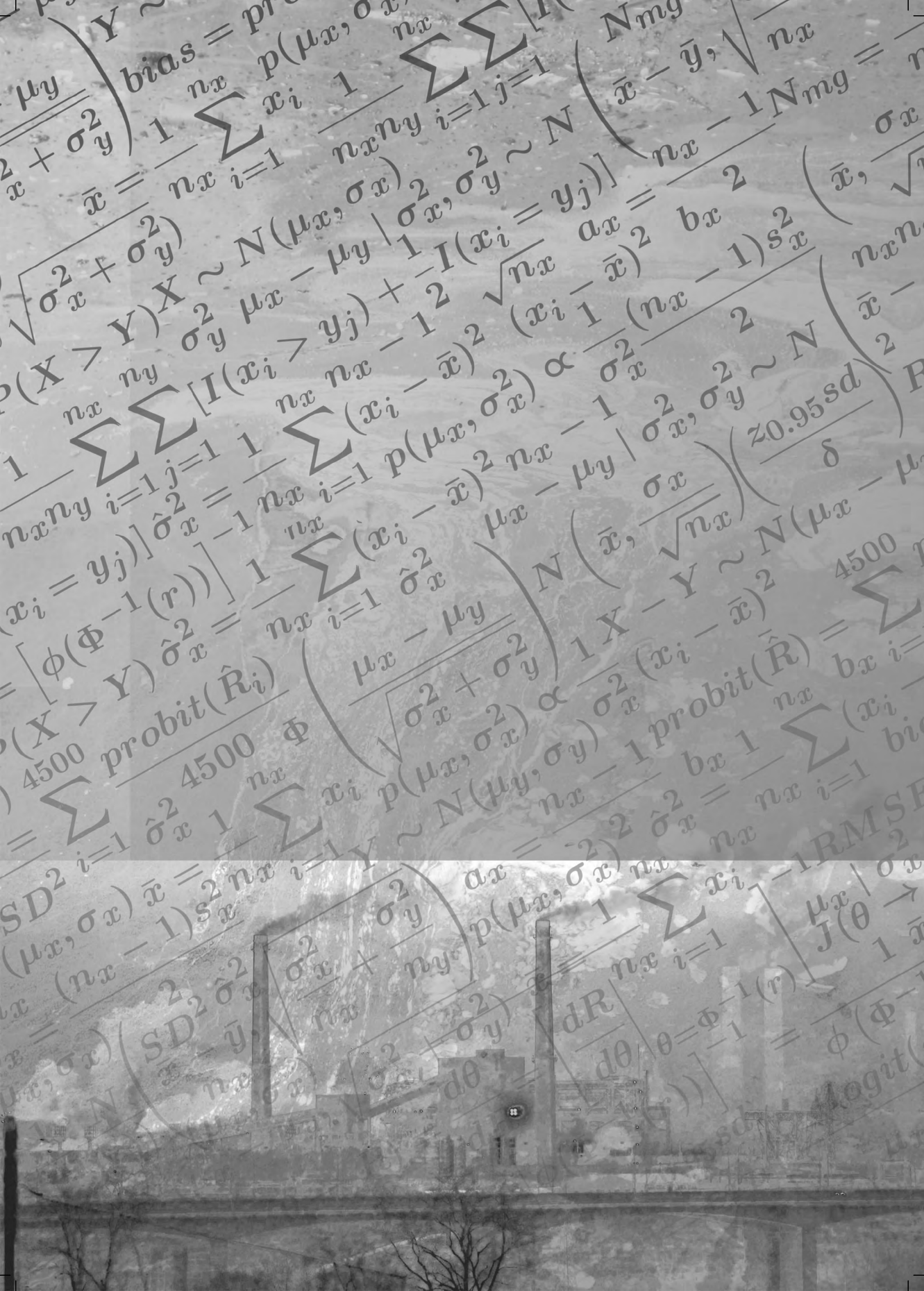
$$f_{R \mid \sigma_x^2, \sigma_y^2}(r \mid \sigma_x^2, \sigma_y^2) = \left[\frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2} \right]^{-\frac{1}{2}} \exp \left[-\frac{\left(\Phi^{-1}(r) - \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right)^2}{2 \frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2}} + \frac{(\Phi^{-1}(r))^2}{2} \right]. \quad (4.A.3)$$

To obtain the marginal posterior density, $f_R(r)$, we integrate σ_x^2 and σ_y^2 out of the joint pdf, $f(r, \sigma_x^2, \sigma_y^2)$ and obtain the required result.

$$\begin{aligned} f_R(r) &= \int_0^\infty \int_0^\infty f(r, \sigma_x^2, \sigma_y^2) d\sigma_x^2 d\sigma_y^2 \\ &= \int_0^\infty \int_0^\infty f(r \mid \sigma_x^2, \sigma_y^2) f(\sigma_x^2) f(\sigma_y^2) d\sigma_x^2 d\sigma_y^2 \\ &= \int_0^\infty \int_0^\infty \left[\frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2} \right]^{-\frac{1}{2}} \exp \left[-\frac{\left(\Phi^{-1}(r) - \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right)^2}{2 \frac{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}{\sigma_x^2 + \sigma_y^2}} + \frac{(\Phi^{-1}(r))^2}{2} \right] \\ &\quad \times \frac{b_x^{a_x}}{\Gamma(a_x)} (\sigma_x^2)^{-a_x-1} \exp \left[-\frac{b_x}{\sigma_x^2} \right] \frac{b_y^{a_y}}{\Gamma(a_y)} (\sigma_y^2)^{-a_y-1} \exp \left[-\frac{b_y}{\sigma_y^2} \right] d\sigma_x^2 d\sigma_y^2 \end{aligned} \quad (4.A.4)$$

with

$$a_x = \frac{n_x - 1}{2}, \quad a_y = \frac{n_y - 1}{2}, \quad b_x = \frac{(n_x - 1)s_x^2}{2} \quad \text{and} \quad b_y = \frac{(n_y - 1)s_y^2}{2}.$$



5

NORMAL-GPD MIXTURES FOR MODELLING EXPOSURE AND EFFECT CONCENTRATIONS IN ENVIRONMENTAL RISK ASSESSMENT OF NANOPARTICLES

Rianne Jacobs, Hilko van der Voet, Cajo J.F. ter Braak

Based on: Jacobs R, van der Voet H, ter Braak CJF (Submitted) Normal-GPD mixtures for modelling exposure and effect concentrations in environmental risk assessment of nanoparticles.

ABSTRACT

IN estimating the risk in a risk assessment, we are interested in the tails of the exposure (X) and effect (Y) distributions. In the case of risk assessment of nanoparticles, we are often confronted with small sample sizes. In this situation, empirical estimation of the risk fails due to the lack of data points in the tails of the exposure and/or effect distributions. Although the normal distribution is customarily the first choice when moving from empirical to parametric estimation, its tails are often found to be too thin. In this chapter, we allow for thicker tails by using the generalised Pareto distribution to estimate the tails of the exposure and effect distributions. We develop a mixture model to estimate the risk, $P(X > Y)$, with the assumption of a normal distribution for the bulk data and generalised Pareto distributions for the tails of X and Y . A sensitivity analysis showed significant influence of the tail thickness on the risk value, especially for low risks. We also studied the effect of small sample sizes on the estimation of the tail index and illustrate the proposed methods on a real data set.

5.1. INTRODUCTION

RISK assessment of engineered nanoparticles (ENPs) is important for the societal acceptance and safe use of ENPs. In the EU, the European Chemical Agency (ECHA) regulates environmental risk assessment and follows a tiered risk assessment approach (ECHA, 2012c). Probabilistic risk assessment is in level 3 of this tiered approach.

In probabilistic risk assessment, the variability in environmental exposure concentrations across environments is modelled by an exposure concentration (X) distribution (ECD). The variability in critical effect concentrations across species is modelled by a species sensitivity distribution (SSD) or critical effect concentration (Y) distribution. The risk, defined as the probability $R = P(X > Y)$, is the area under the curve obtained by multiplying the probability density function (pdf) of the ECD with the cumulative distribution function (cdf) of the SSD. This measure of risk is not unique to the area of nano risk assessment. It is especially well-known in the area of engineering where $R = P(X > Y)$ is known as the reliability in stress-strength models. In the statistics literature, it has been extensively studied for a wide range of distributions having become a standard part of many distributions theory papers (Bekker et al., 2011; Jacobs et al., 2012; Kundu and Gupta, 2006; Mokhlis, 2005; Nadar et al., 2014).

In the context of risk assessment, $R = P(X > Y)$ is a measure of the amount of overlap of the exposure and critical effect distributions. The probability of an exposure concentration being larger than a critical effect concentration decreases as there is less overlap between the two distributions. In the case of small sample sizes, this may lead to problems in the estimation of the risk.

Jacobs et al. (2015a) discuss this problem motivated by a case study of aquatic risk assessment of nanoAg. Figure 5.1 shows the histograms of the exposure and critical effect concentration data (Gottschalk et al., 2013a) as used in Jacobs et al. (2015a). We see that the histograms do not overlap, resulting in an empirical risk estimate of 0. Jacobs et al. (2015a) found that in such a case, parametric modelling is necessary. The authors considered the normal-normal model in which normal distributions are assumed for the log exposure and critical effect concentrations. This is a well-known practice in risk assessment (Aldenberg et al., 2002; Verdonck et al., 2003; Wagner and Løkke, 1991).

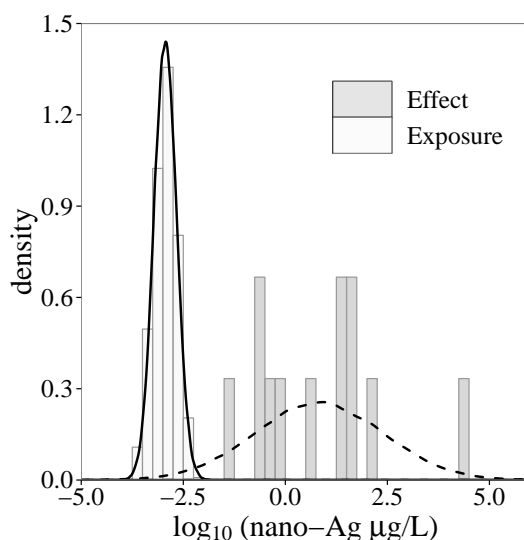


Figure 5.1: Histogram and normal probability density functions of effect (dashed line) and exposure (solid line) concentrations (Jacobs et al., 2015a).

The normal-normal model performs well for small sample sizes and low risk values if the data is in fact normally distributed (Jacobs et al., 2015a). The normal distribution, however, is light-tailed (Carreau and Bengio, 2008). Light tails can cause underestimation of the risk. Therefore, although the normal-normal model improves on empirical estimation, it can still be too optimistic in giving a lower risk than warranted. In addition, exposure and critical effect distributions that lie far apart result in low risk values. It is then the tails of the distributions which become extremely important in the estimation of the risk.

Extreme value theory (Coles, 2001) provides a mathematical basis for the estimation of the tails of a distribution. The tail of a distribution consists of the exceedances of a random variable above a threshold. These exceedances can, for many distributions, be modelled by a generalised Pareto distribution (gpd). This is due to Pickands theorem which states that, for many distributions, the gpd is the limiting distribution of the upper tail of the underlying distribution as the threshold tends to the upper endpoint of the underlying distribution (Pickands III, 1975).

The tail thickness of a distribution is characterised by the tail index, denoted ξ (Carreau

and Bengio, 2008). When $\xi > 0$, the distribution has a thick tail (e.g. Pareto and Student's t -distributions). When $\xi = 0$, the distribution has a light tail (e.g. normal, exponential and log-normal distributions). When $\xi < 0$, the distribution has a finite tail (e.g. beta and uniform distributions) (Carreau and Bengio, 2008). The tail index is a parameter that can be estimated by fitting a gpd to data. In contrast to a single distribution such as the normal or the Student's t -distribution which is just one family of distributions to accommodate varying tail thicknesses, the gpd is a limiting distribution for many different families of distributions. To allow more flexibility in tail thickness, we study the use of the gpd in estimating the tails of the ECD and the SSD.

In this chapter, the gpd is incorporated in a normal-gpd mixture distribution for both the ECD and the SSD. For small R values, where the distributions lie relatively far from one another, the neighbouring (inside) tails are especially important. We, therefore, assume a gpd for the upper tail of the ECD and a gpd for the lower tail of the SSD. The rest of the data is assumed to be normally distributed. With the gpd tails, we can control the tail thicknesses.

In Section 5.2, we provide background on the gpd, introduce the normal-gpd mixture distribution for both the upper and lower tail situation, derive expressions for $R = P(X > Y)$ and explain the implementation of the normal-gpd mixture using the `evmix` package v2.6 (Scarrott and Hu, 2015) in R Software. In Section 5.3, we provide and discuss the simulation results on the sensitivity of $R = P(X > Y)$ towards the tail thickness and the effect of small sample size on the estimation of the tail index, ξ . In Section 5.4, we fit the normal-gpd mixture to the case study data and discuss the results in the context of the simulation results. We close with a discussion on the usage of the gpd and its estimation in practice.

5.2. METHOD

IN this section, we develop a method to estimate the risk, $R = P(X > Y)$, giving special attention to the tails of the distributions.

5.2.1. THE GENERALISED PARETO DISTRIBUTION

The gpd is a threshold model. It arises when one wants to model threshold exceedances, i.e. values above a threshold. The gpd, denoted $GPD^+(\xi, \psi, c)$, is the limiting distribution of excesses over high thresholds (Coles, 2001, p. 75). The + sign indicates that this distribution models the right tail. Its pdf and cdf (Coles, 2001, pp. 75-76) are given by

$$f(x) = \begin{cases} \frac{1}{\psi} \left[1 + \xi \left(\frac{x-c}{\psi} \right) \right]^{-(1/\xi+1)} & \text{if } \xi \neq 0 \\ \frac{1}{\psi} \exp\left(-\frac{x-c}{\psi}\right) & \text{if } \xi = 0 \end{cases} \quad (5.1)$$

and

$$F(x) = \begin{cases} 1 - \left[1 + \xi \left(\frac{x-c}{\psi} \right) \right]^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x-c}{\psi}\right) & \text{if } \xi = 0 \end{cases} \quad (5.2)$$

where ψ is the scale parameter, ξ is the shape parameter or tail index, $x \geq c$ when $\xi \geq 0$, and $c \leq x \leq c - \psi/\xi$ when $\xi < 0$. Although the standard gpd models the right tail, we easily obtain the gpd, denoted $GPD^-(\xi, \psi, c)$, for the left tail by replacing $x - c$ with $c - x$ in the pdf. The pdf and cdf (see derivations in Appendix 5.A) are given by

$$f(x) = \begin{cases} \frac{1}{\psi} \left[1 - \xi \left(\frac{x-c}{\psi} \right) \right]^{-(1/\xi+1)} & \text{if } \xi \neq 0 \\ \frac{1}{\psi} \exp\left(\frac{x-c}{\psi}\right) & \text{if } \xi = 0 \end{cases} \quad (5.3)$$

and

$$F(x) = \begin{cases} \left[1 - \xi \left(\frac{x-c}{\psi} \right) \right]^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp\left(\frac{x-c}{\psi}\right) & \text{if } \xi = 0 \end{cases} \quad (5.4)$$

where ψ is the scale parameter, ξ is the shape parameter or tail index, $x \leq c$ when $\xi \geq 0$, and $c + \psi/\xi \leq x \leq c$ when $\xi < 0$. Note that for $\xi = 0$, the gpd reduces to the exponential distribution.

DERIVATION OF $R = P(X > Y)$

In this section, we derive expressions for the calculation of $R = P(X > Y)$ assuming that both X and Y have a gpd. Rezaei et al. (2010) considered the estimation of $R = P(X > Y)$ when $X \sim GPD^+(\xi_x, \psi_x, c_x)$ and $Y \sim GPD^+(\xi_y, \psi_y, c_y)$, the probability density functions

for X and Y of which are illustrated in Figure 5.2A. In the context of risk assessment, however, we are interested in how the right tail of the exposure distribution overlaps with the left tail of the critical effect distribution. This situation is illustrated in Figure 5.2B.

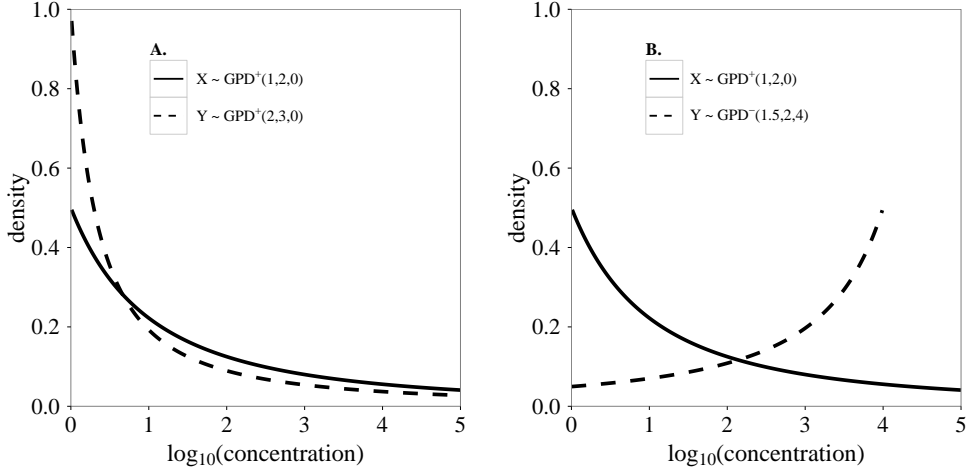


Figure 5.2: Probability density functions for two right tails (A) and one right and one left tail (B).

Let X represent the right tail of a distribution with left endpoint, c_x . Then $X \sim \text{GPD}^+(\phi_x, \xi_x, c_x)$ with pdf and cdf is given by Equations 5.1 and 5.2, respectively. Also, let Y represent the left tail of a distribution with right endpoint, c_y . Then $Y \sim \text{GPD}^-(\phi_y, \xi_y, c_y)$ with pdf and cdf is given by Equations 5.3 and 5.4, respectively.

Assuming $c_x < c_y$, the risk, $R = P(X > Y)$, for $\xi_x \neq 0$ and $\xi_y \neq 0$ is then given by

$$\begin{aligned}
 R &= \int_{-\infty}^{\infty} P(X = x)P(Y < x | X = x)dx \\
 &= \frac{1}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &\quad + \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x}.
 \end{aligned} \tag{5.5}$$

The risk, $R = P(X > Y)$, for $\xi_x = 0$ and $\xi_y = 0$, is given by

$$R = \frac{\psi_x}{\psi_x - \psi_y} \exp\left(\frac{c_x - c_y}{\psi_x}\right) - \frac{\psi_y}{\psi_x - \psi_y} \exp\left(\frac{c_x - c_y}{\psi_y}\right). \quad (5.6)$$

The reader is referred to Appendix 5.A for a detailed derivation of Equations 5.5 and 5.6. For the case where $c_x < c_y$, all X values are greater than all Y values resulting in $R = 1$ (for formal proof see Appendix 5.A).

The parameters of the gpd (i.e. ξ and ψ) can be estimated using the `fgpd` function from the `evmix` package v2.6 (Scarrott and Hu, 2015) in R Software. This function provides maximum likelihood (ml) estimators and standard errors of the parameters. The threshold, c , can (i) also be estimated or (ii) given as a fixed value. To implement the GPD^- distribution, the quantiles and the threshold, c , need to be negated.

5

5.2.2. THE NORMAL-GPD MIXTURE

We now consider all the data and not only the tails. In this case, the right tail of the exposure data and the left tail of the critical effect data are modelled by a gpd and the bulk of the data is modelled by either a left or right truncated normal distribution. The truncated normal distribution, denoted $trN_{c_\ell}^{c_r}(\mu, \sigma^2)$, is given by

$$f(x) = \frac{(1/\sigma)\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{c_r-\mu}{\sigma}\right) - \Phi\left(\frac{c_\ell-\mu}{\sigma}\right)} \quad \text{and} \quad F(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{c_\ell-\mu}{\sigma}\right)}{\Phi\left(\frac{c_r-\mu}{\sigma}\right) - \Phi\left(\frac{c_\ell-\mu}{\sigma}\right)} \quad (5.7)$$

where $c_\ell \leq x \leq c_r$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively (Johnson et al., 1994). For left truncation ($c_r = \infty$), we have $\Phi\left(\frac{c_r-\mu}{\sigma}\right) = 1$ and for right truncation ($c_\ell = -\infty$), we have $\Phi\left(\frac{c_\ell-\mu}{\sigma}\right) = 0$.

Without loss of generality, we assume that the exposure data is located at zero and scaled to one. The exposure data (x) is then modelled by a $trN_{-\infty}^{c_x} - GPD^+(\psi_x, \xi_x, c_x)$ mixture and the effect data (y) by a $trN_{c_y}^{\infty}(\mu, \sigma^2) - GPD^-(\psi_y, \xi_y, c_y)$ mixture. The pdfs and cdfs

are given by

$$f(x) = \begin{cases} (1 - p_x) \text{tr} N_{-\infty}^{c_x} & \text{if } x \leq c_x \\ p_x \text{GPD}^+(\psi_x, \xi_x, c_x) & \text{if } c_x < x \end{cases} \quad (5.8)$$

$$F(x) = \begin{cases} (1 - p_x) \text{tr} N_{-\infty}^{c_x} & \text{if } x \leq c_x \\ 1 - p_x + p_x \text{GPD}^+(\psi_x, \xi_x, c_x) & \text{if } c_x < x \end{cases} \quad (5.9)$$

$$f(y) = \begin{cases} p_y \text{GPD}^-(\psi_y, \xi_y, c_y) & \text{if } y < c_y \\ (1 - p_y) \text{tr} N_{c_y}^{\infty}(\mu, \sigma^2) & \text{if } c_y \leq y \end{cases} \quad (5.10)$$

$$F(y) = \begin{cases} p_y \text{GPD}^-(\psi_y, \xi_y, c_y) & \text{if } y < c_y \\ p_y + (1 - p_y) \text{tr} N_{c_y}^{\infty}(\mu, \sigma^2) & \text{if } c_y \leq y \end{cases} \quad (5.11)$$

where p_x (p_y) is the percentage of data in the upper (lower) tail to the right (left) of c_x (c_y). The tail probabilities, p_x and p_y , can be calculated from c_x and c_y using the bulk distributions,

$$p_x = 1 - \Phi(c_x) \text{ and } p_y = \Phi\left(\frac{c_y - \mu}{\sigma}\right).$$

When considering the normal-gpd mixture distribution, there is the problem of discontinuity of the pdf at the threshold, i.e. the pdf of the gpd at the threshold will not necessarily equal the pdf of the truncated normal at the threshold. This problem can be solved by imposing a continuity constraint. This continuity constraint equates the pdf of the gpd and the truncated normal at the threshold and can be enforced by solving for the scale parameter ψ :

$$\psi_x = \frac{1 - \Phi(c_x)}{\phi(c_x)} \text{ and } \psi_y = \frac{\Phi\left(\frac{c_y - \mu}{\sigma}\right)}{\phi\left(\frac{c_y - \mu}{\sigma}\right)}.$$

DERIVATION OF $R = P(X > Y)$

When deriving an expression for the risk, R , we assume that $c_x < c_y$. Moreover, we assume that $\xi_x \neq 0$ and $\xi_y \neq 0$. Using the definition of $R = P(X > Y)$ and Equations 5.8-5.11,

we obtain R as

$$\begin{aligned}
 R = & \frac{(1-p_x)p_y}{\sqrt{2\pi}\Phi(c_x)} \int_{-\infty}^{c_x} e^{-x^2/2} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 & + \frac{p_x p_y}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 & + \frac{p_x(1-p_y)}{\left(1 - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)} \Phi \left(\frac{c_y-\mu}{\sigma} \right) \left(1 + \xi_x \left(\frac{c_y-c_x}{\psi_x} \right) \right)^{-1/\xi_x} \\
 & + \frac{\xi_x^{1/\xi_x} p_x(1-p_y)}{\psi_x^{1/\xi_x+1} \sigma \sqrt{2\pi} \left(1 - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)} \int_{c_y}^{\infty} e^{-(\frac{x-\mu}{\sigma})^2/2} \left(\frac{\psi_x}{\xi_x} - c_x + x \right)^{-1/\xi_x} dx \\
 & + \frac{p_x \left(p_y - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)}{\left(1 - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)} \left[1 + \xi_x \left(\frac{c_y-c_x}{\psi_x} \right) \right]^{-1/\xi_x}. \tag{5.12}
 \end{aligned}$$

5

For a detailed derivation of Equation 5.12, the reader is referred to Appendix 5.A, where we also derive an expression for R for the case where both the upper and lower tails of the exposure and effect distributions are gpd.

The parameters of the normal-gpd mixture distribution (i.e. ξ , μ and σ) can be estimated using the `fnormgdpcon` function from the `evmix` package v2.6 (Scarrott and Hu, 2015) in R Software. This function provides `ml` estimators and standard errors of the parameters. The threshold, c , can (i) also be estimated or (ii) given as a fixed value. To implement the GPD^- distribution, the quantiles, the threshold, c , and the mean, μ , of the truncated normal need to be negated.

5.2.3. SENSITIVITY OF THE RISK TO THE TAIL INDICES

We investigated the sensitivity of R to the tail indices, ξ_x and ξ_y , by considering three scenarios.

In the three scenarios the reference exposure and the reference critical effect distributions were normal distributions, chosen such that the risk, R , is equal to 10^{-10} , 10^{-5} and 10^{-2} . Motivated by the case study in which the critical effect concentration data has a bigger variance than that of the exposure concentration data, we set the variance of the

critical effect concentration at 4 times that of the exposure concentration distribution. By setting the exposure concentration distribution to standard normal, the critical effect distributions for the three cases are then $N(14.22, 4)$, $N(9.54, 4)$ and $N(5.2, 4)$. These three scenarios are illustrated in the first row of Figure 5.3 with the corresponding R values indicated by a horizontal black line on the graphs in the second row.

We then changed the normal distributions to normal-gpd distributions by placing the exposure threshold, c_x , at the 95th percentile of the exposure distribution and the critical effect threshold, c_y , at the 5th percentile of the critical effect distribution (indicated by the vertical black and red lines in the first row of Figure 5.3). The resulting tails were then assumed to have a gpd distribution with the shape parameters, ξ_x and ξ_y , taken on a grid from -0.2 to 0.5. The value of R was calculated for the resulting normal-gpd mixture distributions. These were plotted as a function of the exposure shape parameter, ξ_x with a colour gradient indicating the value of critical effect shape parameter, ξ_y .

For each of the three scenarios, we also calculated R assuming the exposure and critical effect distributions to be normally distributed with the same expected value and variance as the corresponding normal-gpd mixture distributions. Note that the variance of a gpd is infinite for $\xi \geq 0.5$ (Johnson et al., 1994). That is the reason that the grid of shape parameter values only runs up to 0.5. The R values are plotted in Figure 5.3 as dotted lines. Furthermore, we also plotted the ratio $R_{mixture}/R_{normal}$ in the third row of Figure 5.3 to compare the two values.

5.2.4. ESTIMATION OF THRESHOLD AND TAIL INDEX

The reason for using the gpd to model the tails as opposed to using a normal distribution for all the data, is to allow for thicker tails. For the gpd, this is obtained when the shape parameter, ξ , is non-negative. We are, therefore, only interested in a zero or positive shape parameter. Moreover, in practice it is unknown which threshold provides the best fit to the data. We performed a simulation study to investigate the performance of the ml estimator of the threshold, c , and shape parameter, ξ , for various sample sizes. We considered three values of the shape parameter, namely $\xi = 0$, $\xi = 0.15$ and $\xi = 0.3$. For each value, we considered three sample sizes, namely 12, 100 and 1000. We set the range of values of the threshold in relation to the sample size by choosing a grid of n_{tail} values

from 2 up to 50% of the sample size, where n_{tail} is the expected number of data points in the tail, and calculating the corresponding threshold values. For each ξ , sample size and n_{tail} combination, we simulated 1000 samples from a normal-gpd mixture distribution where the bulk data is modelled by a standard normal distribution and the right tail by a gpd. For each sample, we obtained the ml estimators of the threshold, c , and the shape parameter, ξ .

5.3. SIMULATION RESULTS

5.3.1. SENSITIVITY TO TAIL INDICES

FIGURE 5.3 summarises the analysis of the sensitivity of R to the tail indices, ξ_x and ξ_y . The columns represent the three scenarios illustrated in the first row, where from left to right the R values are equal to 10^{-10} , 10^{-5} and 10^{-2} . The vertical black and red lines indicate the thresholds which are set at the 95th and 5th percentiles for the exposure and critical effect distributions, respectively.

In the second row, we plot $R = P(X > Y)$ as a function of the exposure tail index, ξ_x , with the colour gradient indicating the critical effect tail index, ξ_y . The solid lines indicate the case where the exposure and critical effect distributions are normal-gpd mixture distributions and the dotted lines indicate the case where the exposure and critical effect distributions are normal distributions with expected value and variance equal to that of the corresponding normal-gpd mixtures. The horizontal black line indicates the R value of the normal-normal scenario illustrated in the first row. This line indicates the minimum value R can take assuming tails that are equal or thicker than a normal distribution.

We notice that as the exposure and effect distributions move closer together (moving from the first column to the third column), the tail indices have less influence on the value of R , i.e. the range of possible R values becomes smaller. This is due to the fact that as the exposure and effect distributions move closer together, the tails have less influence on determining the value of R . In the third column the normal distributions in the reference case already overlap considerably. Creating heavier tails does not contribute much extra in increasing the value of R . In the first column, however, the normal distributions in the reference case hardly overlap. Even the slightest change in tail

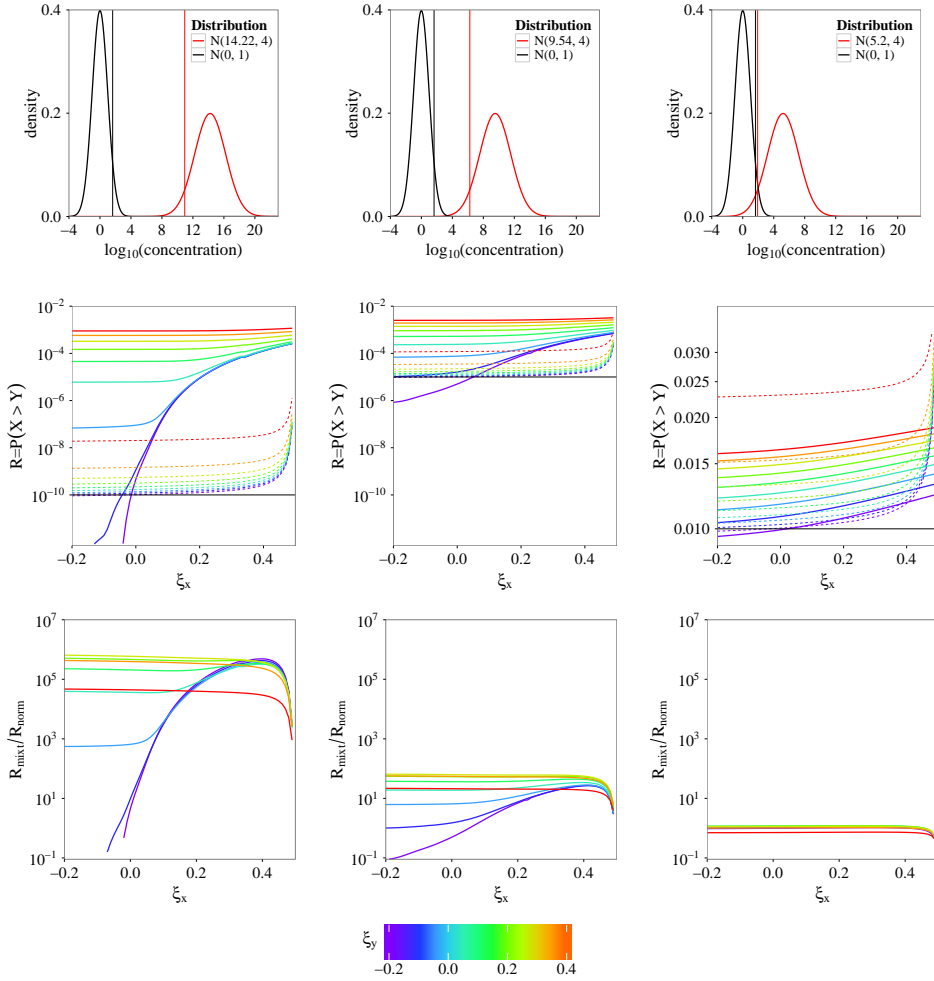


Figure 5.3: Plot of $R = P(X > Y)$ as a function of the tail indices ξ_x and ξ_y . The three columns represent the three scenarios illustrated in the first row. In the second row, R is plotted for the normal-gpd mixture (solid lines) and normal distribution (dashed line). The horizontal black line indicates the R value corresponding to three scenarios in the first row. In the third row the ratio $R_{mixture}/R_{normal}$ is plotted.

thickness, therefore, greatly contributes to the value of R . We also note that when the tail indices become negative, R drops to values below those of the normal distributions. This is due to the fact that for negative ξ , the gpd has lighter tails than a normal distri-

bution and a finite domain. This greatly reduces the overlap of the exposure and critical effect distributions. Finally, we note the extremely large range the R values can take by keeping ξ_x fixed and only changing ξ_y (the difference between the blue/purple and red lines, especially for small risk values).

In the third row, we plot the ratio, $R_{mixture}/R_{normal}$, as a function of the exposure tail index, ξ_x , with the colour gradient indicating the critical effect tail index, ξ_y . Modelling the exposure and critical effect distributions with normal distributions can severely underestimate the risk compared to when the exposure and effect distributions are modelled with normal-gpd mixtures with the same means and variances. We see that the $R_{mixture}/R_{normal}$ ratio runs up to 10^6 . The normal distributions can, therefore, underestimate the risk with a factor of up to 10^6 . This difference becomes less as the distributions move closer together.

5

5.3.2. ESTIMATION OF THRESHOLD AND TAIL INDEX

Figure 5.4 summarises the results of the simulation study on the performance of the `ml` estimator of the tail index, ξ , for various sample sizes. The first column shows the results obtained from (randomly chosen) single data sets, whereas the second column summarises the sampling distribution of $\hat{\xi}$ as estimated from 1000 data sets.

The top-left graph plots $\hat{\xi}$ in single data sets as a function of the expected number of data points in the tail, n_{tail} . The dotted lines indicate a 95% confidence interval calculated using the obtained standard errors from the `evmix` package ($\hat{\xi} \pm 1.96SE(\hat{\xi})$). The horizontal grey line indicates the true $\xi = 0.3$. The missing points in, for example, the case $n = 100$ are due to non-convergence of the `ml` optimisation. The bottom graph plots the $SE(\hat{\xi})$ as a function of n_{tail} . The coloured lines indicate the three sample sizes.

The second column illustrates the sampling distribution of $\hat{\xi}$ as a function of n_{tail} . The top-right graph plots the mean, $\bar{\hat{\xi}}$, with the 2.5th and 97.5th percentiles of 1000 iterations. The iterations in which there was non-convergence of the `ml` optimisation were removed. The bottom graph plots the standard deviation of the sampling distribution of $\hat{\xi}$ for each sample size.

When comparing the left and right graphs, we note that the standard error is not very

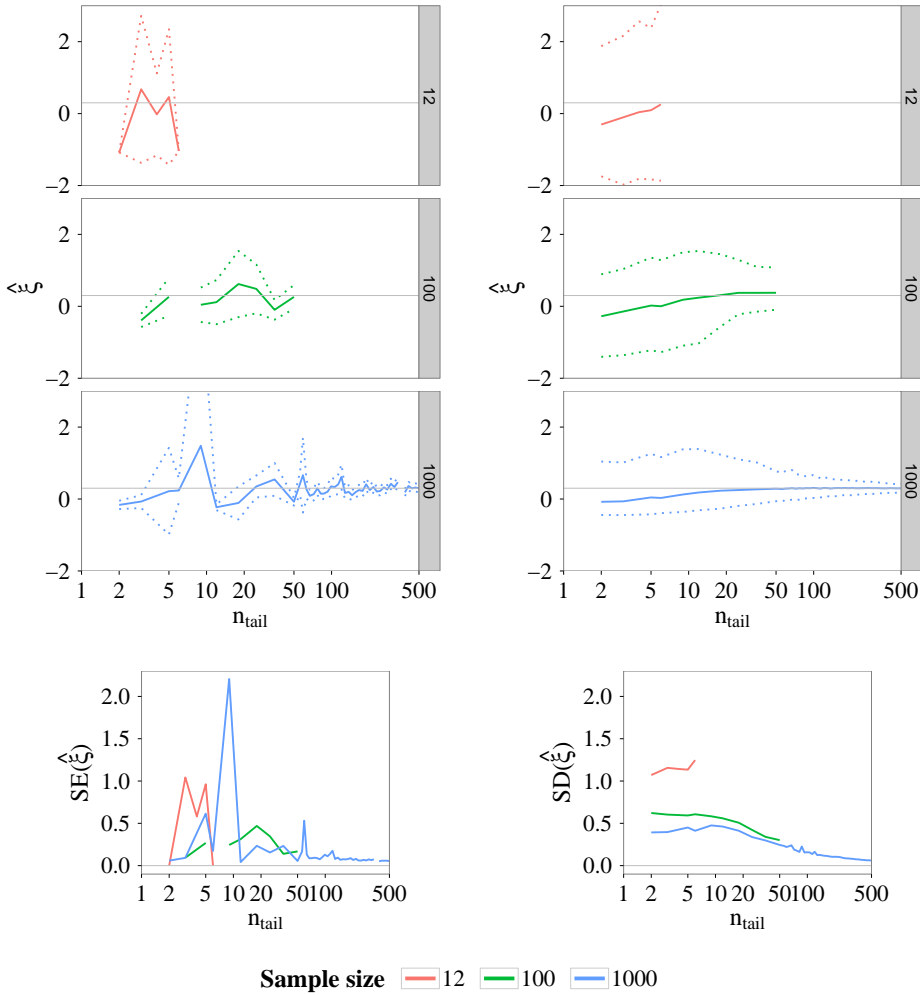


Figure 5.4: Plot of $\hat{\xi}$ and $SE(\hat{\xi})$ as a function of n_{tail} for a single data set (first column) and for 1000 iterations (second column). The dotted lines indicate an approximate 95% confidence interval for $\hat{\xi}$. The three colours indicate the three sample sizes 12, 100 and 1000. The horizontal grey lines in the top two graphs indicate the true value of $\xi = 0.3$ used in the simulations.

reliable when the number of data points in the tail is small. Moreover, we note that ξ is consistently underestimated for smaller number of points in the tail. This bias is greatest in the smallest sample size (the top right graph). The bias could be explained by the fact

that with little data in the tails, there is no evidence for a thick tail and the tail index is underestimated. For larger sample size, the bias is small compared to the range of the 95% interval and the standard deviation. As seen in the bottom right graph, the standard deviations of the sampling distribution of $\hat{\xi}$ reduce for larger sample sizes as is expected.

The graphs for $\xi = 0.15$ (Figure 5.B.1) and $\xi = 0$ (Figure 5.B.2) are given in Appendix 5.B. Here we see similar results as for the $\xi = 0.3$. One thing to note for these lower values of ξ is that for a sample size of 100, ξ seems to be slightly overestimated for larger n_{tail} .

Figure 5.5 summarises the results of the simulation study on the performance of the `ml` estimator of the threshold, c , for various sample sizes. The graphs plot \hat{n}_{tail} instead of the threshold estimate. The left graph shows the results obtained from (randomly chosen) single data sets, whereas the right graph summarises the sampling distribution of \hat{n}_{tail} as estimated from 1000 data sets.

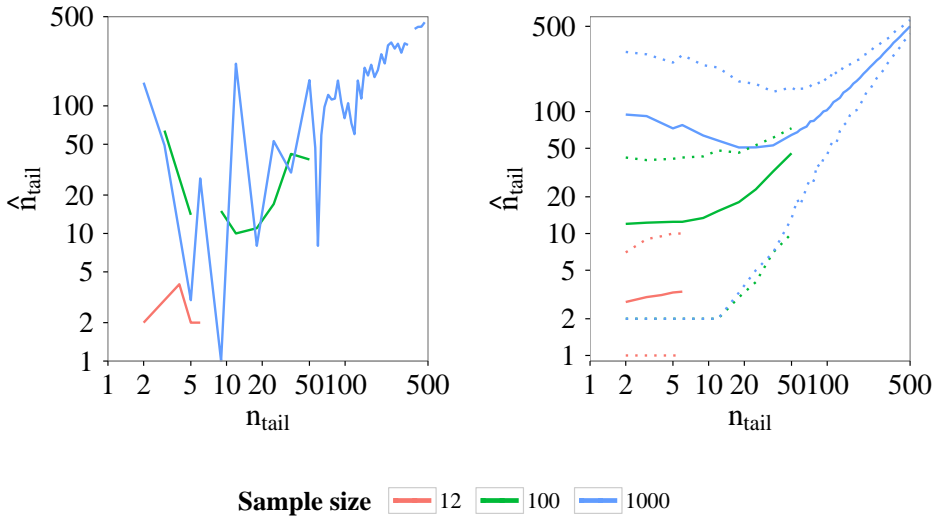


Figure 5.5: Plot of \hat{n}_{tail} as a function of n_{tail} for a single data set (left graph) and for 1000 iterations (right graph). The dotted lines indicate an approximate 95% confidence interval for \hat{n}_{tail} . The three colours indicate the three sample sizes 12, 100 and 1000.

The left graph plots \hat{n}_{tail} in single data sets as a function of the number of data points in the tail, n_{tail} . The missing points in, for example, the case $n = 100$ and $n = 1000$ are

due to non-convergence of the `ml` optimisation. The coloured lines indicate the three sample sizes.

The right graph illustrates the sampling distribution of \hat{n}_{tail} as a function of n_{tail} . We plot the mean, $\bar{\hat{n}}_{tail}$, with the 2.5th and 97.5th percentiles of 1000 iterations. The iterations in which there was non-convergence of the `ml` optimisation were removed.

From the left graph, we see that the estimation of the threshold becomes problematic for n_{tail} values less than 50 for $n = 1000$ and less than 10 for $n = 100$. The right graph confirms this by the extremely wide confidence bands for smaller n_{tail} . As the number of data points in the tail becomes less, the `ml` estimation forces a threshold which results in more data points in the tail. This is most apparent for the large sample size ($n = 1000$), where there are enough data points available to force into the tail. For the smaller sample sizes, especially $n = 12$, there are less data points to play with, and the `ml` estimation procedure is forced to have a small number of data points in the tail.

5.4. CASE STUDY

IN this section, we evaluate the case study results on the basis of the simulation study results. Table 5.1 provides the `ml` estimators for the shape parameter, threshold and corresponding tail probability for the exposure and critical effect data. Note that both data sets have been scaled with the mean and standard deviation of the exposure data.

Table 5.1: Maximum likelihood estimators and standard errors of shape parameter and threshold for exposure and effect data

	$\hat{\xi}$ (SE)	\hat{c} (SE)	\hat{p}
Exposure data	-0.3156 (0.0406)	1.1125 (0.4612)	0.1360
Effect data	-1.3133 (*)	8.6209 (*)	0.2077

* observed information matrix is singular

The exposure sample size is 1000. The estimated tail fraction is 0.136, resulting in 136 data points in the tail. The tail fraction estimate is questionable because it is based on the threshold estimate which has a large standard error (Table 5.1). We, nevertheless, have some idea as to the number of data points in the tail. From the simulation study (Figure

5.4), we have learned that the estimation of the shape parameter is quite good when the sample size is 1000 and $n_{tail} = 136$. The estimated shape parameter of the exposure data is, therefore, reliable and we deduce that it is indeed negative. This is also evident from the 95% confidence interval $\hat{\xi}_x \pm 1.96SE(\xi_x) = (-0.3952, -0.2360)$ which does not contain 0. The data does not provide any evidence for a thick tail. For the exposure data, there is, therefore, no reason to use the normal-gpd mixture instead of the normal distribution.

The estimates of the effect data are not reliable at all. First, the observed information matrix was singular which makes us doubt whether the likelihood optimisation converged to the global maximum. Second, we learn from the simulation results (red lines in Figure 5.4) that the shape parameter is consistently underestimated with very large standard errors. This is expected as a sample size of 12 is basically too small to do any decent parameter estimation.

5

Estimating the threshold from such a small effect data set is actually not possible. We, therefore, also considered the estimation of the tail index of the effect data for some fixed threshold values. The fixed threshold values were chosen to lie halfway between the data points resulting in 2, 3, 4, and 5 data points in the tail. Table 5.2 provides the chosen threshold values, the resulting number of data points in the tail and the tail index estimates. Even for a fixed threshold, the likelihood optimisation had trouble converging. In all four cases, the observed information matrix was singular. Moreover, the tail index estimates are still negative.

Table 5.2: Maximum likelihood estimators and standard errors of the tail index of the critical effect concentration data for fixed thresholds where n_{tail} refers to the number of observations in the data that are smaller than the threshold.

threshold	n_{tail}	$\hat{\xi}$	$SE(\hat{\xi})$
8.5991	2	-1.3188	*
8.9393	3	-1.1306	*
9.8223	4	-1.0283	*
11.8329	5	-1.0119	*

* observed information matrix is singular

Using the obtained parameter estimates of the exposure and effect distributions from Table 5.1 to estimate R , we obtain a value of 0. This is due to the negative shape parameters resulting in very light and finite tails. As mentioned in the Introduction, the

motivation for using the `gpd` is to model tails that are heavier than the normal. Our data does not provide any evidence for heavier tails. The estimate for R assuming normal distributions ranges from 0.0068-0.0108 (Table 2 in Jacobs et al. (2015a)). These correspond to the third column in Figure 5.3 of our simulation study. We note that for negative tail indices, the normal distribution gives a higher estimate of the risk than the mixture distribution.

5.5. DISCUSSION AND CONCLUSION

THE `gpd` is a general distribution incorporating three families of distributions, namely, thick-tailed distributions ($\xi > 0$), light-tailed distributions ($\xi = 0$) and finite-tailed distributions ($\xi < 0$). A negative shape parameter ($\xi < 0$) should, therefore, only be used if a bounded true distribution is considered a realistic option. Apart from the trivial upper concentration bound corresponding to 1 kg/kg, which is a value very outside the range of expected data, there is no compelling theory which leads us to expect a true upper bound on the concentration distribution anywhere in the order of magnitude of the data points. Therefore we consider negative shape parameters as not acceptable.

As mentioned in the Introduction (Section 5.1), the normal distribution is a light-tailed distribution. Modelling the tails by a `gpd` would then allow for thicker tails. Only with sufficient data, however, will the shape parameter of the `gpd` be able to capture a thick tail if it exists. This was shown in our simulation (Figure 5.4) in which we found that one needs at least around 20 data points in the tail of the distribution to obtain a shape parameter that is on average positive, if the true value is as high as $\xi = 0.3$. For a smaller shape parameter of 0.15, the tail sample size has to be around 50 (see Figure 5.B.1 in Appendix 5.B). Kennedy et al. (2011) also found difficulty in estimating the shape parameter. They obtained, in pesticide residue data sets, surprisingly frequent shape parameters that were negative in both `ml` and Bayesian estimation. The Bayesian estimations were less negative because they were estimated in a hierarchical model incorporating many data sets, resulting in larger sample sizes. Our simulation study shows that part of the issue of obtaining negative point estimates for the shape parameters is due to the bias of the estimates.

To better understand the tail behaviour of our case study, we also fitted Student's t -

distributions to both the exposure and the critical effect data. The degrees of freedom of the Student's t -distribution relates to the thickness of the tail with lower degrees giving a heavier tail. Figure 5.6 illustrates what happens when we attempt to fit a t -distribution to the exposure and effect data via `ml` estimation. In Figure 5.6, we plot the profile likelihood as a function of the degrees of freedom. For the exposure data, we note that the profile likelihood is a decreasing function suggesting that the `ml` estimate of the degrees of freedom is infinite. The 95% confidence interval is $[33 - \infty]$. This interval is obtained by adding the 95th percentile of the $\chi^2(1)$ distribution to the infimum of the $-2\log$ likelihood profile on the degrees of freedom (in practice the value at 150 degrees of freedom was used). The value obtained is indicated by the upper horizontal grey line in Figure 5.6. This line cuts the profile at 33 degrees of freedom. This is to be expected as above 20 degrees of freedom, the t -distribution is, for all practical purposes, close to a normal distribution. This confirms our result of a non-positive shape parameter, ξ_x . For the exposure data, there is, therefore, no indication for a thick upper tail.

5

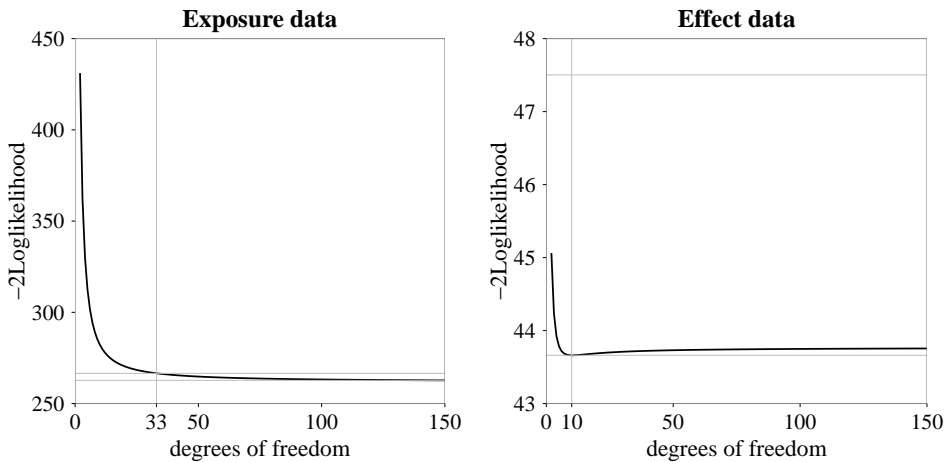


Figure 5.6: Plot of profile likelihood as a function of the degrees of freedom of a t -distribution. The horizontal grey lines indicate the minimum profile likelihood and the minimum profile likelihood + the 95th percentile of a $\chi^2(1)$ distribution. The vertical lines indicate the corresponding degrees of freedom.

For the critical effect data, we have a different situation. In this case, the profile likelihood has an optimum at 10 degrees of freedom, but the confidence interval covers all

possible numbers of degrees of freedom: profile likelihood function nowhere exceeds the critical value (the top horizontal grey line in Figure 5.6). The data, therefore, does not contain any useful information for estimating the number of degrees of freedom. It is, therefore, not possible to determine how thick the tails should be. This corresponds to what we found in the simulation study of the normal-gpd mixture, where we needed at least around 20 data points to estimate a positive shape parameter which implies a thick tail.

For estimation of the gpd parameters we used the *evmix* R package which implements *ml* estimation. Maximum likelihood estimators only exist for $\xi > -1$ and their asymptotic properties only hold for $\xi > -0.5$ (Smith, 1985). Even so, *ml* estimation may fail for small sample sizes (Hosking and Wallis, 1987). This is also evident in our simulation study, where, for small sample sizes, we often encountered non-convergence in the likelihood optimisation. An alternative estimation method is the method of probability-weighted moments or L-moments (Hosking and Wallis, 1997) developed for the generalised extreme-value distribution (Hosking et al., 1985) and the generalised Pareto distribution (Hosking and Wallis, 1987). L-moment estimators outperform *ml* estimators for small sample sizes (Hosking and Wallis, 1987). L-moment estimation was used in a gpd-normal-gpd mixture context by Mendes and Lopes (2004).

In conclusion, the normal-gpd mixture model is a flexible model when it comes to tail estimation. It is, however, not very useful for small sample sizes. A possible solution to the small sample problem could be to estimate a pooled tail index from many data sets as was done by Kennedy et al. (2011). If such data sets are not available, a conservative approach could be followed by choosing a high value for the tail index in the normal-gpd mixture model, resulting in a higher risk value. When conservatism is not required, a less complex model such as the normal-normal model (Jacobs et al., 2015a) might be preferred.

Appendix

5.A. DERIVATIONS

5.A.1. DERIVATION OF PDF AND CDF FOR GDP^- DISTRIBUTION

The pdf of $GDP^-(\xi, \psi, c)$ is trivially obtained by substituting $x-c$ in the pdf of $GDP^+(\xi, \psi, c)$ in Equation 5.1 with $c-x$ as

$$g(x) = \begin{cases} \frac{1}{\psi} \left[1 + \xi \left(\frac{c-x}{\psi} \right) \right]^{-(1/\xi+1)} & \text{if } \xi \neq 0 \\ \frac{1}{\psi} \exp \left(-\frac{c-x}{\psi} \right) & \text{if } \xi = 0 \end{cases}$$

$$= \begin{cases} \frac{1}{\psi} \left[1 - \xi \left(\frac{x-c}{\psi} \right) \right]^{-(1/\xi+1)} & \text{if } \xi \neq 0 \\ \frac{1}{\psi} \exp \left(\frac{x-c}{\psi} \right) & \text{if } \xi = 0 \end{cases}$$

where ψ is the scale parameter, ξ is the shape parameter, $x \leq c$ when $\xi \geq 0$, and $c + \psi/\xi \leq x \leq c$ when $\xi < 0$.

The cdf of $GDP^-(\xi, \psi, c)$ is obtained by integrating the pdf above.

For $\xi \neq 0$:

$$\begin{aligned} G(x) &= \int_{-\infty}^x \frac{1}{\psi} \left[1 - \xi \left(\frac{y-c}{\psi} \right) \right]^{-(1/\xi+1)} dy \\ &= \frac{1}{\psi} \int_{-\infty}^x \left[1 - \xi \left(\frac{y-c}{\psi} \right) \right]^{-(1/\xi+1)} dy \\ &= \frac{1}{\psi} \frac{-\psi}{\xi} \frac{\left[1 - \xi \left(\frac{y-c}{\psi} \right) \right]^{-1/\xi}}{-1/\xi} \Bigg|_{-\infty}^x \\ &= \left[1 - \xi \left(\frac{y-c}{\psi} \right) \right]^{-1/\xi} \Bigg|_{-\infty}^x = \left[1 - \xi \left(\frac{x-c}{\psi} \right) \right]^{-1/\xi} - 0 \\ &= \left[1 - \xi \left(\frac{x-c}{\psi} \right) \right]^{-1/\xi} \end{aligned}$$

For $\xi = 0$:

$$\begin{aligned}
 G(x) &= \int_{-\infty}^x \frac{1}{\psi} \exp\left(\frac{y-c}{\psi}\right) dy \\
 &= \frac{1}{\psi} \int_{-\infty}^x \exp\left(\frac{y-c}{\psi}\right) dy \\
 &= \frac{1}{\psi} \psi \exp\left(\frac{y-c}{\psi}\right) \Big|_{-\infty}^x \\
 &= \exp\left(\frac{x-c}{\psi}\right) - 0 \\
 &= \exp\left(\frac{x-c}{\psi}\right)
 \end{aligned}$$

5.A.2. DERIVATION OF $R = P(X > Y)$ FOR $c_y > c_x$ WHEN X AND Y HAVE GPD DISTRIBUTIONS ($\xi_x \neq 0$ AND $\xi_y \neq 0$)

5

$$\begin{aligned}
 R &= \int_{-\infty}^{\infty} P(X=x)P(Y < x | X=x)dx \\
 &= \int_{-\infty}^{c_x} P(X=x)P(Y < x | X=x)dx + \int_{c_x}^{c_y} P(X=x)P(Y < x | X=x)dx \\
 &\quad + \int_{c_y}^{\infty} P(X=x)P(Y < x | X=x)dx \\
 &= \int_{c_x}^{c_y} \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &\quad + \int_{c_y}^{\infty} \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} dx. \tag{5.A.1}
 \end{aligned}$$

Consider term 2 of Equation 5.A.1.

$$\begin{aligned}
 \int_{c_y}^{\infty} \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} dx &= \frac{1}{\psi_x} \frac{\psi_x}{\xi_x} \frac{\left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-1/\xi_x}}{-1/\xi_x} \Bigg|_{c_y}^{\infty} \\
 &= - \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \Bigg|_{c_y}^{\infty} \\
 &= - \left(0 - \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \right) \\
 &= \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \quad (5.A.2)
 \end{aligned}$$

We now substitute Equation 5.A.2 back into Equation 5.A.1 to obtain

$$\begin{aligned}
 R &= \frac{1}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &\quad + \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x}. \quad (5.A.3)
 \end{aligned}$$

5.A.3. DERIVATION OF $R = P(X > Y)$ FOR $c_y > c_x$ WHEN X AND Y HAVE GPD DISTRIBUTIONS ($\xi_x = 0$ AND $\xi_y = 0$)

$$\begin{aligned}
 R &= \int_{-\infty}^{\infty} P(X = x) P(Y < x | X = x) dx \\
 &= \int_{-\infty}^{c_x} P(X = x) P(Y < x | X = x) dx + \int_{c_x}^{c_y} P(X = x) P(Y < x | X = x) dx \\
 &\quad + \int_{c_y}^{\infty} P(X = x) P(Y < x | X = x) dx \\
 &= \int_{c_x}^{c_y} \frac{1}{\psi_x} \exp\left(-\frac{x - c_x}{\psi_x}\right) \exp\left(\frac{x - c_y}{\psi_y}\right) dx \\
 &\quad + \int_{c_y}^{\infty} \frac{1}{\psi_x} \exp\left(-\frac{x - c_x}{\psi_x}\right) dx.
 \end{aligned}$$

After some rearranging, we integrate:

$$\begin{aligned}
 R &= \frac{1}{\psi_x} \int_{c_x}^{c_y} \exp\left(-\frac{x-c_x}{\psi_x} + \frac{x-c_y}{\psi_y}\right) dx + \frac{1}{\psi_x} \int_{c_y}^{\infty} \exp\left(-\frac{x-c_x}{\psi_x}\right) dx \\
 &= \frac{1}{\psi_x} \int_{c_x}^{c_y} \exp\left(\frac{-\psi_y(x-c_x) + \psi_x(x-c_y)}{\psi_x\psi_y}\right) dx + \frac{1}{\psi_x} \int_{c_y}^{\infty} \exp\left(-\frac{x-c_x}{\psi_x}\right) dx \\
 &= \frac{1}{\psi_x} \int_{c_x}^{c_y} \exp\left(\frac{(\psi_x - \psi_y)x + \psi_y c_x - \psi_x c_y}{\psi_x\psi_y}\right) dx + \frac{1}{\psi_x} \int_{c_y}^{\infty} \exp\left(-\frac{x-c_x}{\psi_x}\right) dx \\
 &= \frac{1}{\psi_x} \frac{\psi_x\psi_y}{\psi_x - \psi_y} \left[\exp\left(\frac{(\psi_x - \psi_y)x + \psi_y c_x - \psi_x c_y}{\psi_x\psi_y}\right) \right]_{c_x}^{c_y} - \frac{1}{\psi_x} \psi_x \left[\exp\left(-\frac{x-c_x}{\psi_x}\right) \right]_{c_y}^{\infty}
 \end{aligned}$$

After integration we simplify to obtain:

$$\begin{aligned}
 R &= \frac{\psi_y}{\psi_x - \psi_y} \left[\exp\left(\frac{(\psi_x - \psi_y)x + \psi_y c_x - \psi_x c_y}{\psi_x\psi_y}\right) \right]_{c_x}^{c_y} - \left[\exp\left(-\frac{x-c_x}{\psi_x}\right) \right]_{c_y}^{\infty} \\
 &= \frac{\psi_y}{\psi_x - \psi_y} \left[\exp\left(\frac{(\psi_x - \psi_y)c_y + \psi_y c_x - \psi_x c_y}{\psi_x\psi_y}\right) - \exp\left(\frac{(\psi_x - \psi_y)c_x + \psi_y c_x - \psi_x c_y}{\psi_x\psi_y}\right) \right] \\
 &\quad - \left[0 - \exp\left(-\frac{c_y - c_x}{\psi_x}\right) \right] \\
 &= \frac{\psi_y}{\psi_x - \psi_y} \left[\exp\left(\frac{-c_y + c_x}{\psi_x}\right) - \exp\left(\frac{c_x - c_y}{\psi_y}\right) \right] + \exp\left(-\frac{c_y - c_x}{\psi_x}\right) \\
 &= \frac{\psi_x}{\psi_x - \psi_y} \exp\left(\frac{c_x - c_y}{\psi_x}\right) - \frac{\psi_y}{\psi_x - \psi_y} \exp\left(\frac{c_x - c_y}{\psi_y}\right) \tag{5.A.4}
 \end{aligned}$$

5.A.4. DERIVATION OF $R = P(X > Y) = 1$ FOR $c_x > c_y$ WHEN X AND Y HAVE GPD DISTRIBUTIONS

$$\begin{aligned}
R &= \int_{-\infty}^{\infty} P(X=x)P(Y < x | X=x)dx \\
&= \int_{-\infty}^{c_y} P(X=x)P(Y < x | X=x)dx + \int_{c_y}^{c_x} P(X=x)P(Y < x | X=x)dx \\
&\quad + \int_{c_x}^{\infty} P(X=x)P(Y < x | X=x)dx \\
&= \int_{-\infty}^{c_y} 0 \cdot \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx + \int_{c_y}^{c_x} 0 \cdot 1 dx + \int_{c_x}^{\infty} \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} \cdot 1 dx \\
&= \int_{c_x}^{\infty} \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} dx \\
&= \frac{1}{\psi_x} \int_{c_x}^{\infty} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} dx \\
&= \frac{1}{\psi_x} \frac{\psi_x}{\xi_x} \frac{\left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-1/\xi_x}}{-1/\xi_x} \Bigg|_{c_x}^{\infty} \\
&= - \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \Bigg|_{c_x}^{\infty} \\
&= -(0 - 1) \\
&= 1
\end{aligned}$$

5.A.5. DERIVATION OF $R = P(X > Y)$ FOR NORMAL-GPD MIXTURE

$$\begin{aligned}
R &= \int_{-\infty}^{c_x} (1 - p_x) \frac{\phi(x)}{\Phi(c_x)} p_y \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
&\quad + \int_{c_x}^{c_y} p_x \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} p_y \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
&\quad + \int_{c_y}^{\infty} p_x \frac{1}{\psi_x} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} \left[p_y + (1 - p_y) \frac{\Phi\left(\frac{x - \mu}{\sigma}\right) - \Phi\left(\frac{c_y - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)} \right] dx
\end{aligned}$$

Move constants to the front:

$$\begin{aligned}
 R &= \frac{(1-p_x)p_y}{\Phi(c_x)} \int_{-\infty}^{c_x} \phi(x) \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x p_y}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x p_y}{\psi_x} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} dx \\
 &+ \frac{p_x(1-p_y)}{\psi_x} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[\frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{c_y-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{c_y-\mu}{\sigma}\right)} \right] dx
 \end{aligned}$$

Multiply out last integral:

$$\begin{aligned}
 R &= \frac{(1-p_x)p_y}{\Phi(c_x)} \int_{-\infty}^{c_x} \phi(x) \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x p_y}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x p_y}{\psi_x} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} dx \\
 &+ \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi\left(\frac{c_y-\mu}{\sigma}\right) \right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \Phi\left(\frac{x-\mu}{\sigma}\right) dx \\
 &- \frac{p_x(1-p_y)\Phi\left(\frac{c_y-\mu}{\sigma}\right)}{\psi_x \left(1 - \Phi\left(\frac{c_y-\mu}{\sigma}\right) \right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} dx
 \end{aligned}$$

Rearrange:

$$R = \frac{(1-p_x)p_y}{\Phi(c_x)} \int_{-\infty}^{c_x} \phi(x) \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \quad (5.A.5)$$

$$+ \frac{p_x p_y}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \quad (5.A.6)$$

$$+ \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} \Phi \left(\frac{x-\mu}{\sigma} \right) dx \quad (5.A.7)$$

$$+ \frac{p_x \left(p_y - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)}{\psi_x \left(1 - \Phi \left(\frac{c_y-\mu}{\sigma} \right) \right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x-c_x}{\psi_x} \right) \right]^{-(1/\xi_x+1)} dx \quad (5.A.8)$$

Consider Integral 5.A.5.

5

$$\begin{aligned} & \frac{(1-p_x)p_y}{\Phi(c_x)} \int_{-\infty}^{c_x} \phi(x) \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\ &= \frac{(1-p_x)p_y}{\sqrt{2\pi}\Phi(c_x)} \int_{-\infty}^{c_x} e^{-x^2/2} \left[1 - \xi_y \left(\frac{x-c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \end{aligned} \quad (5.A.9)$$

Consider Integral 5.A.7. We first rewrite the integral using integration by parts:

$$\begin{aligned}
& \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x}\right)\right]^{-(1/\xi_x + 1)} \Phi\left(\frac{x - \mu}{\sigma}\right) dx \\
&= \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \left\{ \left[\Phi\left(\frac{x - \mu}{\sigma}\right) \frac{\left(1 + \xi_x \left(\frac{x - c_x}{\psi_x}\right)\right)^{-1/\xi_x}}{-1/\xi_x} \frac{\psi_x}{\xi_x} \right]_{c_y}^{\infty} \right. \\
&\quad \left. - \int_{c_y}^{\infty} \frac{\left(1 + \xi_x \left(\frac{x - c_x}{\psi_x}\right)\right)^{-1/\xi_x}}{-1/\xi_x} \frac{\psi_x}{\xi_x} \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) dx \right\} \\
&= \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \left\{ \left[-\psi_x \Phi\left(\frac{x - \mu}{\sigma}\right) \left(1 + \xi_x \left(\frac{x - c_x}{\psi_x}\right)\right)^{-1/\xi_x} \right]_{c_y}^{\infty} \right. \\
&\quad \left. + \frac{\psi_x}{\sigma} \int_{c_y}^{\infty} \phi\left(\frac{x - \mu}{\sigma}\right) \left(1 + \xi_x \left(\frac{x - c_x}{\psi_x}\right)\right)^{-1/\xi_x} dx \right\} \\
&= \frac{p_x(1-p_y)}{\psi_x \left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \left\{ \psi_x \Phi\left(\frac{c_y - \mu}{\sigma}\right) \left(1 + \xi_x \left(\frac{c_y - c_x}{\psi_x}\right)\right)^{-1/\xi_x} \right. \\
&\quad \left. + \frac{1}{\sigma \sqrt{2\pi}} \left(\frac{\xi_x}{\psi_x}\right)^{1/\xi_x} \int_{c_y}^{\infty} e^{-\left(\frac{x - \mu}{\sigma}\right)^2/2} \left(\frac{\psi_x}{\xi_x} - c_x + x\right)^{-1/\xi_x} dx \right\} \\
&= \frac{p_x(1-p_y)}{\left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \Phi\left(\frac{c_y - \mu}{\sigma}\right) \left(1 + \xi_x \left(\frac{c_y - c_x}{\psi_x}\right)\right)^{-1/\xi_x} \\
&\quad + \frac{\xi_x^{1/\xi_x} p_x(1-p_y)}{\psi_x^{1/\xi_x + 1} \sigma \sqrt{2\pi} \left(1 - \Phi\left(\frac{c_y - \mu}{\sigma}\right)\right)} \int_{c_y}^{\infty} e^{-\left(\frac{x - \mu}{\sigma}\right)^2/2} \left(\frac{\psi_x}{\xi_x} - c_x + x\right)^{-1/\xi_x} dx \quad (5A.10)
\end{aligned}$$

Consider Integral 5.A.8. We first apply the transformation ($x \rightarrow t$) with $t = \xi_x \left(\frac{x - c_x}{\psi_x}\right)$ and

jacobian $\frac{dx}{dt} = \frac{\psi_x}{\xi_x}$:

$$\begin{aligned}
 & \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\psi_x \left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \int_{c_y}^{\infty} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} dx \\
 &= \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\psi_x \left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \int_{\xi_x \left(\frac{c_y - c_x}{\psi_x} \right)}^{\infty} [1 + t]^{-(1/\xi_x + 1)} \frac{\psi_x}{\xi_x} dt \\
 &= \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\xi_x \left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \int_{\xi_x \left(\frac{c_y - c_x}{\psi_x} \right)}^{\infty} [1 + t]^{-(1/\xi_x + 1)} dt \\
 &= \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\xi_x \left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \left[\frac{(1 + t)^{-1/\xi_x}}{-1/\xi_x} \right]_{\xi_x \left(\frac{c_y - c_x}{\psi_x} \right)}^{\infty} \\
 &= \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \quad (5.A.11)
 \end{aligned}$$

Substituting Equations 5.A.9, 5.A.10 and 5.A.11 back into Equations 5.A.5, 5.A.7 and 5.A.8 we obtain an expression for the risk:

$$\begin{aligned}
 R &= \frac{(1 - p_x)p_y}{\sqrt{2\pi}\Phi(c_x)} \int_{-\infty}^{c_x} e^{-x^2/2} \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x p_y}{\psi_x} \int_{c_x}^{c_y} \left[1 + \xi_x \left(\frac{x - c_x}{\psi_x} \right) \right]^{-(1/\xi_x + 1)} \left[1 - \xi_y \left(\frac{x - c_y}{\psi_y} \right) \right]^{-1/\xi_y} dx \\
 &+ \frac{p_x(1 - p_y)}{\left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \Phi \left(\frac{c_y - \mu}{\sigma} \right) \left(1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right)^{-1/\xi_x} \\
 &+ \frac{\xi_x^{1/\xi_x} p_x(1 - p_y)}{\psi_x^{1/\xi_x + 1} \sigma \sqrt{2\pi} \left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \int_{c_y}^{\infty} e^{-(\frac{x - \mu}{\sigma})^2/2} \left(\frac{\psi_x}{\xi_x} - c_x + x \right)^{-1/\xi_x} dx \\
 &+ \frac{p_x \left(p_y - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)}{\left(1 - \Phi \left(\frac{c_y - \mu}{\sigma} \right) \right)} \left[1 + \xi_x \left(\frac{c_y - c_x}{\psi_x} \right) \right]^{-1/\xi_x} \quad (5.A.12)
 \end{aligned}$$

5.A.6. DERIVATION OF $R = P(X > Y)$ FOR NORMAL-GPD MIXTURE - GENERAL CASE

The more general mixture case occurs when the left and right tails of the exposure and effect data are modelled by gpds. In this case, the centre part of the exposure and effect data are modelled by a two-sided truncated standard normal distribution. This distribution is denoted $trN_{c_\ell}^{c_r}$ with pdf and cdf given by

$$f(x) = \frac{\phi(x)}{\Phi(c_r) - \Phi(c_\ell)} \quad (5.A.13)$$

and

$$F(x) = \frac{\Phi(x) - \Phi(c_\ell)}{\Phi(c_r) - \Phi(c_\ell)} \quad (5.A.14)$$

where $c_\ell \leq x \leq c_r$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

The pdf and the cdf of the mixture distribution is then given by

$$f(x) = \begin{cases} p_\ell GPD^-(\psi_\ell, \xi_\ell, c_\ell) & \text{if } x \leq c_\ell \\ (1 - p_\ell - p_r) trN_{c_\ell}^{c_r} & \text{if } c_\ell < x < c_r \\ p_r GPD^+(\psi_r, \xi_r, c_r) & \text{if } c_r \leq x \end{cases} \quad (5.A.15)$$

and

$$F(x) = \begin{cases} p_\ell GPD^-(\psi_\ell, \xi_\ell, c_\ell) & \text{if } x \leq c_\ell \\ p_\ell + (1 - p_\ell - p_r) trN_{c_\ell}^{c_r} & \text{if } c_\ell < x < c_r \\ 1 - p_r + p_r GPD^+(\psi_r, \xi_r, c_r) & \text{if } c_r \leq x \end{cases} \quad (5.A.16)$$

where p_ℓ (p_r) is the percentage of data in the lower (upper) tail to the left (right) of c_ℓ (c_r). Note that the cut-off points, c_ℓ and c_r can belong to either the gpd part or the truncated normal part. Here we assume that the cut-off points belong to the gpd part.

Substituting the pdfs and cdfs of the GPD^- , $trN_{c_\ell}^{c_r}$ and GPD^+ (Equations 5.1-5.4, 5.A.13 and 5.A.14) into Equations 5.A.15 and 5.A.16, we obtain the pdf and cdf of the mixture distributions of X and Y .

When deriving an expression for the risk, R , we assume that $c_{\ell,x} < c_{r,x} < c_{\ell,y} < c_{r,y}$. Using

the definition of $R = P(X > Y)$ and Equations 5.A.15 and 5.A.16, we obtain R as

$$\begin{aligned}
 R &= \int_{-\infty}^{\infty} P(X = x)P(Y < x | X = x)dx \\
 &= \int_{-\infty}^{c_{\ell,x}} P(X = x)P(Y < x | X = x)dx + \int_{c_{\ell,x}}^{c_{r,x}} P(X = x)P(Y < x | X = x)dx \\
 &\quad + \int_{c_{r,x}}^{c_{\ell,y}} P(X = x)P(Y < x | X = x)dx + \int_{c_{\ell,y}}^{c_{r,y}} P(X = x)P(Y < x | X = x)dx \\
 &\quad + \int_{c_{r,y}}^{\infty} P(X = x)P(Y < x | X = x)dx
 \end{aligned}$$

Substitute in the correct pdfs and cdfs:

$$\begin{aligned}
 R &= \int_{-\infty}^{c_{\ell,x}} p_{\ell,x} \frac{1}{\psi_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} p_{\ell,y} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
 &\quad + \int_{c_{\ell,x}}^{c_{r,x}} (1 - p_{\ell,x} - p_{r,x}) \frac{\phi(x)}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} p_{\ell,y} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
 &\quad + \int_{c_{r,x}}^{c_{\ell,y}} p_{r,x} \frac{1}{\psi_{r,x}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} p_{\ell,y} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
 &\quad + \int_{c_{\ell,y}}^{c_{r,y}} p_{r,x} \frac{1}{\psi_{r,x}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left(p_{\ell,y} + (1 - p_{\ell,y} - p_{r,y}) \frac{\Phi(x) - \Phi(c_{\ell,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \right) dx \\
 &\quad + \int_{c_{r,y}}^{\infty} p_{r,x} \frac{1}{\psi_{r,x}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \\
 &\quad \times \left(1 - p_{r,y} + p_{r,y} \left(1 - \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} \right) \right) dx
 \end{aligned}$$

Move constants to the front:

$$\begin{aligned}
R = & \frac{p_{\ell,x}p_{\ell,y}}{\psi_{\ell,x}} \int_{-\infty}^{c_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{(1 - p_{\ell,x} - p_{r,x})p_{\ell,y}}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} \int_{c_{\ell,x}}^{c_{r,x}} \phi(x) \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}p_{\ell,y}}{\psi_{r,x}} \int_{c_{r,x}}^{c_{\ell,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}p_{\ell,y}}{\psi_{r,x}} \int_{c_{\ell,y}}^{c_{r,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& + \frac{p_{r,x}(1 - p_{\ell,y} - p_{r,y})}{\psi_{r,x}(\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} (\Phi(x) - \Phi(c_{\ell,y})) dx \\
& + \frac{p_{r,x}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left(1 - p_{r,y} \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} \right) dx
\end{aligned}$$

5

Each of the last two integrals is multiplied out:

$$\begin{aligned}
R = & \frac{p_{\ell,x}p_{\ell,y}}{\psi_{\ell,x}} \int_{-\infty}^{c_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{(1 - p_{\ell,x} - p_{r,x})p_{\ell,y}}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} \int_{c_{\ell,x}}^{c_{r,x}} \phi(x) \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}p_{\ell,y}}{\psi_{r,x}} \int_{c_{r,x}}^{c_{\ell,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}p_{\ell,y}}{\psi_{r,x}} \int_{c_{\ell,y}}^{c_{r,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& + \frac{p_{r,x}(1 - p_{\ell,y} - p_{r,y})}{\psi_{r,x}(\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \Phi(x) \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& - \frac{p_{r,x}(1 - p_{\ell,y} - p_{r,y})\Phi(c_{\ell,y})}{\psi_{r,x}(\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& + \frac{p_{r,x}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& - \frac{p_{r,x}p_{r,y}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} dx
\end{aligned}$$

Lines 4 and 6 contain the same integral and are combined:

$$\begin{aligned}
R = & \frac{p_{\ell,x}p_{\ell,y}}{\psi_{\ell,x}} \int_{-\infty}^{c_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{(1 - p_{\ell,x} - p_{r,x})p_{\ell,y}}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} \int_{c_{\ell,x}}^{c_{r,x}} \phi(x) \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}p_{\ell,y}}{\psi_{r,x}} \int_{c_{r,x}}^{c_{\ell,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x}}{\psi_{r,x}} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1 - p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \int_{c_{\ell,y}}^{c_{r,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& + \frac{p_{r,x}(1 - p_{\ell,y} - p_{r,y})}{\psi_{r,x}(\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \Phi(x) \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& + \frac{p_{r,x}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& - \frac{p_{r,x}p_{r,y}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} dx
\end{aligned}$$

The integrals in Lines 4 and 6 can be evaluated. We first apply the transformation ($x \rightarrow t$) with $t = \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right)$ and jacobian $\frac{dx}{dt} = \frac{\psi_{r,x}}{\xi_{r,x}}$:

$$\begin{aligned}
R = & \frac{p_{\ell,x} p_{\ell,y}}{\psi_{\ell,x}} \int_{-\infty}^{c_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{(1 - p_{\ell,x} - p_{r,x}) p_{\ell,y}}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} \int_{c_{\ell,x}}^{c_{r,x}} \phi(x) \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x} p_{\ell,y}}{\psi_{r,x}} \int_{c_{r,x}}^{c_{\ell,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x} (1 - p_{\ell,y} - p_{r,y})}{\psi_{r,x} (\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \Phi(x) \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& - \frac{p_{r,x} p_{r,y}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} dx \\
& + \frac{p_{r,x}}{\psi_{r,x}} \frac{\Phi(c_{r,y}) p_{\ell,y} - \Phi(c_{\ell,y}) (1 - p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \int_{\xi_{r,x} \left(\frac{c_{\ell,y} - c_{r,x}}{\psi_{r,x}} \right)}^{\xi_{r,x} \left(\frac{c_{r,y} - c_{r,x}}{\psi_{r,x}} \right)} (1+t)^{-(1/\xi_{r,x}+1)} \frac{\psi_{r,x}}{\xi_{r,x}} dt \\
& + \frac{p_{r,x}}{\psi_{r,x}} \int_{\xi_{r,x} \left(\frac{c_{r,y} - c_{r,x}}{\psi_{r,x}} \right)}^{\infty} (1+t)^{-(1/\xi_{r,x}+1)} \frac{\psi_{r,x}}{\xi_{r,x}} dt
\end{aligned}$$

Next we evaluate the last two integrals:

$$\begin{aligned}
R &= \dots \\
&+ \frac{p_{r,x}}{\xi_{r,x}} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1-p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \int_{\xi_{r,x}\left(\frac{c_{\ell,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)} (1+t)^{-(1/\xi_{r,x}+1)} dt \\
&+ \frac{p_{r,x}}{\xi_{r,x}} \int_{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\infty} (1+t)^{-(1/\xi_{r,x}+1)} dt \\
&= \dots \\
&+ \frac{p_{r,x}}{\xi_{r,x}} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1-p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \frac{(1+t)^{-(1/\xi_{r,x})}}{-(1/\xi_{r,x})} \Bigg|_{\xi_{r,x}\left(\frac{c_{\ell,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)} \\
&+ \frac{p_{r,x}}{\xi_{r,x}} \frac{(1+t)^{-(1/\xi_{r,x})}}{-(1/\xi_{r,x})} \Bigg|_{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\infty} \\
&= \dots \\
&- p_{r,x} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1-p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} (1+t)^{-(1/\xi_{r,x})} \Bigg|_{\xi_{r,x}\left(\frac{c_{\ell,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)} \\
&- p_{r,x} (1+t)^{-(1/\xi_{r,x})} \Bigg|_{\xi_{r,x}\left(\frac{c_{r,y}-c_{r,x}}{\psi_{r,x}}\right)}^{\infty} \\
&= \dots \\
&- p_{r,x} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1-p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \left(1 + \xi_{r,x} \left(\frac{c_{r,y} - c_{r,x}}{\psi_{r,x}}\right)\right)^{-(1/\xi_{r,x})} \\
&+ p_{r,x} \frac{\Phi(c_{r,y})p_{\ell,y} - \Phi(c_{\ell,y})(1-p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \left(1 + \xi_{r,x} \left(\frac{c_{\ell,y} - c_{r,x}}{\psi_{r,x}}\right)\right)^{-(1/\xi_{r,x})} \\
&+ p_{r,x} \left(1 + \xi_{r,x} \left(\frac{c_{r,y} - c_{r,x}}{\psi_{r,x}}\right)\right)^{-(1/\xi_{r,x})}
\end{aligned}$$

Finally, we combine the first and the last term and obtain the expression for R :

$$\begin{aligned}
R = & \frac{p_{\ell,x} p_{\ell,y}}{\psi_{\ell,x}} \int_{-\infty}^{c_{\ell,x}} \left[1 - \xi_{\ell,x} \left(\frac{x - c_{\ell,x}}{\psi_{\ell,x}} \right) \right]^{-(1/\xi_{\ell,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{(1 - p_{\ell,x} - p_{r,x}) p_{\ell,y}}{\Phi(c_{r,x}) - \Phi(c_{\ell,x})} \int_{c_{\ell,x}}^{c_{r,x}} \phi(x) \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x} p_{\ell,y}}{\psi_{r,x}} \int_{c_{r,x}}^{c_{\ell,y}} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 - \xi_{\ell,y} \left(\frac{x - c_{\ell,y}}{\psi_{\ell,y}} \right) \right]^{-1/\xi_{\ell,y}} dx \\
& + \frac{p_{r,x} (1 - p_{\ell,y} - p_{r,y})}{\psi_{r,x} (\Phi(c_{r,y}) - \Phi(c_{\ell,y}))} \int_{c_{\ell,y}}^{c_{r,y}} \Phi(x) \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} dx \\
& - \frac{p_{r,x} p_{r,y}}{\psi_{r,x}} \int_{c_{r,y}}^{\infty} \left[1 + \xi_{r,x} \left(\frac{x - c_{r,x}}{\psi_{r,x}} \right) \right]^{-(1/\xi_{r,x}+1)} \left[1 + \xi_{r,y} \left(\frac{x - c_{r,y}}{\psi_{r,y}} \right) \right]^{-1/\xi_{r,y}} dx \\
& + p_{r,x} \left(1 - \frac{\Phi(c_{r,y}) p_{\ell,y} - \Phi(c_{\ell,y}) (1 - p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \right) \left(1 + \xi_{r,x} \left(\frac{c_{r,y} - c_{r,x}}{\psi_{r,x}} \right) \right)^{-(1/\xi_{r,x})} \\
& + p_{r,x} \frac{\Phi(c_{r,y}) p_{\ell,y} - \Phi(c_{\ell,y}) (1 - p_{r,y})}{\Phi(c_{r,y}) - \Phi(c_{\ell,y})} \left(1 + \xi_{r,x} \left(\frac{c_{\ell,y} - c_{r,x}}{\psi_{r,x}} \right) \right)^{-(1/\xi_{r,x})}
\end{aligned}$$

5

5.A.7. MAXIMUM LIKELIHOOD ESTIMATOR

In this section, we derive the maximum likelihood estimator for the parameters of the GDP^+ and GPD^- . The likelihood function is given by

$$L(\alpha, \beta, \lambda_x, \lambda_y, c_x, c_y) = \prod_{i=1}^{n_x} \alpha \lambda_x (1 + \lambda_x (x_i - c_x))^{-(\alpha+1)} \prod_{j=1}^{n_y} \beta \lambda_y (1 - \lambda_y (y_j - c_y))^{-(\beta+1)} \quad (5.A.17)$$

and the log-likelihood by

$$\begin{aligned}
\ell(\alpha, \beta, \lambda_x, \lambda_y, c_x, c_y) = & n_x \ln \alpha + n_x \ln \lambda_x - (\alpha + 1) \sum_{i=1}^{n_x} \ln(1 + \lambda_x (x_i - c_x)) \\
& + n_y \ln \beta + n_y \ln \lambda_y - (\beta + 1) \sum_{j=1}^{n_y} \ln(1 - \lambda_y (y_j - c_y)) \quad (5.A.18)
\end{aligned}$$

The cut-off point, $c_x(c_y)$, is the left(right) end point of the sample $x_1, x_2, \dots, x_{n_x} (y_1, y_2, \dots, y_{n_y})$ and can, therefore, be easily estimated as the minimum(maximum) of the sample. To

obtain the estimators of the remaining parameters, we find the first derivatives of the log-likelihood with respect to the parameters, α , β , λ_x and λ_y as

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{n_x}{\alpha} - \sum_{i=1}^{n_x} \ln(1 + \lambda_x(x_i - c_x)) \\ \frac{\partial \ell}{\partial \beta} &= \frac{n_y}{\beta} - \sum_{j=1}^{n_y} \ln(1 - \lambda_y(y_j - c_y)) \\ \frac{\partial \ell}{\partial \lambda_x} &= \frac{n_x}{\lambda_x} - (\alpha + 1) \sum_{i=1}^{n_x} \frac{x_i - c_x}{1 + \lambda_x(x_i - c_x)} \\ \frac{\partial \ell}{\partial \lambda_y} &= \frac{n_y}{\lambda_y} - (\beta + 1) \sum_{j=1}^{n_y} \frac{y_j - c_y}{1 - \lambda_y(y_j - c_y)}.\end{aligned}$$

Finding the solution of the above equations we get

$$\hat{\alpha} = \frac{n_x}{\sum_{i=1}^{n_x} \ln(1 + \hat{\lambda}_x(x_i - \hat{c}_x))} \text{ and } \hat{\beta} = \frac{n_y}{\sum_{j=1}^{n_y} \ln(1 - \hat{\lambda}_y(y_j - \hat{c}_y))}$$

and $\hat{\lambda}_x$ and $\hat{\lambda}_y$ are obtained as solutions of

$$\begin{aligned}0 &= \frac{n_x}{\lambda_x} - \frac{n_x}{\sum_{i=1}^{n_x} \ln(1 + \lambda_x(x_i - \hat{c}_x))} \sum_{i=1}^{n_x} \frac{x_i - \hat{c}_x}{1 + \lambda_x(x_i - \hat{c}_x)} + \sum_{i=1}^{n_x} \frac{x_i - \hat{c}_x}{1 + \lambda_x(x_i - \hat{c}_x)} \\ 0 &= \frac{n_y}{\lambda_y} - \frac{n_y}{\sum_{j=1}^{n_y} \ln(1 - \lambda_y(y_j - \hat{c}_y))} \sum_{j=1}^{n_y} \frac{y_j - \hat{c}_y}{1 - \lambda_y(y_j - \hat{c}_y)} + \sum_{j=1}^{n_y} \frac{y_j - \hat{c}_y}{1 - \lambda_y(y_j - \hat{c}_y)}.\end{aligned}$$

The latter two equations can be written in the form $g(\lambda) = \lambda$ with

$$\begin{aligned}g(\lambda_x) &= n_x \left(\frac{n_x}{\sum_{i=1}^{n_x} \ln(1 + \lambda_x(x_i - \hat{c}_x))} \sum_{i=1}^{n_x} \frac{x_i - \hat{c}_x}{1 + \lambda_x(x_i - \hat{c}_x)} + \sum_{i=1}^{n_x} \frac{x_i - \hat{c}_x}{1 + \lambda_x(x_i - \hat{c}_x)} \right)^{-1} \\ g(\lambda_y) &= n_y \left(\frac{n_y}{\sum_{j=1}^{n_y} \ln(1 - \lambda_y(y_j - \hat{c}_y))} \sum_{j=1}^{n_y} \frac{y_j - \hat{c}_y}{1 - \lambda_y(y_j - \hat{c}_y)} + \sum_{j=1}^{n_y} \frac{y_j - \hat{c}_y}{1 - \lambda_y(y_j - \hat{c}_y)} \right)^{-1}.\end{aligned}$$

We then obtain λ_x and λ_y in an iterative manner as

$$\lambda_{x(k+1)} = g(\lambda_{x(k)}) \text{ and } \lambda_{y(k+1)} = g(\lambda_{y(k)}),$$

stopping when the absolute difference between consecutive λ 's is sufficiently small.

5.B. FIGURES

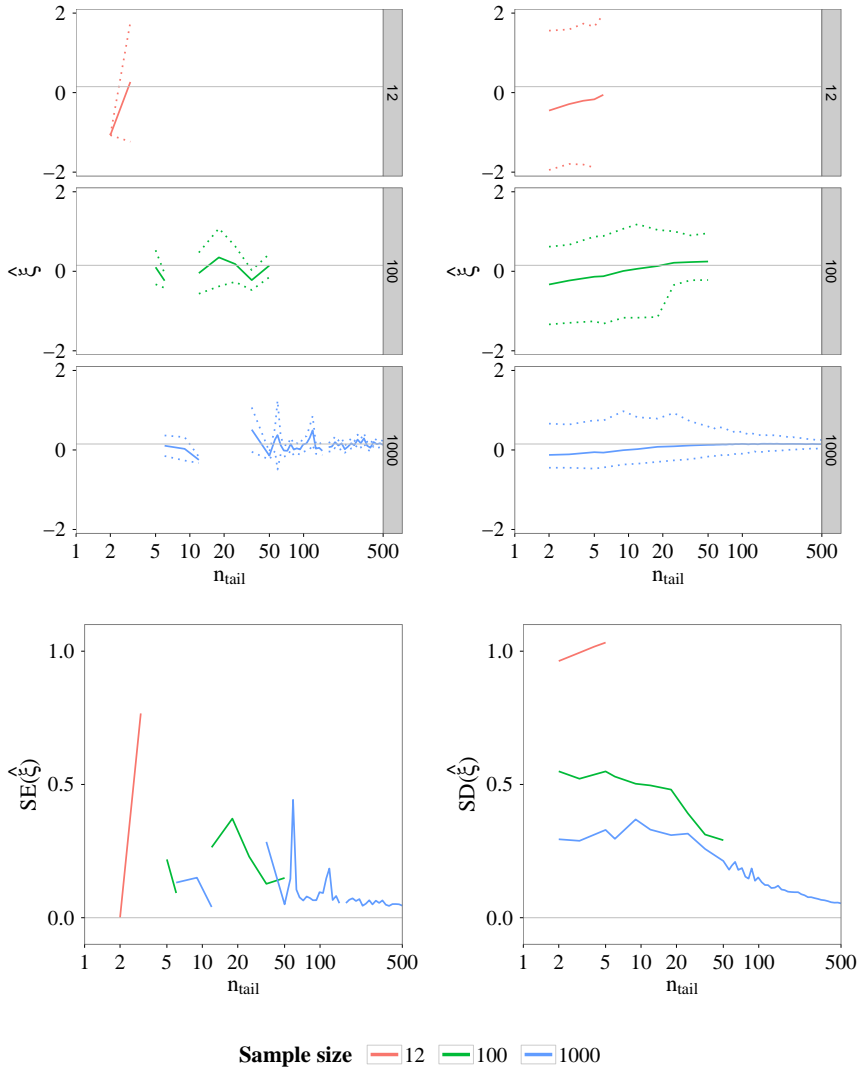


Figure 5.B.1: Plot of $\hat{\xi}$ and $SE(\hat{\xi})$ as a function of n_{tail} for a single iteration (first column) and for 1000 iterations (second column). The dotted lines indicate an approximate 95% confidence interval for $\hat{\xi}$. The three colours indicate the three sample sizes 12, 100 and 1000. The horizontal grey lines in the top two graphs indicate the true value of $\xi = 0.15$.

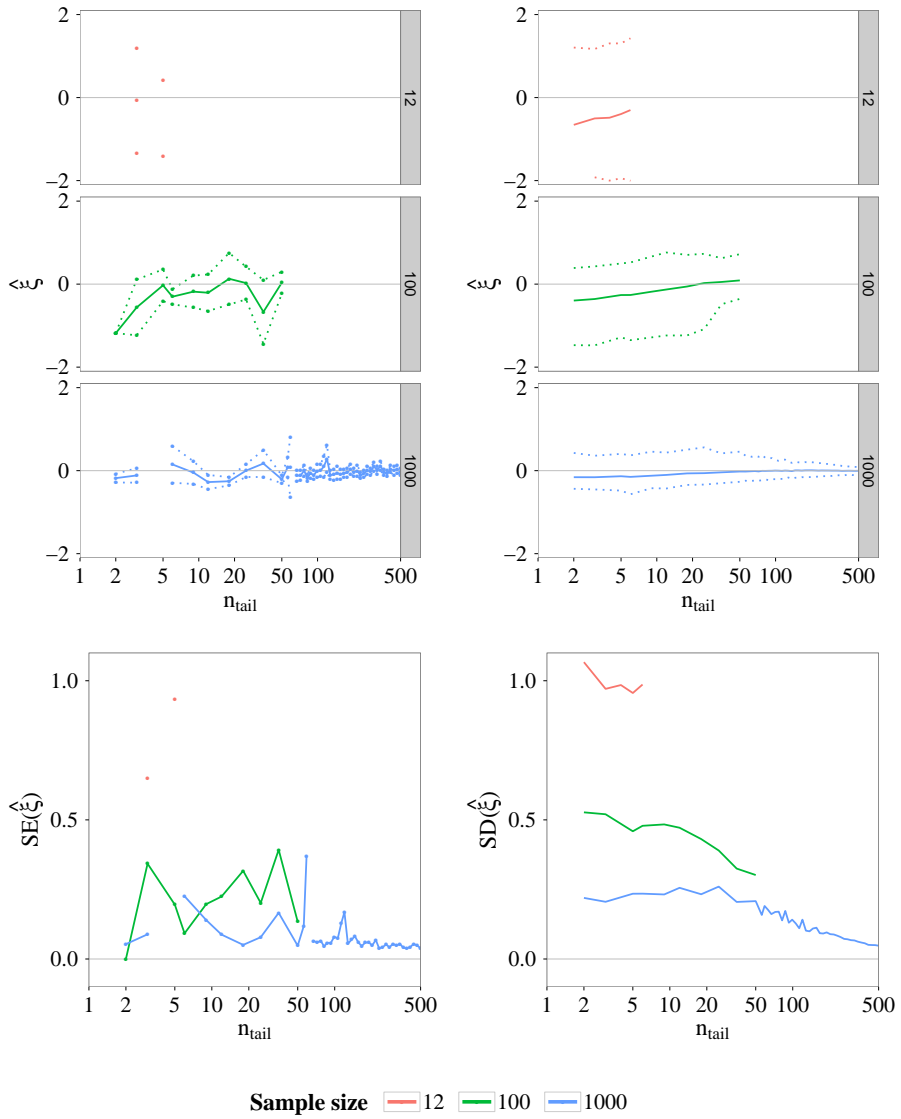
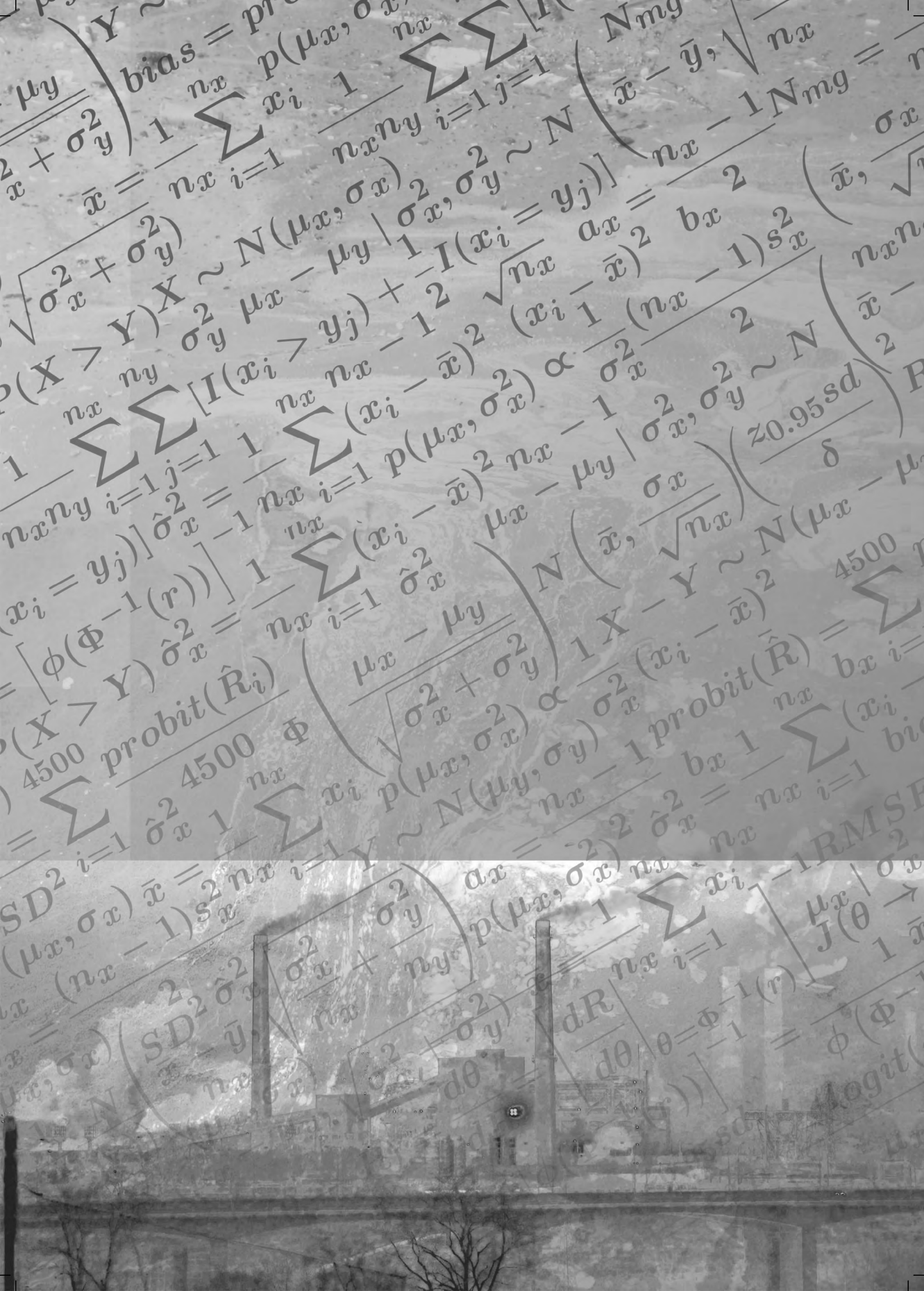


Figure 5.B.2: Plot of $\hat{\xi}$ and $SE(\hat{\xi})$ as a function of n_{tail} for a single iteration (first column) and for 1000 iterations (second column). The dotted lines indicate an approximate 95% confidence interval for $\hat{\xi}$. The three colours indicate the three sample sizes 12, 100 and 1000. The horizontal grey lines in the top two graphs indicate the true value of $\xi = 0$.



6

GENERAL DISCUSSION

Rianne Jacobs

6.1. INTRODUCTION

In this thesis, I developed statistical methods for risk assessment of nanoparticles. These methods deal with the challenges of quantifying uncertainty and variability and of small sample sizes. Before I discuss the methods studied and conclusions obtained in this thesis, I first provide the basic requirements, in the context of nanoparticle risk assessment, for any statistical model development to be meaningful.

The results obtained from statistical modelling are only as good as the data on which they are applied. Although this is a well-known fact, we find that, in practice, it is often not considered when presenting results and drawing conclusions. In the field of risk assessment, the quality of the results and the conclusions is of special importance. The results of a risk assessment are crucial for risk managers in setting up guidelines and protocols for the safe production and use of chemicals, conventional or nano.

Nanotechnology is a novel technology with promising applications but with no history of safe use. Risk assessment of ENPs should result in the assurance that ENPs pose no harm to humans and the environment or, if harmful to a certain extent, should provide guidelines for safe production and use. At the same time, risk assessors must guard against an over-conservative risk assessment which, although protective, may result in unnecessarily stringent regulation hampering the development of a promising new technology. Due to the novelty of nanotechnology, nanoparticle risk assessment is hampered by many uncertainties. These uncertainties can be quantitative or qualitative in nature.

Many of the uncertainties surrounding the risk assessment of ENPs are due to uncertainties in each of the variables used for intermediate calculations in the exposure assessment, hazard assessment and risk characterisation. These uncertainties are quantitative and are not difficult to deal with from a statistical point of view. They can be quantified by distributions, provided there is enough knowledge available to provide a valid basis for the choice of the distributions and their parameters.

A more difficult and fundamental issue surrounding the risk assessment of ENPs is that of qualitative uncertainty. Qualitative uncertainty is not quantifiable, either due to its nature or due to lack of knowledge. Let us consider the uncertainty surrounding the measurement of ENPs. Although the final risk measure that is calculated is dimension-

less, the two parts constituting the risk assessment, namely exposure and hazard assessment are not dimensionless. For any risk assessment to be meaningful and reliable, the measurement unit should be the same for the exposure and the hazard assessment. If this is not the case, then the risk assessment results are meaningless, regardless of the quality of the statistical methods used. I will discuss three measurement aspects: the actual unit of measurement or dose metric, the form of the substance being measured and the time scale of the measurements.

First, I discuss the actual unit of measurement, such as mass (μg), surface area (nm^2) or number of particles. As mentioned in Chapter 1, risk assessors are still uncertain as to which metric is the best to use in the case of ENPs. Is increasing toxic effect of ENPs related to increasing mass, surface area or number of ENPs? The answer to this question may depend on the risk assessment area, such as environmental toxicity, human inhalation toxicity or human oral toxicity. This is a question that toxicologists need to answer before the correct exposure assessments can be conducted. The relevant metric for toxicity must also be the metric that is used to measure or calculate the exposure concentrations.

Second, we need to consider the form of the substance being measured. The question that needs to be answered is: Are we talking about free single ENPs, larger aggregated ENPs or ENPs attached to natural colloids or larger natural particles? Again this goes back to the toxicologists. What form of the ENPs is causing the toxicity? Toxicity testing of ENPs is done in controlled laboratory environments using free ENPs. In the environment or in complex food matrices, however, ENPs rarely occur in free particle form. Moreover, the ENP form can also change during the exposure period. This was, for example, found during *in vitro* digestion of foods containing nanosilica in which ENPs were found in different forms during different stages of the digestion process (Peters et al., 2012). Toxicity testing is, therefore, performed on a substance (free ENPs) to which humans and the environment may not be exposed at all. At the same time, we do not know the toxicity of the substance (aggregated and attached ENPs) that humans and the environment are exposed to in reality. Applying any form of statistical modelling on such incompatible data is meaningless. Some discussion on this topic can be found in Chapter 3 Section 3.4.1.

Third, I discuss the time scale of the measurements. The concentrations of the hazard

assessment are usually obtained in a laboratory setting while exposure concentrations relate to field data (Aldenberg et al., 2002). Chronic effect concentrations cannot be compared with acute exposure concentrations. Specifically, 96-h toxicity test endpoints cannot be compared to hourly fluctuating concentrations at a discharge point (Aldenberg et al., 2002). The time interval of exposure measurements or simulation results should be equal to (or larger than) the time interval of toxicity testing, otherwise the resulting risk assessment cannot be interpreted (Verdonck et al., 2003) even when using the most sophisticated statistical methods.

The three measurement aspects discussed above are examples of qualitative uncertainty. These cannot be accounted for statistically and are considered as conditions needed before applying any statistical model. In this thesis, therefore, I assumed that (i) the exposure and critical effect concentrations are measured in the same toxicologically relevant unit, (ii) the exposure and critical effect concentrations are actually referring to the same ENP form and (iii) the temporal scale of the exposure and the hazard assessment is the same. The quality of the results and the usefulness of the conclusions drawn in this thesis concerning the risk assessment of ENPs depend on these assumptions. The statistical methods developed in this thesis may, in themselves, be good methods, but without these assumptions, no meaningful or useful risk assessment can be performed. A possible way to deal with these qualitative uncertainties is via scenarios. In Chapters 2 and 3, I paid attention to this by considering different dose metrics (Chapter 2) and different ENP forms (Chapter 3).

6.2. QUANTIFYING UNCERTAINTY AND VARIABILITY

6.2.1. TWO-DIMENSIONAL MONTE CARLO

The first objective was to perform a transparent risk assessment of nanoparticles in the face of large uncertainty in such a way that it can guide future research to reduce the uncertainty.

An intuitive way of quantifying uncertainty and variability is by using two-dimensional Monte Carlo (2DMC) analysis (Cohen et al., 1996; Frey, 1992; Frey and Rhodes, 1996; Simon, 1999). This method is schematically represented in Figure 6.1. Uncertainty and

variability of an exposure or effect variable are described by distributions. This is done by specifying the distribution of the variable as variability and specifying the parameters of that distribution as uncertain (Cohen et al., 1996). We can then specify a joint variability distribution for all the variables and a joint uncertainty distribution for the parameters of the joint variability distribution.

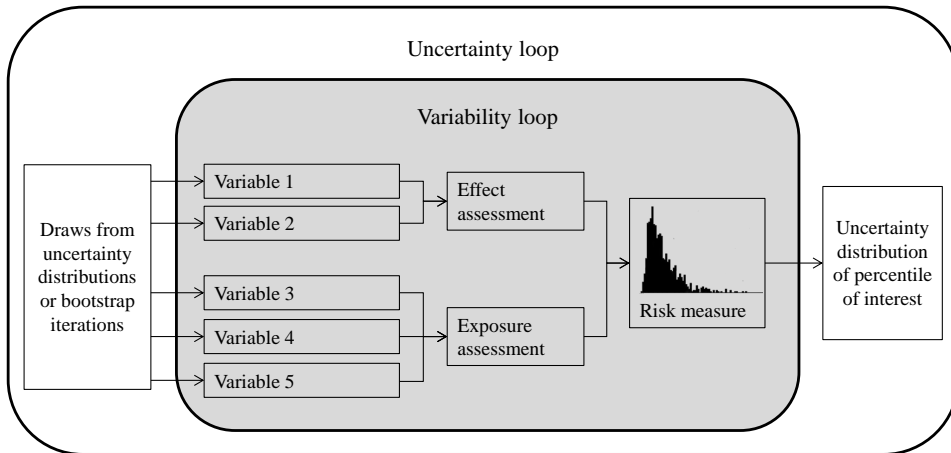


Figure 6.1: Schematic representation of a nested two-dimensional Monte Carlo algorithm for the separate quantification of uncertainty and variability.

The 2DMC algorithm starts by drawing a sample of size N from the joint uncertainty distribution of the parameters. In each uncertainty iteration (outer loop of Figure 6.1), the joint variability distribution is specified for a given set of parameters from the uncertainty sample. In the variability loop (inner loop of Figure 6.1), a Monte Carlo simulation is performed by repeatedly sampling (M times) from the specified joint variability distribution and calculating the risk measure of interest. This simulation results in M risk measure values representing the variability distribution of the risk measure for a given set of the uncertain parameters. Repeating this simulation for each of the N uncertainty iterations, results in a $N \times M$ matrix of risk measures where each row represents the variability distribution of the risk measure given a particular draw from the joint uncertainty distribution. If each row of the matrix is sorted from small to large, then each column represents the uncertainty distribution of M equally spaced percentiles of the variability distribution of the risk measure. In the case of large data sets, the variability distribution

of the exposure and effect input variables can be obtained empirically. The uncertainty is then quantified by resampling using the bootstrap (Efron and Tibshirani, 1993).

The 2DMC method has primarily been applied in the human health area of risk assessment with applications to exposure assessment of copper arsenate (Xue et al., 2006) and to risk assessment of chemicals (van der Voet et al., 2009), insecticides (Schleier et al., 2009), pesticides (Bosgra et al., 2009; Müller et al., 2009) and fungicides (Muri et al., 2009), to name a few. A quick scan of the literature found only one environmental application, namely, to risk assessment of mercury (Lohman et al., 2000). I implemented the 2DMC method for the case of nanosilica in food (Chapter 2) and nanoTiO₂ in the aquatic environment (Chapter 3). To my knowledge, this is the first comprehensive attempt at applying the 2DMC method to risk assessment of ENPs.

In Chapter 2, I applied the 2DMC method in health risk assessment. The 2DMC method is implemented in MCRA software (van der Voet et al., 2015) as the Integrated Probabilistic Risk Assessment (IPRA) method (van der Voet and Slob, 2007; van der Voet et al., 2009) for health risk assessment of chemicals. For most of the variables, variability and uncertainty were quantified using parametric distributions. For the consumption variable, consumption data from the Dutch National Food Consumption Survey was used to estimate the empirical variability distribution. Uncertainty was quantified via the bootstrap. For the concentration variable, I also used the empirical method to quantify variability and uncertainty.

In Chapter 3, I applied the 2DMC method to aquatic risk assessment. I implemented the model using R Software. Variability and uncertainty were quantified using parametric distributions, with the exception of two of the exposure variables for which large datasets were available. Because of the amount of variables and the complexity of the exposure model, I quantified, for each variable, either the variability or the uncertainty, whichever one was thought to constitute the biggest source of variation for that variable. Furthermore, due to inadequate information from literature about the possible ranges of the variables, it was not possible to determine which part of the variation of a variable was due to variability and which was due to uncertainty. Expert judgement (Flari et al., 2011) could possibly be used in future to improve the model.

2DMC simulation is an intuitive method to quantify uncertainty and variability. It is a

transparent way of following the effect of uncertainty and variability from the source to the final risk measure. It is also relatively easy to implement and, therefore, appropriate for complex models (Kelly and Campbell, 2000). The main challenges of 2DMC are (i) the choice of whether to quantify both uncertainty and variability for all variables and (ii) which uncertainty and variability distributions to use.

Concerning the first challenge, it is, in theory, possible to represent all variables as having both uncertainty and variability (Kelly and Campbell, 2000). In practice, however, this can become problematic, especially for complex models (Kelly and Campbell, 2000). The SimpleBox4Nano (SB4N) model used in Chapter 3 for the exposure assessment is an example of such a complex model. One, therefore, has to make a choice about which variables to include in the 2D analysis and for which ones to only consider a 1D analysis (i.e. only uncertainty or only variability). Although such simplifications are practical, care must be taken when interpreting the results, because not all variation might be captured correctly (Frey and Rhodes, 1998). Regardless of the choices that are made, they must be clearly reported, and the results of the analysis should be interpreted in the light of these choices.

The second challenge relates to the parametric quantification of uncertainty and variability. In this case, the choice of distributions and parameters can have a significant effect on the results of the risk assessment. Frey and Rhodes (1998) provide a detailed investigation on the choice of distributions and their effect on the model outputs.

6.2.2. BAYESIAN ANALYSIS

Bayesian analysis is well suited for quantifying uncertainty of parameters. In the Bayesian framework, the uncertainty of the parameters of a distribution are quantified by a prior distribution. The prior distribution can be obtained from historical data or some prior (subjective) belief about the scope of the uncertainty of the parameter. The prior distribution is then updated with the data. This updated distribution is known as the posterior distribution.

The quantification of uncertainty and variability using Bayesian statistics has been discussed by various authors (Kennedy and Hart, 2009; Nayak and Kundu, 2001). Nayak and Kundu (2001) developed a general Bayesian framework for the quantification of uncer-

tainty. The authors do not mention variability, although it is implicitly accounted for in their method. The separate quantification of uncertainty and variability in a Bayesian framework was developed in detail by Kennedy and Hart (2009). They developed the method specifically for the dietary risk assessment of pesticides.

The method developed by Kennedy and Hart (2009) constitutes a Bayesian analysis implemented in a 2DMC framework. Consider, for example, an exposure model that calculates the exposure concentration (ExpC). Variability in the ExpC is due to variability in the input variables of the exposure model. This variability is quantified by a joint variability distribution. The parameters of this variability distribution are uncertain. Historical data or expert judgement can be used to construct a joint prior distribution which quantifies the uncertainty of the parameters. Updating this prior with the data, we obtain the joint posterior distribution. Markov Chain Monte Carlo (MCMC) can then be used to draw a sample from the joint posterior distribution of the parameters. This sample serves as the input for the 2DMC algorithm described in the previous section.

The Bayesian implementation as explained above can be extended for the case where we deal with data that has a hierarchical structure. Consider, for example, the SSD model of Chapter 3. This model calculates the chronic critical effect concentration (CEC) as

$$CEC_{chronic} = \frac{CONC}{AF_{time} \cdot AF_{no-effect}}$$

where CONC is the limit concentration (e.g. LC₁₀ or EC₂₀) obtained from a toxicological study, AF_{time} is the assessment factor to extrapolate from acute to chronic studies and AF_{no-effect} is the assessment factor to extrapolate from the limit concentration to the CEC. We are interested in the variability distribution of $CEC_{chronic}$, i.e. the SSD. In other words, to obtain the SSD, we need a $CEC_{chronic}$ value for each species under consideration. When considering the data (see Table 3.B.1 in Chapter 3 Appendix 3.B), we note that the data has a hierarchical structure. The limit concentrations can be divided into groups according to the species. It is possible to define a one-way random effects model for such hierarchical data. A one-way random effects model has been applied in a similar situation for modelling occupational exposure (Harvey and van der Merwe, 2014; Krishnamoorthy and Mathew, 2002). I will discuss the model and its possible implementation in a 2DMC structure. Note that the model is built on $CEC_{chronic}$ values,

implying that the assessment factors should already have been applied. The uncertainty of the assessment factors can be quantified by a joint uncertainty distribution.

Let X_{ij} denote the j^{th} $CEC_{chronic}$ for the i^{th} species, $i = 1, 2, \dots, k; j = 1, 2, \dots, n$. Assume X_{ij} to be log-normally distributed. Then $Y_{ij} = \ln(X_{ij})$ is normally distributed. The one-way random effects model for Y_{ij} is given by

$$Y_{i,j} = \mu + \Theta_i + e_{ij}$$

where μ is the overall mean, $\Theta_i \sim N(0, \sigma_s)$ is the random effect due to the i^{th} species and $e_{ij} \sim N(0, \sigma_e)$ is the random error. Let

$$\mu_{x_i} = E(X_{ij}|\Theta_i) = E(\ln(Y_{ij})|\Theta_i) = \exp(\mu + \Theta_i + \sigma_e^2/2)$$

be the mean $CEC_{chronic}$ for the i^{th} species and $\ln(\mu_{x_i}) \sim N(\mu + \sigma_e^2/2, \sigma_s^2)$. The μ_{x_i} represent the values needed to construct the SSD.

Harvey and van der Merwe (2014) provided a Bayesian approach for fitting this model. The likelihood function is obtained from the normal distribution. We require a joint prior for $(\mu, \sigma_s^2, \sigma_e^2)$ quantifying the uncertainty in these parameters. Combining the prior distribution and the likelihood, Harvey and van der Merwe (2014) estimated the posterior distribution of $\mu + \Theta_i + \sigma_e^2/2$. MCMC can then be used to draw a sample from the posterior distribution which then serves as the input for the 2DMC algorithm described in the previous section.

In this example, we assumed a balanced random effects model in which we have the same number of $CEC_{chronic}$ values for each species. In light of the data used in Chapter 3, this is not a realistic scenario. The Bayesian analysis of the random effects model can be extended to the unbalanced case with unequal number of $CEC_{chronic}$ values for each species (see Harvey and van der Merwe (2015)).

In this section, I discussed how the 2DMC algorithm can be extended to include a Bayesian analysis of uncertainty. Especially the use of hierarchical Bayes models may prove to be very useful in the risk assessment context. Hierarchical Bayes models can also be used to pool data from various sources when individual datasets are too small. Kennedy

et al. (2011) developed such a model which pools information from various field trials to model extreme pesticide residues.

6.3. ESTIMATING THE RISK PROBABILITY, $P(X > Y)$

The second objective was to evaluate empirical and parametric methods to estimate the risk probability, $R = P(\text{exposure} > \text{critical effect})$, in the case of small sample sizes.

In risk assessment, we compare the exposure with the hazard. This comparison is quantified as $CR = \frac{ExpC}{CEC}$ or $MoE = \frac{\text{limit concentration}}{\text{human exposure}}$. In Chapters 2 and 3, I used either parametric (i.e. from a theoretical distribution) or empirical (i.e. from data, as in resampling or bootstrap) sampling methods to quantify uncertainty and variability in the concentration ratio (CR) (Chapter 3) and margin of exposure (MoE) (Chapter 2). When using sampling methods, the CR is calculated by repeatedly dividing a randomly drawn ExpC by a randomly drawn CEC. The risk probability, $P(CR > 1) = P(ExpC > CEC)$, is then estimated by counting the number of times the CR is greater than 1. This method requires infinite or at least a very large number of ExpCs and CECs. This was the case for the methods implemented in Chapters 2 and 3. Either large datasets were available (e.g. the Dutch National Food Consumption Survey) or a sample of the required size was drawn from a parametric distribution. The question arises, are (re)sampling methods also appropriate to use when large datasets are unavailable? To what extent do empirical methods provide valid results as the sample size of, for example, the CEC decreases? These questions were answered in Chapter 4 by comparing empirical methods with a simple parametric model to estimate $P(ExpC > CEC)$.

6.3.1. THE NORMAL-NORMAL MODEL

In environmental risk assessment, it is common practice to assume a normal distribution to describe the variability in log ExpCs and log CECs (Aldenberg et al., 2002; Verdonck et al., 2003; Wagner and Løkke, 1991). In Chapter 4, I used the normal-normal

model to estimate $R = P(ExpC > CEC)$. In the normal-normal model we have

$$\begin{aligned} ExpC &\sim N(\mu_x, \sigma_x^2) \\ CEC &\sim N(\mu_y, \sigma_y^2) \end{aligned} \Rightarrow R = \Phi \left(\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right) \quad (6.1)$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. I compared two ml type estimators, a Bayesian and an empirical estimator in the estimation of R for small sample sizes and low risks. The simulation study showed that the empirical estimator fails. The parametric estimators performed better. I found that the Bayesian estimator with non-informative prior performed better than the ml estimators.

In Chapter 4, I considered methods based on the standard likelihood function. The ml estimator uses the invariance property of ml estimators, in which the ml estimator of a function of parameters is equal to the function of the ml estimators of the parameters. As is expected, I found that the ml estimators did not perform well for small samples sizes.

Not only the ml estimators, but also the Bayesian estimator used the standard likelihood function, using a variable transformation method to obtain the joint posterior distribution of $(R, \sigma_x^2, \sigma_y^2)$. The marginal posterior distribution of R was then obtained by integrating out σ_x^2 and σ_y^2 . One of the limitations of this method is that the priors are specified on the nuisance parameters μ_x, μ_y, σ_x^2 and σ_y^2 instead of directly on R . Using non-informative priors on the nuisance parameters, resulted in a prior tendency of R towards 0.5. For small sample sizes, this prior tendency results in bias in the estimation of R towards 0.5. This bias can either be viewed as unwanted or as a way of obtaining a conservative risk assessment when there is too little data available.

Inference on R can also be based on a likelihood function depending only on R instead of indirectly via the likelihood function of the parameter $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$. Such methods are widely available in the literature and make use of pseudo-likelihood functions such as marginal likelihoods, profile likelihoods and modified profile likelihoods (Severini, 2000, p. 279). In this context, we reparameterise θ as $\theta = (R, \lambda)$ with R the parameter of interest (Equation 6.1) and, for example, $\lambda = (\mu_y, \sigma_x^2, \sigma_y^2)$ the nuisance parameter (Cortese and Ventura, 2013). Inference on R , such as confidence intervals, is then based on the profile likelihood, $l_p(R, \hat{\lambda}_R)$, with $\hat{\lambda}_R$ the constrained maximum likelihood estimator (Cortese and Ventura, 2013). The problem with $l_p(R, \hat{\lambda}_R)$ is that it is not a genuine

likelihood function because it is not based on the density function of a random variable (Severini, 2000, p. 323). Replacing λ by its ml estimate can have a large impact on inferences in small samples (Severini, 2000, p. 324).

Ideally, one would want to base inference on a genuine likelihood which is only a function of R . The marginal or conditional likelihood can be used in this context. In this discussion, I will focus on the marginal likelihood. Suppose there exists a statistic, T , such that the density of the data, Y , can be written as

$$p(y; R, \lambda) = p(t; R)p(y|t; R, \lambda).$$

The marginal distribution, $p(t; R)$, of T depends only on the parameter of interest, R . The marginal likelihood function (Severini, 2000, p. 298) is then obtained as

$$l_m(R; t) = p(t; R).$$

If this marginal likelihood exists and can be computed exactly, it gives exact inference for the parameter R (Reid and Fraser, 2003). The marginal likelihood may, however, not exist or be difficult to calculate (Severini, 2000, p. 323). The modified profile likelihood is a good alternative in this regard.

The modified profile likelihood can be derived as an approximation to the marginal likelihood when that likelihood exists. Moreover, the calculation of the modified profile likelihood does not require the existence of the marginal likelihood and is, therefore, more versatile than the marginal likelihood (Severini, 2000, p. 323). The modified profile likelihood can be expressed as

$$l_{mo}(R) = l_p(R, \hat{\lambda}_R)c(R)$$

where $c(R)$ is a correction term. The modified profile likelihood with correction term according to Barndorff-Nielsen (1983) results from an application of the Laplace approximation (Cox and Reid, 1987; Reid, 1996). This is closely linked to the use of the Laplace approximation for computing marginal posterior densities (Cox and Reid, 1987; Reid, 1996; Tierney and Kadane, 1986). The well-known Cox-Reid correction (Cox and Reid, 1987) is a special case of the Barndorff-Nielsen correction term. The Cox-Reid correction term requires orthogonality of the parameters and is not invariant to reparametri-

sation of the nuisance parameters (Cox and Reid, 1987). The modified profile likelihood improves the classical likelihood-based inference by reducing the effect of estimation of nuisance parameters and improving the approximations for small sample sizes (Cortese and Ventura, 2013). The improved performance of the modified profile likelihood over the profile likelihood may be partially explained by the fact that the modified profile likelihood is an approximation of the marginal likelihood, which the profile likelihood is not (Severini, 2000, pp. 327-330).

Standard Bayesian analysis is based on a prior and the classical likelihood function. The joint posterior for (R, λ) can be obtained by a suitable transformation as was done in Chapter 4. Two drawbacks in this standard approach are that (i) it requires specification of a prior distribution on the complete parameter space and (ii) numerical integration is required to obtain the marginal posterior density, $\pi(R|x, y)$, of R (Ventura and Racugno, 2011). The computational problem can be solved by using the Laplace approximation (Tierney and Kadane, 1986) to approximate the marginal posterior density (Ventura and Racugno, 2011). Alternatively, the modified profile likelihood can be used to obtain the marginal posterior distribution of R , which, in addition to solving the computational problem, also does not require prior specification on the complete parameter space. Treating the modified profile likelihood as a genuine likelihood, we can obtain the posterior density as

$$\pi^*(R|x, y) \propto \pi^*(R) l_{mo}(R) \quad (6.2)$$

where $l_{mo}(R)$ is the modified profile likelihood and $\pi^*(R)$ is referred to as the matching prior for R (Ventura and Racugno, 2011). If the modified profile likelihood with the Barndorff-Nielsen correction term is used, the matching prior is proportional to the square root of the inverse of the asymptotic variance of the maximum likelihood estimator of R (Ventura et al., 2009; Ventura and Racugno, 2011). Note that using such a prior is related to empirical Bayesian analysis in which the prior distribution is based on the data (Gelman et al., 2014). The approximate posterior tail probabilities obtained from a Bayesian analysis based on Equation 6.2 correspond to the approximate frequentist tail probabilities when using a modified profile likelihood (Ventura and Racugno, 2011). The prior, $\pi^*(R)$, is, therefore, a strong matching prior (Ventura and Racugno, 2011). Note also, that the prior is now only specified on R and fixed (i.e. The Bayesian analysis cannot benefit from prior information).

Ventura and Racugno (2011) derived the posterior distribution of R using a strong matching prior in the normal-normal model with equal variances. Moreover, Cortese and Ventura (2013) provide details for the unequal variance case using the R package ProbYX (Cortese, 2015).

Although the literature on strong matching priors seem to promote the idea that it is good to have matching Bayesian and frequentist probabilities, I would like to argue to the contrary. Using a strong matching prior undermines the Bayesian philosophy and removes the advantage of Bayesian analysis. The advantage of Bayesian analysis is that the current data can be combined with other sources of knowledge via the prior distribution. With this in mind, it is not necessarily the case that posterior probabilities correspond to their frequentist counterpart. I want to go a step further and claim that this correspondence is not even desirable nor to be expected. The whole idea behind the Bayesian framework is the possibility to include more information than is possible in a frequentist setting. In the context of this thesis, such information could come from expert judgement or data from, for example, conventional chemicals. Including extra information in a method will give results that are different to those obtained from a method which does not include that information. I would like to conclude with the following thought: If the goal of Bayesian analysis is to produce results identical to those obtained in a frequentist setting, why not stick to frequentist statistics in the first place?

6.3.2. ALTERNATIVE TO NORMAL-NORMAL MODEL

Following the evaluation of estimators for R of the normal-normal model, the question arose as to the appropriateness of the use of normal distributions for the exposure and critical effect. The normal distribution is considered a light-tailed distribution (Carreau and Bengio, 2008). This might not always be appropriate when modelling data. A natural alternative to the normal distribution to allow heavier tails is Student's t -distribution. The t -distribution has the normal distribution as limiting distribution for increasing degrees of freedom, ν . For smaller ν , the t -distribution becomes flatter with heavier tails, being equal to the Cauchy distribution for $\nu = 1$ (Bain and Engelhardt, 1992, p. 274).

Although the t -distribution allows for heavy tails, it is still a symmetric distribution. In the context of risk assessment, this is not necessarily useful. When considering the es-

timination of $R = P(\text{Exp}C > \text{CEC})$ for low risks, it is especially the upper tail of the $\text{Exp}C$ distribution (i.e. exposure concentration distribution (ECD)) and the lower tail of the CEC distribution (i.e. species sensitivity distribution (SSD)) that determine the value of R . These tails are referred to as the inner tails.

In Chapter 5, I developed a model to allow more flexibility in the estimation of the inner tails of the ECD and SSD . The model assumes gpds for the inner tails of the ECD and SSD and a normal distribution for the bulk of the data and is referred to as the normal-gpd model. The gpd allows for a wide range of tail heaviness from very light finite tails, through the normal tail to heavy tails such as those of the t and cauchy distributions. The normal-gpd model was implemented using the R package *evmix* (Scarrott and Hu, 2015). This package also allows for a variety of other parametric models (gamma, weibull, log-normal, beta) and a nonparametric model (kernel density estimation) for the bulk data.

Via simulation, I showed that the risk, $R = P(\text{Exp}C > \text{CEC})$, can be severely underestimated if the normal-normal model is used instead of the normal-gpd model, especially for low risks. I found that it is not possible to estimate the tail index for the normal-gpd model when sample sizes are small. This implies that heavy tails cannot be estimated from the data for small sample sizes. This is also seen in the t -distribution case. I found that, for small sample sizes, the data did not contain any useful information for estimating the number of degrees of freedom.

In the case of small sample sizes for the CECs , when the normal-gpd model cannot be estimated, the methods in Chapter 5 can be used as a sensitivity analysis. In this context, it is possible to determine the range of possible R values for a given tail heaviness of the $\text{Exp}Cs$. Moreover, there seems to be an upper bound on the risk probability of around 0.02 for the cases investigated.

6.4. PROBABILISTIC RISK ASSESSMENT IN A REGULATORY CONTEXT

This thesis work is about probabilistic methods for risk assessment of ENPs . I have shown that probabilistic methods are necessary for proper risk assessment of ENPs . Such risk assessment should prevent over-conservatism and be transparent to allow for

focused research into the risks of ENPs. The next question that arises is that of interpretation and integration into regulatory guidelines. In this section, I will discuss the importance of correct interpretation of probabilistic risk assessment results. Furthermore, I will attempt to answer the question: How do we make the results of a probabilistic risk assessment understandable to risk managers and regulators?

In probabilistic risk assessment, we compare the ECD with the SSD. This comparison is done via the risk probability $R = P(ExpC > CEC)$. The danger of interpreting a probabilistic risk assessment result such as $R = P(ExpC > CEC)$ is that it can be interpreted as in a deterministic risk assessment, i.e. as a single value. It is worrying that this is exactly what the ECHA is suggesting: “The only number that needs to be communicated is the expected risk [i.e. $P(ExpC > CEC)$], which is a single number” (ECHA, 2012c)

Verdonck et al. (2003) discuss this problem of single number interpretation by showing that different scenarios for exposure and species sensitivity result in the same numerical value for R . Suppose we have an ECD describing variability in exposure concentrations between environmental regions and a SSD describing variability in species sensitivity. A risk probability value of 0.5 can correspond with many different situations, from (i) 50% of species are at risk in all regions to (ii) all species are at risk in 50% of the regions. To differentiate which interpretation is more relevant, we need more than only the R value. In this case, the variances of the ECD and the SSD are necessary to draw a better conclusion. If the variance of the ECD is greater than that of the SSD, all species have more similar sensitivities, but the exposure concentrations vary greatly between regions. In this case, scenario (ii) is the more correct interpretation as all species will be affected similarly, but only in 50% of the regions. Alternatively, if the variance of the SSD is greater than that of the ECD, all regions have more or less similar concentrations, but the species sensitivities vary greatly. In this case, scenario (i) is the more correct interpretation as only 50% of all species will be affected, but in all regions. Although we have the same R value for different scenarios, one scenario may have more serious consequences for the environment than the other. Scenario (i) can lead to lower overall biodiversity as 50% of species are affected in all regions. In Scenario (ii), although all species are affected in 50% of the regions, there are regions in which no species are affected. These interpretations can greatly influence the decisions risk managers need to make. To add to this, we are talking here about spatial variability in the exposure concentrations. The ECD could

also describe the temporal variability of concentrations. In that case, Scenario (i) might be the better environmental outcome of the two because then 50% of the species will be affected all of the time, but the other 50% will not.

The above example illustrates the importance of reporting the whole probabilistic risk assessment and not just a single value. This is also the strength of probabilistic risk assessment. Although you can report only a single value as in deterministic risk assessment, probabilistic risk assessment provides a more complete and transparent picture of the risk and the contributing variabilities and uncertainties should risk managers require these to make informed regulatory decisions.

For example, consider the CR distribution as reported in the CR bars in Chapter 3 Figure 3.4. We focus on the top bar taking into account exposure due to all forms of nanoTiO₂ particles. If risk managers want a single number for the CR, we can provide them with the value of 214.77, which is the 95th upper uncertainty bound of the 99th percentile of the variability distribution. This is typically a worst case value similar to what a deterministic risk assessment could have produced by making use of assessment factors to account for all the possible sources of variability and uncertainty. Were this a deterministic risk assessment, risk managers would be shocked by such a high value. There would be no way of assessing how realistic this CR of 214.77 is. Will all aquatic species now start dying? Must we put a halt to the production and use of nanoTiO₂? There is no way to determine how conservative this value is and if with more data and increasing knowledge we can reduce this value.

Now consider the probabilistic risk assessment again. Risk managers, as non-statisticians, only want to see a single number. So, again we provide them with 214.77. In the appendix of the risk assessment report, however, we provide the whole distribution as given in Chapter 3 Figure 3.4 with an explanation. Again risk managers will be shocked when seeing this large number. When referred to the appendix of the risk assessment report, however, they read the following explanation. The CR of 214.77 only applies to the most sensitive species which by accident finds itself in a region with an extreme high concentration of nanoTiO₂. Let's place this value in a more realistic context. A random species in a random region can have a CR anywhere between 3×10^{-6} and 214.77 with a best guess of between 2.45×10^{-5} and 23.8. The median CR is 0.03. Taking into account that the variance of the species sensitivity distribution is much larger than that of the

exposure distribution (see Chapter 3 Figure 3.7 bottom and top bars), we can deduce from the median that around 50% of the species in all regions will have a CR of up to 0.03. Going one step further, only between 5% and 25% of species in all regions have a CR greater than one (see Chapter 3 Figure 3.5). If we can reduce the uncertainty by obtaining more data in the form of toxicity studies and more knowledge on the processes determining the fate of nanoTiO₂ in the environment, we can narrow it down to approximately 11% of species having a CR greater than 1. This means that 11% (5%-25%) of species can potentially be adversely affected by nanoTiO₂. Based on this knowledge, risk managers may still find the risk too high and decide to reduce the production of nanoTiO₂, but they might not take the drastic decision to stop the production completely as they might have done had they only had the deterministic risk assessment at their disposal. Naturally, the probabilistic risk assessment interpretation is more technical in nature and takes more time to grasp. Nevertheless, it provides a more transparent picture and can possibly prevent unnecessarily stringent regulations which can hamper the development of nanotechnology.

6

In 2015, European Food Safety Authority (EFSA) brought out a draft guidance on how to characterise, document and explain all types of uncertainty arising in EFSA's scientific assessments (EFSA, 2015). I quote from the summary of the draft guidance:

In all types of assessment, the primary information on uncertainty needed by decision-makers is: what is the range of possible answers, and how likely are they? Assessors should also describe the nature and causes of the main sources of uncertainty, for use in communication with stakeholders and the public and to inform targeting of further work to reduce uncertainty, when needed. (EFSA, 2015)

This is very much in line with the message of this thesis. Although the scientific committee mentions that uncertainties can be expressed qualitatively or quantitatively, it is reassuring (as a statistician) to read that “assessors should always aim to express overall uncertainty in quantitative terms to the extent that is scientifically achievable”. In Figure S.2 of the draft guidance, they summarise the methods available for qualitative and quantitative uncertainty analysis. Most of the quantitative methods are also used or discussed in this thesis, for example, confidence intervals, Bayesian inference, bootstrap, 1D and 2D Monte Carlo, sensitivity analysis, Bayesian modelling and expert elicitation.

In terms of reporting, the draft guidance suggests a layered approach in which the probabilistic risk assessment results are reported in various levels of technicality. This layered approach confirms my discussion above in which I argue for a bottom up approach in which a full probabilistic risk assessment is performed and results are reported depending on the knowledge level of the reader.

6.5. CONCLUSION AND THE WAY FORWARD

In this thesis, I investigated the use of probabilistic methods in risk assessment of ENPs with special attention to the quantification of uncertainty and variability. I showed that probabilistic methods allow us to deal with uncertainty and variability in a transparent way. Although this is applicable to risk assessment of any chemical substance, it is especially relevant for ENPs. This is due to the fact that risk assessment of ENPs is surrounded by much uncertainty.

In this Chapter, I already mentioned the possibilities of using methods such as Bayesian 2DMC, Bayesian hierarchical modelling, modified profile likelihoods and strong matching priors. The use of such methods may present improvements on the efficiency and accuracy of the methods developed in this thesis. More research into these methods and their applicability to risk assessment of ENPs is a future research topic.

Throughout this thesis, I made the simplistic assumption that the environmental exposure distribution only quantifies the variability in exposure concentrations on a spatial scale. As already mentioned in Chapter 1 and in the previous section, there is also temporal variability in exposure concentrations. It would be interesting to investigate this further. Would quantifying temporal variability require an exposure distribution of its own? If yes, then one would have to choose which source of variability is the most important in the risk assessment under consideration and define the exposure distribution for either temporal or spatial variability. In reality, however, both sources of variability exist simultaneously and ideally, one would define an exposure distribution which quantifies both sources. This could be obtained by making use of data that is both spatially and temporally variable (Gottschalk et al., 2011; Sun et al., 2015). This idea could be further investigated taking special care of the time scale used in the species sensitivity data as it should be compatible with the temporal scale of the exposure distribution.

Methods such as 2DMC with corresponding uncertainty analysis should be implemented in the risk assessment of nanoparticles to a much larger extent than is currently the case. With these methods, we can identify the greatest sources of uncertainty. Based on such identification, research can be focused on those areas that need it most, thereby making large leaps in reducing the uncertainty that is currently hampering risk assessment of ENPs.

REFERENCES

- Aitchison, J. and Shen, S. (1980). Logistic-normal distributions : Some properties and uses. *Biometrika*, 67(2):261–272. DOI:10.1093/biomet/67.2.261.
- Aldenberg, T. and Jaworska, J. S. (2000). Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46:1–18. DOI:10.1006/eesa.1999.1869.
- Aldenberg, T., Jaworska, J. S., and Traas, T. P. (2002). Normal Species Sensitivity Distributions and probabilistic Ecological Risk Assessment. In *Species Sensitivity Distributions in Ecotoxicology*, pages 49–102. CRC Press, London.
- Aldenberg, T. and Slob, W. (1993). Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, 25(1):48–63. DOI:10.1006/eesa.1993.1006.
- Bain, L. J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*. Duxbury Press, Belmont, CA, second edition.
- Bamber, D. (1975). The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. *Journal of Mathematical Psychology*, 12:387–415. DOI:10.1016/0022-2496(75)90001-2.
- Barbiero, A. (2011). Confidence Intervals for Reliability of Stress-Strength Models in the Normal Case. *Communications in Statistics - Simulation and Computation*, 40(6):907–925. DOI:10.1080/03610918.2011.560728.
- Barlow, S., Chesson, A., Collins, J. D., Flynn, A., Hardy, A., Knaap, A., Kuiper, H., Larsen, J.-C., Lovell, D., Neindre, P. L., Schlatter, J., Silano, V., Skerfving, S., and Vannier, P. (2009). Guidance of the scientific committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. *The EFSA Journal*, 1150:1–72. DOI:10.2903/j.efsa.2009.1150.
- Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika Trust*, 70(2):343–365. DOI:10.1093/biomet/70.2.343.
- Batke, M., Escher, S., Hoffmann-Doerr, S., Melber, C., Messinger, H., and Mangelsdorf, I. (2011). Evaluation of time extrapolation factors based on the database RepDose. *Toxicology Letters*, 205(2):122–129. DOI:10.1016/j.toxlet.2011.05.1030.
- Baudrit, C. and Dubois, D. (2006). Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108. DOI:10.1016/j.csda.2006.02.009.

- Bekker, A., Roux, J., and Arashi, M. (2011). Wishart ratios with dependent structure: New members of the bimatrix beta type IV. *Linear Algebra and its Applications*, 435(12):3243–3260. DOI:10.1016/j.laa.2011.06.007.
- Bokkers, B., Bakker, M., Boon, P., Bos, P., Bosgra, S., van der Heijden, G., Janer, G., Slob, W., and van der Voet, H. (2009). The practicability of the integrated probabilistic risk assessment (IPRA) approach for substances in food. Technical report, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
- Bokkers, B. G. H. and Slob, W. (2005). A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. *Toxicological Sciences*, 85(2):1033–40. DOI:10.1093/toxsci/kfi144.
- Bos, P. M. J., Boon, P. E., van der Voet, H., Janer, G., Piersma, A. H., Brüschweiler, B. J., Nielsen, E., and Slob, W. (2009). A semi-quantitative model for risk appreciation and risk weighing. *Food and Chemical Toxicology*, 47(12):2941–2950. DOI:10.1016/j.fct.2009.03.009.
- Bosgra, S., van der Voet, H., Boon, P. E., and Slob, W. (2009). An integrated probabilistic framework for cumulative risk assessment of common mechanism chemicals in food: an example with organophosphorus pesticides. *Regulatory Toxicology and Pharmacology*, 54(2):124–33. DOI:10.1016/j.yrtph.2009.03.004.
- Bouwmeester, H., Dekkers, S., Noordam, M. Y., Hagens, W. I., Bulder, A. S., de Heer, C., ten Voorde, S. E. C. G., Wijnhoven, S. W. P., Marvin, H. J. P., and Sips, A. J. A. M. (2009). Review of health safety aspects of nanotechnologies in food production. *Regulatory Toxicology and Pharmacology*, 53:52–62. DOI:10.1016/j.yrtph.2008.10.008.
- Bowman, D. M. and Hodge, G. A. (2007). A small matter of regulation: an international review of nanotechnology regulation. *The Columbia Science and Technology Law Review*, 8:1–36.
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York.
- Boysen, E. (2015). UnderstandingNano. <http://www.understandingnano.com/>.
- Brandes, L., den Hollander, H. A., and van de Meent, D. (1996). Simplebox 2.0: a nested multimedia fate model for evaluating the environmental fate of chemicals. Technical report, RIVM Report no. 719101029, Bilthoven, The Netherlands.
- Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25:4279–4292. DOI:10.1002/sim.2673.
- Canty, A. and Ripley, B. (2013). boot: Bootstrap R (S-Plus) Functions.
- Cardwell, R. D., Parkhurst, B. R., Warren-Hicks, W., and Volosin, J. (1993). Aquatic ecological risk. *Water Environment and Technology*, 5:47–51.

- Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19(August 1999):1141–1164. DOI:10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.
- Carreau, J. and Bengio, Y. (2008). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76. DOI:10.1007/s10687-008-0068-0.
- Cervone, G., Franzese, P., Ezber, Y., and Boybeyi, Z. (2008). Risk assessment of atmospheric emissions using machine learning. *Natural Hazards and Earth System Science*, 8(5):991–1000. DOI:10.5194/nhess-8-991-2008.
- Chen, G. and Pham, T. (2001). *Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems*. CRC Press, Boca Raton.
- Chio, C.-P., Chen, W.-Y., Chou, W.-C., Hsieh, N.-H., Ling, M.-P., and Liao, C.-M. (2012). Assessing the potential risks to zebrafish posed by environmentally relevant copper and silver nanoparticles. *Science of the Total Environment*, 420:111–118. DOI:10.1016/j.scitotenv.2012.01.023.
- Church, J. and Harris, B. (1970). The Estimation of Reliability from Stress-Strength Relationships. *Technometrics*, 12(1):49–54. DOI:10.1080/00401706.1970.10488633.
- Ciffroy, P., Charlatchka, R., Ferreira, D., and Marang, L. (2013). Probabilistic approaches to accounting for data variability in the practical application of bioavailability in predicting aquatic risks from metals. *Integrated Environmental Assessment and Management*, 9(3):508–516. DOI:10.1002/ieam.1417.
- Cockburn, A., Bradford, R., Buck, N., Constable, A., Edwards, G., Haber, B., Hepburn, P., Howlett, J., Kampers, F., Klein, C., Radomski, M., Stamm, H., Wijnhoven, S., and Wildemann, T. (2012). Approaches to the safety assessment of engineered nanomaterials (ENM) in food. *Food and Chemical Toxicology*, 50:2224–2242. DOI:10.1016/j.fct.2011.12.029.
- Cohen, J. T., Lampson, M. a., and Bowers, T. S. (1996). The use of two-stage Monte Carlo simulation techniques to characterize variability and uncertainty in risk analysis. *Human and Ecological Risk Assessment: An International Journal*, 2(4):939–971. DOI:10.1080/10807039609383657.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- Coll, C., Notter, D., Gottschalk, E., Sun, T., Som, C., and Nowack, B. (2015). Probabilistic environmental risk assessment of five nanomaterials (nano-TiO₂, nano-Ag, nano-ZnO, CNT, and fullerenes). *Nanotoxicology*, pages 1–9. DOI:10.3109/17435390.2015.1073812.
- Cortese, G. (2015). *ProbYX: Inference for the Stress-Strength Model $R = P(Y < X)$* .
- Cortese, G. and Ventura, L. (2013). Accurate higher-order likelihood inference on $P(Y < X)$. *Computational Statistics*, 28(3):1035–1059. DOI:10.1007/s00180-012-0343-z.

- Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, 49(1):1–39.
- Darbra, R. M., Eljarrat, E., and Barceló, D. (2008). How to measure uncertainties in environmental risk assessment. *TrAC - Trends in Analytical Chemistry*, 27(4):377–385. DOI:10.1016/j.trac.2008.02.005.
- Datta, G. S. and Sweeting, T. J. (2005). Probability Matching Priors. *Handbook of Statistics*, 25(252):91–114. DOI:10.1016/S0169-7161(05)25003-4.
- Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.
- De Klein, J., Quik, J., Bauerlein, P., and Koelmans, A. (2016). Towards validation of the NanoDUFLOW nanoparticle fate model for the river Dommel, The Netherlands. *Environmental Science: Nano*, 3:434–441. DOI:10.1039/C5EN00270B.
- Dekkers, S., Bouwmeester, H., Bos, P. M., Peters, R. J., Rietveld, A. G., and Oomen, A. G. (2013). Knowledge gaps in risk assessment of nanosilica in food: evaluation of the dissolution and toxicity of different forms of silica. *Nanotoxicology*, 7(4):367–377. DOI:10.3109/17435390.2012.662250.
- Dekkers, S., Krystek, P., Peters, R. J., Lankveld, D. P., Bokkers, B. G., van Hoeven-Arentzen, P. H., Bouwmeester, H., and Oomen, A. G. (2011). Presence and risks of nanosilica in food products. *Nanotoxicology*, 5(3):393–405. DOI:10.3109/17435390.2010.519836.
- den Hollander, H., van Eijkeren, J., and van de Meent, D. (2004). SimpleBox 3.0: multimedia mass balance model for evaluating the fate of chemical in the environment. Technical report, RIVM Report No. 601200003, Bilthoven, The Netherlands.
- Derjaguin, B. V. and Landau, L. D. (1941). Theory of stability of strongly charged lyophobic sols and the adhesion of strongly charged particles in solutions of electrolytes. *Acta Physicochimica U.R.S.S.*, 14:633–662. DOI:10.1016/0079-6816(93)90013-L.
- Downtown, F. (1973). The Estimation of $Pr(Y > X)$ in the Normal Case. *Technometrics*, 15(3):551–558. DOI:10.1080/00401706.1973.10489081.
- EC (2011). COMMISSION RECOMMENDATION on the definition of nanomaterial L 275. *Official Journal of the European Union*, 54:38–40.
- ECETOC (1995). Assessment factors in human health risk assessment. Technical report, ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium).
- ECHA (2008a). Guidance on information requirements and chemical safety assessment Appendix R.7.13-2: Environmental risk assessment for metals and metal compounds. Technical report, European Chemicals Agency.
- ECHA (2008b). Guidance on information requirements and chemical safety assessment Chapter R.10: Characterisation of dose [concentration]-response for environment. Technical Report May, European Chemicals Agency.

- ECHA (2009). Guidance in a nutshell - Chemical Safety Assessment. Technical report, European Chemical Agency.
- ECHA (2011). Guidance on information requirements and chemical safety assessment Part B: Hazard assessment. Technical Report December, European Chemicals Agency.
- ECHA (2012a). Guidance on information requirements and chemical safety assessment Chapter R. 15 : Consumer exposure estimation. Technical Report October, European Chemicals Agency.
- ECHA (2012b). Guidance on information requirements and chemical safety assessment Chapter R.16: Environmental Exposure Estimation. Technical Report October, European Chemicals Agency.
- ECHA (2012c). Guidance on information requirements and chemical safety assessment Chapter R.19: Uncertainty analysis. Technical Report November, European Chemicals Agency.
- ECHA (2012d). Guidance on information requirements and chemical safety assessment Chapter R.8: Characterisation of dose [concentration]-response for human health. Technical Report November, European Chemicals Agency.
- ECHA (2012e). Guidance on information requirements and chemical safety assessment Part E: Risk Characterisation. Technical report, European Chemicals Agency.
- EC/JRC (2003). Technical Guidance Document on Risk Assessment. Technical report, European Chemicals Bureau, Ispra.
- ECOFRAM (1999). Ecological Committee on FIFRA risk Assessment Methods Aquatic Report, peer review draft. Technical Report 508, U.S. Environmental Protection Agency, Washington, D.C., USA.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- EFSA (2005). Opinion of the Scientific Committee on a request from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic (Request No EFSA-Q-2004-020). *The EFSA Journal*, 282:1–31. DOI:10.1021/bk-2010-1048.
- EFSA (2015). Draft Guidance on Uncertainty in EFSA Scientific Assessment. *The EFSA Journal*. DOI:10.2903/j.efsa.20YY.NNNN.
- ELC (2011). ELC's updated position on the basis of the recent EU regulatory developments on the definition and labelling of nano-ingredients. Technical Report ii, Federation of European Specialty Ingredient Industries, Brussels, Belgium.
- Enis, P. and Geisser, S. (1971). Estimation of the Probability that $Y < X$. *Journal of the American Dietetic Association*, 66(333):162–168. DOI:10.2307/2284867.
- European Commission (2013). Science Standards : a driver for for innovation. Technical report, European Commission.

- FAO/WHO (2009). Dose-response assessment and derivation of health-based guidance values. In *Principles and Methods for the Risk Assessment of Chemicals in Food*, chapter 5. A joint publication of the Food and Agriculture Organization of the United Nations and the World Health Organization, Geneva: World Health Organization (Environmental Health Criteria 240).
- Flari, V., Chaudhry, Q., Neslo, R., and Cooke, R. (2011). Expert judgment based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products. *Journal of Nanoparticle Research*, 13(5):1813–1831. DOI:10.1007/s11051-011-0335-x.
- Franco, A., Struijs, J., Gouin, T., and Price, O. R. (2013). Evolution of the sewage treatment plant model SimpleTreat: Applicability domain and data requirements. *Integrated Environmental Assessment and Management*, 9(4):569–579. DOI:10.1002/ieam.1414.
- Frey, H. C. (1992). Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making. Technical report, Fellowship Program for Environmental Science and Engineering, American Association for the Advancement of Science, Washington, DC.
- Frey, H. C. and Rhodes, D. S. (1996). Characterizing, simulating, and analyzing variability and uncertainty: An illustration of methods using an air toxics emissions example. *Human and Ecological Risk Assessment: An International Journal*, 2(4):762–797. DOI:10.1080/10807039609383650.
- Frey, H. C. and 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: An International Journal*, 4(2):423–468. DOI:10.1080/10807039891284406.
- Gaillardet, J., Viers, J., and Dupré, B. (2003). Trace elements in river waters. In Drever, J., Holland, H., and Turekian, K. K., editors, *Surface and Ground Water, Weathering, and Soils: Treatise on Geochemistry Vol. 5*, pages 225–227. Elsevier, Oxford.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. D., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press, Boca Raton, third edition.
- Gibbons, J. D. and Chakraborti, S. (2011). *Nonparametric Statistical Inference*. CRC Press, Boca Raton, 5th edition.
- Gilbertson, L. M., Wender, B. A., Zimmerman, J. B., and Eckelman, M. J. (2015). Coordinating modeling and experimental research of engineered nanomaterials to improve life cycle assessment studies. *Environmental Science: Nano*. DOI:10.1039/C5EN00097A.
- Gottschalk, F., Kost, E., and Nowack, B. (2013a). Engineered nanomaterials in water and soils: a risk quantification based on probabilistic exposure and effect modeling. *Environmental Toxicology and Chemistry / SETAC*, 32(6):1278–1287. DOI:10.1002/etc.2177.

- Gottschalk, F. and Nowack, B. (2012). A Probabilistic Method for Species Sensitivity Distributions Taking into Account the Inherent Uncertainty and Variability of Effects to Estimate Environmental Risk. *Integrated Environmental Assessment and Management*, 9(1):79–86. DOI:10.1002/ieam.1334.
- Gottschalk, F., Ort, C., Scholz, R., and Nowack, B. (2011). Engineered nanomaterials in rivers - Exposure scenarios for Switzerland at high spatial and temporal resolution. *Environmental Pollution*, 159(12):3439–3445. DOI:10.1016/j.envpol.2011.08.023.
- Gottschalk, F., Scholz, R., and Nowack, B. (2010a). Probabilistic material flow modeling for assessing the environmental exposure to compounds: Methodology and an application to engineered nano-TiO₂ particles. *Environmental Modelling and Software*, 25(3):320–332. DOI:10.1016/j.envsoft.2009.08.011.
- Gottschalk, F., Sonderer, T., Scholz, R., and Nowack, B. (2010b). Possibilities and limitations of modeling environmental exposure to engineered nanomaterials by probabilistic material flow analysis. *Environmental Toxicology and Chemistry / SETAC*, 29(5):1036–1048. DOI:10.1002/etc.135.
- Gottschalk, F., Sun, T., and Nowack, B. (2013b). Environmental concentrations of engineered nanomaterials: review of modeling and analytical studies. *Environmental Pollution*, 181:287–300. DOI:10.1016/j.envpol.2013.06.003.
- Govidarajulu, Z. (1967). Two-Sided confidence limits for $P(X < Y)$ based on normal samples of X and Y. *Sankhya: The Indian Journal of Statistics, Series B*, 29:35–40.
- Grieger, K. D., Hansen, S. F., and Baun, A. (2009). The known unknowns of nanomaterials: Describing and characterizing uncertainty within environmental, health and safety risks. *Nanotoxicology*, 3(3):222–233. DOI:10.1080/17435390902944069.
- Grieger, K. D., Linkov, I., Hansen, S. F., and Baun, A. (2012). Environmental risk analysis for nanomaterials: Review and evaluation of frameworks. *Nanotoxicology*, 6(2):196–212. DOI:10.3109/17435390.2011.569095.
- Hammes, J., Gallego-Urrea, J. a., and Hassellöv, M. (2013). Geographically distributed classification of surface water chemical parameters influencing fate and behavior of nanoparticles and colloid facilitated contaminant transport. *Water Research*, 47(14):5350–5361. DOI:10.1016/j.watres.2013.06.015.
- Handy, R., Owen, R., and Valsami-Jones, E. (2008). The ecotoxicology of nanoparticles and nanomaterials: Current status, knowledge gaps, challenges, and future needs. *Ecotoxicology*, 17(5):315–325. DOI:10.1007/s10646-008-0206-0.
- Handy, R. D., Cornelis, G., Fernandes, T., Tsyusko, O., Decho, A., Sabo-Attwood, T., Metcalfe, C., Steevens, J. A., Klaine, S. J., Koelmans, A. A., and Horne, N. (2012). Ecotoxicity test methods for engineered nanomaterials: Practical experiences and recommendations from the bench. *Environmental Toxicology and Chemistry / SETAC*, 31(1):15–31. DOI:10.1002/etc.706.

- Hansen, S. F. and Baun, A. (2012). European regulation affecting nanomaterials - review of limitations and future recommendations. *Dose-Response*, 10(3):364–383. DOI:10.2203/dose-response.10-029.Hansen.
- Harvey, J. and van der Merwe, A. (2014). Modelling occupational exposure using a random effects model: a bayesian approach. *South African Statistical Journal*, 48(1):61–71.
- Harvey, J. and van der Merwe, A. (2015). Analysis of unbalanced occupational exposure data using a bayesian random effect model. *South African Statistical Journal*, 49(2):169–185.
- Hayashi, T. I. and Kashiwagi, N. (2011). A Bayesian approach to probabilistic ecological risk assessment: Risk comparison of nine toxic substances in Tokyo surface waters. *Environmental Science and Pollution Research*, 18(3):365–375. DOI:10.1007/s11356-010-0380-5.
- Ho, C. K. (2008). Analytical risk-based model of gaseous and liquid-phase radon transport in landfills with radium sources. *Environmental Modelling and Software*, 23(9):1163–1170. DOI:10.1016/j.envsoft.2008.01.002.
- Hodge, G. and Bowman, D. (2004). Governing nanotechnology : setting the regulatory agenda. *The Journal of Contemporary Issues in Business and Government*, 10:18–33.
- Hollander, A., Hauck, M., Cousins, I. T., Huijbregts, M. A. J., Pistocchi, A., Ragas, A. M. J., and van de Meent, D. (2012). Assessing the Relative Importance of Spatial Variability in Emissions Versus Landscape Properties in Fate Models for Environmental Exposure Assessment of Chemicals. *Environmental Modeling and Assessment*, 17(6):577–587. DOI:10.1007/s10666-012-9315-5.
- Hosking, J. and Wallis, J. (1997). *Regional Frequency Analysis: An approach based on L-moments*. Cambridge University Press, Cambridge.
- Hosking, J., Wallis, J., and Wood, E. (1985). Estimation of the generalized value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261. DOI:10.1080/00401706.1985.10488049.
- Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):339–349. DOI:10.1080/00401706.1987.10488243.
- Jacobs, R., Bekker, A., and Human, S. (2012). On the Bivariate Kummer-Beta Type IV Distribution. *Communications in Statistics - Theory and Methods*, 41(18):3339–3354. DOI:10.1080/03610926.2011.560737.
- Jacobs, R., Bekker, A., van der Voet, H., and ter Braak, C. J. (2015a). Parametric estimation of $P(X > Y)$ for normal distributions in the context of probabilistic environmental risk assessment. *PeerJ*, 3:e1164. DOI:10.7717/peerj.1164.

- Jacobs, R., Meesters, J. A., ter Braak, C. J., van de Meent, D., and van der Voet, H. (2016). Combining exposure and effect modelling into an integrated probabilistic environmental risk assessment for nanoparticles. *Environmental Toxicology and Chemistry*, Accepted Author Manuscript. DOI:10.1002/etc.3476.
- Jacobs, R., van der Voet, H., and ter Braak, C. (2015b). Integrated probabilistic risk assessment for nanoparticles : the case of nanosilica in food. *Journal of Nanoparticle Research*, 17(6):251. DOI:10.1007/s11051-015-2911-y.
- Jager, T., den Hollander, H. A., van der Poel, P., Rikken, M. G. J., and Vermeire, T. (2001a). Probabilistic Environmental Risk Assessment for Dibutylphthalate (DBP). *Human and Ecological Risk Assessment: An International Journal*, 7(6):1681–1697. DOI:10.1080/20018091095302.
- Jager, T., Vermeire, T., Rikken, M. G. J., and van der Poel, P. (2001b). Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere*, 43(2):257–64. DOI:10.1016/S0045-6535(00)00087-4.
- JMPR (1999). Procedures for estimating acute reference dose. Pesticide residues in food - 1998. Technical report, Report of the JMPR 1998, FAO Plant Production and Protection Paper, 148, FAO, Rome.
- Johnson, A. C., Bowes, M. J., Crossley, A., Jarvie, H. P., Jurkschat, K., Jürgens, M. D., Lawlor, A. J., Park, B., Rowland, P., Spurgeon, D., Svendsen, C., Thompson, I. P., Barnes, R. J., Williams, R. J., and Xu, N. (2011). An assessment of the fate, behaviour and environmental risk associated with sunscreen TiO₂ nanoparticles in UK field scenarios. *Science of the Total Environment*, 409(13):2503–2510. DOI:10.1016/j.scitotenv.2011.03.040.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. John Wiley & Sons, New York, second edition.
- Kelly, E. J. and Campbell, K. (2000). Separating Variability and Uncertainty in Environmental Risk Assessment—Making Choices. *Human and Ecological Risk Assessment: An International Journal*, 6(1):1–13. DOI:10.1080/10807030091124419.
- Kennedy, M. and Hart, A. (2009). Bayesian modeling of measurement errors and pesticide concentration in dietary risk assessments. *Risk Analysis*, 29(10):1427–1442. DOI:10.1111/j.1539-6924.2009.01265.x.
- Kennedy, M. C., Roelofs, V. J., Anderson, C. W., and Salazar, J. D. (2011). A hierarchical Bayesian model for extreme pesticide residues. *Food and Chemical Toxicology*, 49(1):222–232. DOI:10.1016/j.fct.2010.10.020.
- Klaine, S. J., Koelmans, A. A., Horne, N., Carley, S., Handy, R. D., Kapustka, L., Nowack, B., and von der Kammer, F. (2012). Paradigms to assess the environmental impact of manufactured nanomaterials. *Environmental Toxicology and Chemistry / SETAC*, 31(1):3–14. DOI:10.1002/etc.733.

- Koelmans, A., Diepens, N., Velzeboer, I., Besseling, E., Quik, J., and van de Meent, D. (2015). Guidance for the prognostic risk assessment of nanomaterials in aquatic ecosystems. *Science of the Total Environment*, 535:141–149. DOI:10.1016/j.scitotenv.2015.02.032.
- Kooijman, S. (1987). A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Research*, 21(3):269–276. DOI:10.1016/0043-1354(87)90205-3.
- Krishnamoorthy, K. and Mathew, T. (2002). Assessing occupational exposure via the one-way random effects model with unbalanced data. *Journal of Statistical Planning and Inference*, 7(3):440–451. DOI:10.1198/108571102311.
- Krzanowski, W. J. and Hand, D. J. (2009). *ROC Curves for Continuous Data*. CRC Press, Boca Raton.
- Kundu, D. and Gupta, R. D. (2006). Estimation of $P[Y < X]$ for Weibull Distribution. *IEEE Transactions on Reliability*, 55(2):270–280. DOI:10.1109/TR.2006.874918.
- Lehman, A. and Fitzhugh, O. (1954). 100-Fold margin of safety. *Association of Food and Drug Quarterly Bulletin*, 18:33–35.
- Lesaffre, E. and Lawson, A. B. (2012). *Bayesian Biostatistics*. John Wiley & Sons, Chichester.
- Lester, R. R., Green, L. C., and Linkov, I. (2007). Site-specific applications of probabilistic health risk assessment: review of the literature since 2000. *Risk analysis*, 27(3):635–58. DOI:10.1111/j.1539-6924.2007.00890.x.
- Li, J. and Ma, S. (2011). Time-dependent ROC analysis under diverse censoring patterns. *Statistics in Medicine*, 30(11):1266–77. DOI:10.1002/sim.4178.
- Lindley, D. V. (1971). *Making Decisions*. Wiley, London.
- Lindley, D. V. (2006). *Understanding uncertainty*. John Wiley & Sons, Hoboken.
- Linkov, I., Satterstrom, F. K., Monica Jr, J. C., Foss Hansen, S., and Davis, T. A. (2009). Nano risk governance: current developments and future perspectives. *Nanotechnology Law & Business*, 6:203–220.
- Liu, H. H., Bilal, M., Cohen, Y., Lazareva, A., and Keller, A. (2014). Regional multimedia distribution of nanomaterials and associated exposures: A software platform. *Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014*, pages 10–17. DOI:10.1109/BIBM.2014.6999368.
- Lohman, K., Pai, P., Seigneur, C., Mitchell, D., Heim, K., Wandland, K., and Levin, L. (2000). A Probabilistic Analysis of Regional Mercury Impacts on Wildlife. *Human and Ecological Risk Assessment: An International Journal*, 6(1):103–130. DOI:10.1080/10807030091124464.
- Mackay, D. (2003). *Multimedia Environmental Models: The Fugacity Approach*. Lewis Publishers, London.

- Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30. DOI:10.1145/272991.272995.
- Maynard, A. D. and Kuempel, E. D. (2005). Airborne Nanostructured Particles and Occupational Health. *Journal of Nanoparticle Research*, 7(6):587–614. DOI:10.1007/s11051-005-6770-9.
- Mckone, T. E. and Eshpande, A. W. (2005). Can Fuzzy logic bring Complex Environmental Problems into Focus? *Environmental Science & Technology*, 39(2):42A–47A. DOI:10.1021/es0531632.
- MCRA (2013). MCRA 8.0 Reference Manual, Report Dec 2013. Technical report, WUR/Biometris, FERA and RIVM, available at <https://mcra.rivm.nl>.
- Meesters, J. A. J., Koelmans, A. A., Quik, J. T. K., and Hendriks, A. J. (2014). Multimedia Modeling of Engineered Nanoparticles with SimpleBox4nano: Model Definition and Evaluation. *Environmental Science & Technology*, 48(10):5726–5736. DOI:10.1021/es500548h.
- Meesters, J. A. J., Quik, J. T. K., Koelmans, A. A., Hendriks, A. J., and van de Meent, D. (2016). Multimedia Environmental Fate and Speciation of Engineered Nanoparticles: A probabilistic Modeling Approach. *Environmental Science: Nano*, (Accepted).
- Mendes, B. V. D. M. and Lopes, H. F. (2004). Data driven estimates for mixtures. *Computational Statistics and Data Analysis*, 47(3):583–598. DOI:10.1016/j.csda.2003.12.006.
- Moerbeek, M., Piersma, A. H., and Slob, W. (2004). A comparison of three methods for calculating confidence intervals for the benchmark dose. *Risk analysis*, 24(1):31–40. DOI:10.1111/j.0272-4332.2004.00409.x.
- Mokhlis, N. (2005). Reliability of a Stress-Strength Model with Burr Type III Distributions. *Communications in Statistics - Theory and Methods*, 34(7):1643–1657. DOI:10.1081/STA-200063183.
- Money, E., Reckhow, K., and Wiesner, M. (2012). The use of Bayesian networks for nanoparticle risk forecasting: Model formulation and baseline evaluation. *Science of the Total Environment*, 426:436–445. DOI:10.1016/j.scitotenv.2012.03.064.
- Money, E. S., Barton, L. E., Dawson, J., Reckhow, K. H., and Wiesner, M. R. (2014). Validation and sensitivity of the FINE Bayesian network for forecasting aquatic exposure to nano-silver. *Science of the Total Environment*, 473-474:685–91. DOI:10.1016/j.scitotenv.2013.12.100.
- Morgan, K. (2005). Development of a preliminary framework for informing the risk analysis and risk management of nanoparticles. *Risk Analysis*, 25(6):1621–1635. DOI:10.1111/j.1539-6924.2005.00681.x.
- Morgan, M. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge.

- Mueller, N. C. and Nowack, B. (2008). Exposure modeling of engineered nanoparticles in the environment. *Environmental Science & Technology*, 42(12):4447–4453. DOI:10.1021/es7029637.
- Müller, A. K., Bosgra, S., Boon, P. E., van der Voet, H., Nielsen, E., and Ladefoged, O. (2009). Probabilistic cumulative risk assessment of anti-androgenic pesticides in food. *Food and Chemical Toxicology*, 47(12):2951–2962. DOI:10.1016/j.fct.2009.07.039.
- Muri, S. D., van der Voet, H., Boon, P. E., van Klaveren, J. D., and Brüschweiler, B. J. (2009). Comparison of human health risks resulting from exposure to fungicides and mycotoxins via food. *Food and Chemical Toxicology*, 47(12):2963–2974. DOI:10.1016/j.fct.2009.03.035.
- Nadar, M., Kızılaslan, F., and Papadopoulos, A. (2014). Classical and Bayesian estimation of $P(Y < X)$ for Kumaraswamy's distribution. *Journal of Statistical Computation and Simulation*, 84(7):1505–1529. DOI:10.1080/00949655.2012.750658.
- Nandi, S. and Aich, A. (1996). Hypothesis-test for reliability in a stress-strength model, with prior information. *IEEE transactions on reliability*, 45(1):129–131.
- NanoNextNL (2016). Theme 1: Risk analysis and technology assessment. <http://www.nanonextnl.nl/themes/risk-analysis-and-technology-assessment/>.
- Nayak, T. K. and Kundu, S. (2001). Calculating and Describing Uncertainty in Risk Assessment: The Bayesian Approach. *Human and Ecological Risk Assessment: An International Journal*, 7(2):307–328. DOI:10.1080/20018091094385.
- Nowack, B., Baalousha, M., Bornhöft, N., Chaudhry, Q., Cornelis, G., Cotterill, J., Gondikas, A., Hassellöv, M., Lead, J., Mitrano, D. M., von der Kammer, F., and Wontner-Smith, T. (2015). Progress towards the validation of modeled environmental concentrations of engineered nanomaterials by analytical measurements. *Environ. Sci.: Nano*, 2:421–428. DOI:10.1039/C5EN00100E.
- OECD (1992). Report of the OECD workshop on the extrapolation of laboratory aquatic toxicity data to the real environment. Technical report, Organisation for Economic Co-operation and Development (OECD), OECD Environment Monographs No 59, Paris.
- OECD (2004). SIDS Dossier on Synthetic Amorphous Silica and Silicates. Technical report, UNEP.
- Pasupuleti, S., Alapati, S., Ganapathy, S., Anumolu, G., Pully, N. R., and Prakhya, B. M. (2012). Toxicity of zinc oxide nanoparticles through oral route. *Toxicology and Industrial Health*, 28(8):675–86. DOI:10.1177/0748233711420473.
- Peters, R., Kramer, E., Oomen, A. G., Herrera Rivera, Z. E., Oegema, G., Tromp, P. C., Fokkink, R., Rietveld, A., Marvin, H. J. P., Weigel, S., Peijnenburg, A. A. C. M., and Bouwmeester, H. (2012). Presence of Nano-Sized Silica during In Vitro Digestion of Foods Containing Silica as a Food Additive. *ACS NANO*, 6(3):2441–2451. DOI:10.1021/nn204728k.

- Peters, S. (2011). Feasibility and challenges of human health risk assessment for nanomaterials. http://www.safenano.org/media/69877/SNFeatureArticle_Peters2011-Humanhealthriska.
- Petersen, E. J., Diamond, S. A., Kennedy, A. J., Goss, G. G., Ho, K., Lead, J., Hanna, S. K., Hartmann, N. B., Hund-Rinke, K., Mader, B., Manier, N., Pandard, P., Salinas, E. R., and Sayre, P. (2015). Adapting OECD Aquatic Toxicity Tests for Use with Manufactured Nanomaterials: Key Issues and Consensus Recommendations. *Environmental Science & Technology*, 49(16):9532–9547. DOI:10.1021/acs.est.5b00997.
- Petersen, E. J., Henry, T. B., Zhao, J., MacCuspie, R. I., Kirschling, T. L., Dobrovolskaia, M. A., Hackley, V., Xing, B., and White, J. C. (2014). Identification and avoidance of potential artifacts and misinterpretations in nanomaterial ecotoxicity measurements. *Environmental Science and Technology*, 48(8):4226–4246. DOI:10.1021/es4052999.
- Pickands III, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, 6:7–10. DOI:10.1159/000323281.
- Praetorius, A., Arvidsson, R., Molander, S., and Scheringer, M. (2013). Facing complexity through informed simplifications: a research agenda for aquatic exposure assessment of nanoparticles. *Environmental Science: Processes & Impacts*, 15(1):161–168. DOI:10.1039/c2em30677h.
- Pronk, M., Wijnhoven, S., Bleeker, E., Heugens, E., Peijnenburg, W., Luttik, R., and Hakkert, B. (2009). Nanomaterials in REACH Nanosilver as a case study. Technical report, RIVM report, 601780003.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rasmussen, K., Mech, A., Mast, J., de Temmerman, P.-J., Waegeneers, N., Steen, F. V., Pizzolon, J. C., de Temmerman, L., van Doren, E., Jensen, A., Birkedal, R., Levin, M., Nielsen, H., Koponen, I. K., Axel, P., Kembouche, Y., Thieriet, N., Spalla, O., Giuot, C., Rousset, D., Bau, S., Bianchi, B., Gilliland, D., Pianella, E., Ceccone, G., Cotogno, G., Gibson, N., and Stamm, H. (2013). Synthetic Amorphous Silicon Dioxide (NM-200, NM-201, NM-202, NM-203, NM-204): Characterisation and Physico-Chemical Properties. Technical report, Publications Office of the European Union.
- Reid, N. (1996). Likelihood and higher-order approximations to tail areas: a review and annotated bibliography. *The Canadian Journal of Statistics*, 24(2):141–166. DOI:10.2307/3315622.
- Reid, N. and Fraser, D. a. S. (2003). Likelihood inference in the presence of nuisance parameters. *arXiv preprint physics/0312079*.
- Reiser, B. and Guttman, I. (1986). Statistical Inference for $Pr(Y < X)$: The Normal Case. *Technometrics*, 28(3):253–257. DOI:10.1080/00401706.1986.10488133.

- Rezaei, S., Tahmasbi, R., and Mahmoodi, M. (2010). Estimation of $P(Y < X)$ for generalized Pareto distribution. *Journal of Statistical Planning and Inference*, 140(2):480–494. DOI:10.1016/j.jspi.2009.07.024.
- Royal Commision on Enviromental Pollution (2008). Novel Materials in the Environment: The case of nanotechnology. Technical report, Royal Commision on Enviromental Pollution, London, UK.
- Sala, S., Migliorati, S., Monti, G., and Vighi, M. (2012). SSD-based rating system for the classification of pesticide risk on biodiversity. *Ecotoxicology*, 21(4):1050–1062. DOI:10.1007/s10646-012-0858-7.
- Sani-Kast, N., Scheringer, M., Slomberg, D., Labille, J., Praetorius, A., Ollivier, P., and Hungerbühler, K. (2015). Addressing the complexity of water chemistry in environmental fate modeling for engineered nanoparticles. *Science of the Total Environment*, 535:150–159. DOI:10.1016/j.scitotenv.2014.12.025.
- Savage, S. L. (2012). *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. John Wiley & Sons, Hoboken.
- Scarrott, C. J. and Hu, Y. (2015). evmix: Extreme Value Mixture Modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation.
- Schad, T. and Schulzz, R. (2011). Xplicit, a novel approach in probabilistic spatiotemporally explicit exposure and risk assessment for plant protection products. *Integrated Environmental Assessment and Management*, 7(4):612–623. DOI:10.1002/ieam.205.
- Schleier, J. J., Macedo, P. A., Davis, R. S., Shama, L. M., and Peterson, R. K. D. (2009). A two-dimensional probabilistic acute human-health risk assessment of insecticide exposure after adult mosquito management. *Stochastic Environmental Research and Risk Assessment*, 23(5):555–563. DOI:10.1007/s00477-008-0227-5.
- Schneider, K., Oltmanns, J., and Hassauer, M. (2004). Allometric principles for interspecies extrapolation in toxicological risk assessment—empirical investigations. *Regulatory Toxicology and Pharmacology*, 39(3):334–347. DOI:10.1016/j.yrtph.2004.03.001.
- Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford.
- Shegokar, R. (2015). Nanotoxicity: Must Consider Aspect of Nanoparticle Development. In Lungu, M., Neculae, A., Bunoiu, M., and Biris, C., editors, *Nanoparticles' Promises and Risks: Characterization, Manipulation, and Potential Hazards to Humanity and the Environment*, chapter 6, pages 87–102. Springer, New York.
- Simon, T. W. (1999). Two-Dimensional Monte Carlo Simulation and Beyond: A Comparison of Several Probabilistic Risk Assessment Methods Applied to a Superfund Site. *Human and Ecological Risk Assessment: An International Journal*, 5(4):823–843. DOI:10.1080/10807039.1999.9657762.

- Slob, W. (2002). Dose-response modeling of continuous endpoints. *Toxicological Sciences*, 66(2):298–312. DOI:10.1093/toxsci/66.2.298.
- Slob, W., Bakker, M. I., te Biesebeek, J. D., and Bokkers, B. G. H. (2014). Exploring the Uncertainties in Cancer Risk Assessment Using the Integrated Probabilistic Risk Assessment (IPRA) Approach. *Risk Analysis*, 34(8):1401–1422. DOI:10.1111/risa.12194.
- Slob, W. and Cotton, R. (2013). proast38.9: Benchmark Dose Modelling.
- Slob, W. and Pieters, M. N. (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. *Risk analysis*, 18(6):787–798. DOI:10.1111/j.1539-6924.1998.tb01121.x.
- Smith, R. (1985). Maximum likelihood estimation in a class of non regular cases. *Biometrika*, 72(1):67–90. DOI:10.1093/biomet/72.1.67.
- So, S. J., Jang, I. S., and Han, C. S. (2008). Effect of Micro/Nano Silica Particle Feeding for Mice. *Journal of Nanoscience and Nanotechnology*, 8(10):5367–5371. DOI:10.1166/jnn.2008.1347.
- Solecki, R., Davies, L., Dellarco, V., Dewhurst, I., Van Raaij, M., and Tritscher, A. (2005). Guidance on setting of acute reference dose (ARfD) for pesticides. *Food and Chemical Toxicology*, 43(11):1569–1593. DOI:10.1016/j.fct.2005.04.005.
- Solomon, K., Giesy, J., and Jones, P. (2000). Probabilistic risk assessment of agrochemicals in the environment. *Crop Protection*, 19(8-10):649–655. DOI:10.1016/S0261-2194(00)00086-7.
- Solomon, K. and Takacs, P. (2002). Probabilistic Risk Assessment Using Species Sensitivity Distributions. In *Species Sensitivity Distributions in Ecotoxicology*, pages 285–313. CRC Press, London.
- Stephen, C. E., Mount, D. I., Hansen, D. J., Gentile, J. R., Chapman, G. A., and Brungs, W. A. (1985). Guidelines for Deriving Numerical National Water Quality Criteria for the Protection Of Aquatic Organisms and Their Uses. Technical report, U.S. Environmental Protection Agency, National Technical Information Service, Springfield, VA, USA.
- Strambeanu, N., Demetrovici, L., Dragos, D., and Lungu, M. (2015). Nanoparticles: Definition, Classification and General Physical Properties. In Lungu, M., Neculae, A., Bunoiu, M., and Biris, C., editors, *Nanoparticles' Promises and Risks: Characterization, Manipulation, and Potential Hazards to Humanity and the Environment*, chapter 1, pages 3–8. Springer, New York.
- Sun, T. Y., Conroy, G., Donner, E., Hungerbühler, K., Lombi, E., and Nowack, B. (2015). Probabilistic modelling of engineered nanomaterial emissions to the environment: a spatio-temporal approach. *Environmental Science: Nano*, 2(4):340–351. DOI:10.1039/C5EN00004A.

- Sun, T. Y., Gottschalk, F., Hungerbühler, K., and Nowack, B. (2014). Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials. *Environmental Pollution*, 185:69–76. DOI:10.1016/j.envpol.2013.10.004.
- Suter, G. W., Vaughan, D. S., and Gardner, R. H. (1983). Risk Assessment By Analysis of Extrapolation Error: a Demonstration for Effects of Pollutants on Fish. *Environmental Toxicology and Chemistry / SETAC*, 2(3):369–378. DOI:10.1897/1552-8618(1983)2[369:RABAOE]2.0.CO;2.
- Tian, L. (2008). Confidence intervals for $P(Y_1 < X_2)$ with normal outcomes in linear models. *Statistics in Medicine*, 27:4221–4237. DOI:10.1002/sim.3290.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86. DOI:10.1080/01621459.1986.10478240.
- Tooze, J. A., Midthune, D., Dodd, K. W., Freedman, L. S., Krebs-Smith, S. M., Subar, A. F., Guenther, P. M., Carroll, R. J., and Kipnis, V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association*, 106(10):1575–1587. DOI:10.1016/j.jada.2006.07.003.
- Traas, T., van de Meent, D., Posthuma, L., Hamers, T., Kater, B., de Zwart, D., and Aldenberg, T. (2002). The Potentially Affected Fraction as a Measure of Ecological Risk. In *Species Sensitivity Distributions in Ecotoxicology*. CRC Press, London.
- Upton, G. and Cook, I. (2011). *Oxford Dictionary of Statistics*. Oxford University Press, Oxford, 2nd edition.
- van de Meent, D. (1993). SimpleBox: a generic multimedia fate evaluation model. Technical report, RIVM Report No. 222501002, Bilthoven, The Netherlands.
- van de Meent, D., Hollander, A., Peijnenburg, W., and Breure, T. (2011). Fate and transport of contaminants. In Sanchez-Bayo, F., van den Brink, P. J., and Mann, R. M., editors, *Ecological Impacts of Toxic Chemicals*, pages 13–42. Bentham Science Publishers, Oak Park.
- van der Voet, H., de Boer, W. J., Kruisselbrink, J. W., Goedhart, P. W., van der Heijden, G. W. A. M., Kennedy, M. C., Boon, P. E., and van Klaveren, J. D. (2015). The MCRA model for probabilistic single-compound and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79:5–12. DOI:10.1016/j.fct.2014.10.014.
- van der Voet, H. and Slob, W. (2007). Integration of probabilistic exposure assessment and probabilistic hazard characterization. *Risk Analysis*, 27(2):351–371. DOI:10.1111/j.1539-6924.2007.00887.x.
- van der Voet, H., van der Heijden, G. W., Bos, P. M., Bosgra, S., Boon, P. E., Muri, S. D., and Bruschweiler, B. J. (2009). A model for probabilistic health impact assessment of exposure to food chemicals. *Food and Chemical Toxicology*, 47(12):2926–2940. DOI:10.1016/j.fct.2008.12.027.

- van der Werf, H. M. and Zimmer, C. (1998). An indicator of pesticide environmental impact based on a fuzzy expert system. *Chemosphere*, 36(10):2225–2249. DOI:10.1016/S0045-6535(97)10194-1.
- van der Zande, M., Vandebriel, R. J., Groot, M. J., Kramer, E., Herrera Rivera, Z. E., Rasmussen, K., Ossenkoppele, J. S., Tromp, P., Gremmer, E. R., Peters, R. J. B., Hendriksen, P. J., Marvin, H. J. P., Hoogenboom, R. L. P., Peijnenburg, A. C. M., and Bouwmeester, H. (2014). Sub-chronic toxicity study in rats orally exposed to nanostructured silica. *Particle and Fibre Toxicology*, 11(8). DOI:10.1186/1743-8977-11-8.
- van Rossum, C. T., Fransen, H. P., Verkaik-Kloosterman, J., Buurma-Rethans, E. J., and Ocké, M. C. (2011). National Food Consumption Survey 2007–2010. Technical report, National Institute for Public Health and the Environment.
- van Straalen, N. (2002). Theory of Ecological Risk Assessment Based on Species Sensitivity Distributions. In *Species Sensitivity Distributions in Ecotoxicology*, pages 37–48. CRC Press, London.
- van Straalen, N. M. and Denneman, C. a. (1989). Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety*, 18(3):241–51. DOI:10.1016/0147-6513(89)90018-3.
- Ventura, L., Cabras, S., and Racugno, W. (2009). Prior Distributions From Pseudo-Likelihoods in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, 104(486):768–774. DOI:10.1198/jasa.2009.0133.
- Ventura, L. and Racugno, W. (2011). Recent advances on Bayesian inference for $P(X < Y)$. *Bayesian Analysis*, 6(3):411–428. DOI:10.1214/11-BA616.
- Verdonck, F., Sioen, I., Baert, K., Van Thuyne, N., Bilau, M., Matthys, C., De Henauw, S., De Meulenaer, B., Devlieghere, F., Van Camp, J., Vanrolleghem, P., Van Sprang, P., Verbeke, W., and Willems, J. (2006). Uncertainty and variability modelling of chemical exposure through food. *Archives of Public Health*, 64(4):159–173.
- Verdonck, F., Van Sprang, P., and Vanrolleghem, P. (2005). Uncertainty and precaution in European environmental risk assessment of chemicals. *Water Science & Technology*, 52(6):227–34.
- Verdonck, F. A. M., Aldenberg, T., Jaworska, J., and Vanrolleghem, P. A. (2003). Limitations of current risk characterization methods in probabilistic environmental risk assessment. *Environmental Toxicology and Chemistry / SETAC*, 22(9):2209–2213. DOI:10.1897/02-435.
- Vermeire, T., Stevenson, H., Peiters, M. N., Rennen, M., Slob, W., and Hakkert, B. C. (1999). Assessment factors for human health risk assessment: a discussion paper. *Critical Reviews in Toxicology*, 29(5):439–90. DOI:10.1080/10408449991349249.
- Verwey, E. J. W. and Overbeek, J. T. G. (1948). *Theory of the Stability of Lyophobic Colloids*. Elsevier, Amsterdam.

- Voinov, V. (1986). Unbiased estimation of $P(Y < X)$ in the normal case. *Journal of Soviet Mathematics*, 33(1):701–706.
- Voss, S., Charrondiere, U., Slimani, N., Kroke, A., Riboli, E., Wahrendorf, J., and Boeing, H. (1998). EPIC-SOFT ein europäisches Computerprogramm für 24-Stunden-Erinnerungsprotokolle. *Zeitschrift für Ernährungswissenschaft*, 37(3):227–233.
- Wagner, C. and Løkke, H. (1991). Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Research*, 25(10):1237–1242. DOI:10.1016/0043-1354(91)90062-U.
- Warren-Hicks, W., Parkhurst, B., and Butcher, J. (2002). Methodology for Aquatic Ecological Risk Assessment. In *Species Sensitivity Distributions in Ecotoxicology*. CRC Press, London.
- Weerahandi, S. and Johnson, A. (1992). Testing Reliability in a Stress-Strength Model when X and Y are Normally Distributed. *Technometrics*, 34(1):83–91. DOI:10.1080/00401706.1992.10485236.
- Westerhoff, P., Song, G., Hristovski, K., and Kiser, M. a. (2011). Occurrence and removal of titanium at full scale wastewater treatment plants: implications for TiO₂ nanomaterials. *Journal of Environmental Monitoring*, 13(5):1195–1203. DOI:10.1039/c1em10017c.
- WHO (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. Technical report, IPCS Harmonization Project Document 11.
- Wiesner, M. R., Lowry, G. V., Alvarez, P., Dionysiou, D., and Biswas, P. (2006). Assessing the Risks of Manufactured Nanomaterials. *Environmental Science & Technology*, 40:4336–4345. DOI:10.1021/es062726m.
- Xue, J., Zartarian, V. G., Özkaynak, H., Dang, W., Glen, G., Smith, L., and Stallings, C. (2006). A probabilistic arsenic exposure assessment for children who contact chromated copper arsenate (CCA)-treated playsets and decks, part 2: Sensitivity and uncertainty analyses. *Risk Analysis*, 26(2):533–541. DOI:10.1111/j.1539-6924.2006.00748.x.
- Zabell, S. L. (1989). The rule of succession. *Erkenntnis*, 31:283–321. DOI:10.1007/BF01236567.
- Zhang, K., Achari, G., and Li, H. (2009). A comparison of numerical solutions of partial differential equations with probabilistic and possibilistic parameters for the quantification of uncertainty in subsurface solute transport. *Journal of Contaminant Hydrology*, 110(1-2):45–59. DOI:10.1016/j.jconhyd.2009.08.005.

ACRONYM LIST

2DMC	two-dimensional Monte Carlo
ADI	acceptable daily intake
AF	assessment factor
AIC	Akaike Information Criterion
ARfD	acute reference dose
BCa	Bias corrected and accelerated
BMD	benchmark dose
cdf	cumulative distribution function
CEC	critical effect concentration
CR	concentration ratio
DNEL	derived no-effect level
DNFCS	Dutch National Food Consumption Survey
EC10	10% effect concentration
EC20	20% effect concentration
EC25	25% effect concentration
EC50	50% effect concentration
ECD	exposure concentration distribution
ECHA	European Chemical Agency
EFSA	European Food Safety Authority
ENP	engineered nanoparticle
ER	ecological risk
ExpC	exposure concentration
gpd	generalised Pareto distribution
HC5	5% hazard concentration
HONEC	highest observed no effect concentration

HPD highest posterior density
IBMD Individual Benchmark Dose
IDEXP Individual-day Exposure
IEXP Individual Exposure
IMoE Individual Margin of Exposure
IPRA Integrated Probabilistic Risk Assessment
JPC joint probability curve
LC10 10% lethal concentration
LC20 20% lethal concentration
LC25 25% lethal concentration
LC50 50% lethal concentration
LED lowest effective dose
LNN LogisticNormal-Normal
LOAEL Lowest Observed Adverse Effect Level
LOEC lowest observed effect concentration
MC Monte Carlo
MCMC Markov Chain Monte Carlo
ml maximum likelihood
MLE maximum likelihood estimator
MoE margin of exposure
MoS margin of safety
NOAEL no observed adverse effect level
NOEC no observed effect concentration
pdf probability density function
PEC predicted environmental concentration
PMF probabilistic material flow
PNEC predicted no-effect concentration
pSSD probabilistic species sensitivity distribution
QMLE quasi maximum likelihood estimator
RCR risk characterisation ratio

REACH Registration, Evaluation, Authorisation and Restriction of Chemicals

ROC receiver operating characteristic

SAS Synthetic Amorphous Silica

SB4N SimpleBox4Nano

SSD species sensitivity distribution

T25 25% tumor chronic dose rate

WHO World Health Organization

SUMMARY

Engineered nanoparticles (ENPs) are used everywhere and have large technological and economic potential. Like all novel materials, however, ENPs have no history of safe use. Insight into risks of nanotechnology and the use of nanoparticles is an essential condition for the societal acceptance and safe use of nanotechnology.

Risk assessment of ENPs has been hampered by lack of knowledge about ENPs, their environmental fate, toxicity, experimental testing considerations, characterisation of nanoparticles and human and environmental exposure routes. This lack of knowledge results in uncertainty in the risk assessment. Moreover, due to the novelty of nanotechnology, risk assessors are often confronted with small samples of data on which to perform a risk assessment. Dealing with this uncertainty and the small sample sizes are major challenges when it comes to risk assessment of ENPs. In this context, the main research question of this thesis was: How can statistics assist risk assessors in dealing with the large uncertainty and small sample sizes in the risk assessment of nanoparticles?

In this thesis, I contributed to the methodology needed to understand potential environmental and human risk of ENPs in the presence of uncertainty. I approached the risk assessment of ENPs with probabilistic methods. Besides the large uncertainty, another form of variation present in risk assessment is variability. Variability is the natural inherent variation that is present in all natural processes and living organisms and is not reducible. Uncertainty, on the other hand, is the, in principle, reducible variation that exists due to lack of data and knowledge. The separating and quantifying of variability and uncertainty constituted a central part of this thesis.

The research in this thesis aimed to address the research question by means of two objectives: (i) to perform a transparent risk assessment of nanoparticles in the face of large uncertainty in such a way that it can guide future research to reduce the uncertainty and (ii) to evaluate empirical and parametric methods to estimate the risk probability in the case of small sample sizes.

In **Chapter 2**, I adapted an existing Integrated Probabilistic Risk Assessment (IPRA) method developed for human health effects due to chemicals and applied it to the case of nanosilica in food. In IPRA, statistical distributions and bootstrap methods are used to quantify uncertainty and variability in the risk assessment in a two-dimensional Monte Carlo algorithm. I found that a probabilistic approach confirmed that the results from the deterministic study are worst-case, and allowed to quantify by how much. Moreover, the method allowed the identification of uncertainty sources that contributed most to the total uncertainty in the risk assessment.

In **Chapter 3**, I developed a similar IPRA method for environmental risk assessment of ENPs. I illustrated the use of the method using a simplified aquatic risk assessment of nanoTiO₂ and showed that IPRA leads to a more transparent risk assessment and can also direct further environmental and toxicological research to the areas in which it is most needed.

Estimating the risk probability, $R = P(ExpC > CEC)$ (*ExpC*: exposure concentration, *CEC*: critical effect concentration), is a problem when confronted by potentially small risks and small sample sizes of the exposure concentration and/or the effect concentration. A nonparametric estimator based on data alone is not sufficient as it is limited by sample size. In **Chapter 4**, I investigated the maximum gain possible when making strong parametric assumptions as opposed to making no parametric assumptions at all. I compared maximum likelihood and Bayesian estimators with the non-parametric estimator and studied the influence of sample size and risk on the (interval) estimators via simulation. I found that, compared to the non-parametric estimators, the parametric estimators enable us to better estimate and bound the risk when sample sizes and/or risks are small. Also, the Bayesian estimator outperformed the maximum likelihood estimators in terms of coverage and interval lengths and was, therefore, preferred in our motivating case study of aquatic risk assessment of nanoAg.

Although the normal distribution is customarily the first choice when moving from empirical to parametric estimation, its tails are often found to be too thin. In **Chapter 5**, I allow for thicker tails by using the generalised Pareto distribution to estimate the tails of the exposure and species sensitivity distributions. I developed a mixture model to estimate the risk probability, $R = P(ExpC > CEC)$, with the assumption of a normal distribution for the bulk data and generalised Pareto distributions for the tails. A sensitivity

analysis showed significant influence of the tail heaviness on the risk probability, R , especially for low risks. I also studied the effect of small sample sizes on the estimation of the tail index and illustrated the proposed methods on a real data set.

In the General Discussion, **Chapter 6**, I discussed the results obtained in the preceding chapters and some possible adjustments and extensions to the methods developed in this thesis. To really be able to focus the research into the risks of ENPs to the most needed areas, probabilistic methods as used and developed in this thesis need to be implemented on a larger scale. With these methods, it is possible to identify the greatest sources of uncertainty. Based on such identification, research can be focused on those areas that need it most, thereby making large leaps in reducing the uncertainty that is currently hampering risk assessment of ENPs.

SAMENVATTING

Nanotechnologie is een betrekkelijk nieuw, maar snelgroeiend gebied. Door de mens gemaakte nanodeeltjes worden overal gebruikt. Ze bieden veel interessante technologische en economische mogelijkheden. Net als bij alle nieuwe materialen hebben wij nog geen ervaring wat betreft gebruikersveiligheid. Dat maakt het schatten van de risico's lastig. Inzicht in de risico's van nanotechnologie en het gebruik van nanodeeltjes is een essentiële voorwaarde voor het bepalen van de gebruikersveiligheid en daarmee de maatschappelijke aanvaarding van nanotechnologie.

Er zijn twee belangrijke redenen waarom risicobeoordeling van nanodeeltjes moeilijk is. De eerste is gebrek aan kennis over nanodeeltjes. Over welke deeltjes hebben we het precies, hoe gedragen ze zich, wat is hun verspreiding in het milieu, zijn ze giftig en zo ja, hoe giftig en voor welke organismen (waaronder de mens), hoe komen die organismen met de deeltjes in aanraking? In veel gevallen is ook de vraag hoe je dat kan meten. Dit gebrek aan kennis geeft onzekerheid in de risicobeoordeling. De tweede reden is dat risicobeoordelaars vaak met kleine steekproeven moeten werken om een risicobeoordeling op uit te voeren. Weinig gegevens betekent een extra marge in de risicoschatting. De belangrijkste vraag is hoe je, gegeven deze onzekerheid en de kleine steekproeven, toch een zo goed mogelijke risicobeoordeling van nanodeeltjes kunt geven. Dat is precies de onderzoeksvraag van dit proefschrift: hoe kan statistiek risicobeoordelaars helpen in het omgaan met grote onzekerheid en kleine steekproeven in de risicobeoordeling van nanodeeltjes?

In dit proefschrift lever ik een bijdrage aan de methodologie die nodig is om potentiële risico's van nanodeeltjes voor mens en milieu te begrijpen, gegeven de onzekerheid. Ik benader de risicobeoordeling van nanodeeltjes met probabilistische methoden. Het gebrek aan kennis zorgt voor variatie in de data, bijvoorbeeld omdat we de metingen niet helemaal goed doen. Naast de grote onzekerheid is er een tweede vorm van variatie aanwezig bij de risicobeoordeling, namelijk variabiliteit. Variabiliteit is de van nature

aanwezige variatie in alle natuurlijke processen en levende organismen. Een mens bijvoorbeeld reageert heel anders op allerlei stoffen dan een bacterie. Deze variatie is niet reduceerbaar, zo is de natuur nu eenmaal. Onzekerheid, daarentegen, is de (in principe) reduceerbare variatie die ontstaat door gebrek aan gegevens en kennis. In onderzoek is niet altijd duidelijk met welke variatie we te maken hebben en waar dus verbetering mogelijk is, en hoeveel. Het onderscheiden en kwantificeren van variabiliteit en onzekerheid vormt een centraal onderdeel van dit proefschrift.

Het onderzoek in dit proefschrift beantwoordt de onderzoeksvraag op twee manieren: (i) door het uitvoeren van een transparante risicobeoordeling van nanodeeltjes, in aanwezigheid van grote onzekerheid, om zo toekomstig onderzoek een handvat te geven om de onzekerheid te reduceren en (ii) door het evalueren van empirische en parametrische methoden voor het schatten van de kans op risico bij kleine steekproeven.

In **Hoofdstuk 2**, pas ik de bestaande Integrated Probabilistic Risk Assessment (IPRA) methode, die ontwikkeld is voor humane gezondheidseffecten van chemische stoffen, aan voor toepassing op nanosilica in voedsel. In IPRA worden statistische verdelingen en bootstrap methoden gebruikt om de onzekerheid en de variabiliteit in de risicobeoordeling te kwantificeren door middel van een tweedimensionaal Monte Carlo algoritme. De probabilistische aanpak bevestigt niet alleen dat de deterministische studie een worst-case resultaat geeft, maar ook in hoeverre dat resultaat te verbeteren valt. Bovendien geeft deze methode de mogelijkheid te bepalen welke bronnen van onzekerheid het meest aan de totale onzekerheid in de risicobeoordeling bijdragen. Door onderzoek te doen aan juist die bronnen kan forse winst worden geboekt voor verbetering van die risicobeoordeling.

In **Hoofdstuk 3**, ontwikkel ik een soortgelijke IPRA methode voor de beoordeling van de milieurisico's van nanodeeltjes. Ik illustreer het gebruik van de methode met een vereenvoudigde risicobeoordeling van titaniumdioxide (TiO_2) nanodeeltjes in water. TiO_2 nanodeeltjes komen voor in allerlei gebruiksartikelen, van pigment in verf tot UV blokker in zonnebrandcrème. Daarbij toon ik aan dat de IPRA methode die ik heb ontwikkeld tot een transparante risicobeoordeling leidt en bovendien kan aangeven op welk gebied verder milieutechnisch en toxicologisch onderzoek het hardst nodig is om die te verbeteren.

Een belangrijk onderdeel bij de risicobeoordeling van een stof is de risicokans, de kans dat iets of iemand zoveel van een stof binnenkrijgt dat er schade ontstaat, de kans (R) dat de blootstellingsconcentratie ($ExpC$) boven de kritieke effectconcentratie (CEC) ligt. Bovendien kunnen die grootheden voor verschillende organismen onder verschillende omstandigheden heel verschillend zijn. Het schatten van die kans is een probleem als we geconfronteerd worden met mogelijke lage risico's en kleine steekproefgroottes. Een niet-parametrische schatter die alleen op de data gebaseerd is, is niet toereikend omdat die beperkt wordt door de steekproefgrootte. In **Hoofdstuk 4**, onderzoek ik de maximale winst van sterke parametrische aannames tegenover helemaal geen parametrische aannames. Via simulaties vergelijk ik maximum-likelihood en Bayesiaanse schatters met de niet-parametrische schatter en bestudeer ik de invloed van de steekproefgrootte en lage risico's op de (interval) schatters. Ik toon aan dat, in vergelijking met de niet-parametrische schatter, de parametrische schatters de risicokans beter kunnen schatten en begrenzen voor kleine steekproeven en/of lage risico's. Verder presteert de Bayesiaanse schatter beter dan de maximum-likelihood-schatters wat betreft interval-dekking en -lengte. De Bayesiaanse schatter heeft daarom de voorkeur in de motiverende casus van een milieu risicobeoordeling van zilver nanodeeltjes in water. Zilver nanodeeltjes worden veel gebruikt in experimenten, maar ook voor antibacteriële toepassingen in allerlei gebruiksartikelen en kunnen daardoor desastreus zijn voor organismen in het milieu.

Naarmate de afwijking van het gemiddelde groter is, neemt de kans daarop af, dat heet de staart van de verdeling. Hoewel de normale verdeling doorgaans de eerste keuze is bij de overstap van empirische naar parametrische schatting, wordt de staart van die verdeling vaak te dun bevonden. De kans op een afwijking neemt sneller af dan reëel wordt gevonden. In **Hoofdstuk 5** laat ik dikkere staarten toe, door gebruik te maken van de gegeneraliseerde Pareto verdeling om de staarten van de blootstellings- en effectverdeling te schatten. Ik ontwikkel een mixture model voor het schatten van de risicokans, $R = P(ExpC > CEC)$, met de aanname van normaliteit voor de bulkdata en de gegeneraliseerde Pareto verdeling voor de staarten. Een gevoeligheidsanalyse toont significante invloed van de dikte van de staart op de kans, R , vooral voor lage risico's. Ik bestudeer ook het effect van kleine steekproeven op het schatten van de staartindex en

illustreer de voorgestelde methoden met een echte dataset. De conclusie is dat zeker voor lage risico's er met de dikte van de staart rekening gehouden moet worden.

In de algemene discussie, **Hoofdstuk 6**, bespreek ik de resultaten van de voorafgaande hoofdstukken en enkele mogelijke aanpassingen en uitbreidingen van de in dit proefschrift ontwikkelde methodes. Om daadwerkelijk aan te kunnen geven op welke gebieden onderzoek naar de risico's van nanodeeltjes het meest nodig is, moeten de methoden die in dit proefschrift ontwikkeld zijn op een grotere schaal worden toegepast. Met deze methoden is het mogelijk om de belangrijkste bronnen van onzekerheid te identificeren. Op basis van deze identificatie kan onderzoek beter worden gericht op de gebieden die dat het meest nodig hebben, waardoor grote vooruitgang gemaakt kan worden in het reduceren van de onzekerheid die momenteel de risicobeoordeling van nanodeeltjes belemmert.

ABOUT THE AUTHOR



Rianne Jacobs was born on 4 November 1983 in Amstelveen, The Netherlands. In 2009, Rianne obtained her BSc (Honours) degree in Mathematical Statistics with distinction from the University of Pretoria, South Africa. Her bachelor thesis was in statistical process control in which she studied the design of attribute control charts. Rianne subsequently obtained her MSc degree in Mathematical Statistics with distinction in 2011 at the same university. Her master thesis was in statistical distribution theory in which she presented the product method for finding bivariate Kummer beta distributions. During her master studies, Rianne held a lecturing position at the University of Pretoria where she lectured first year statistics students.

In 2012, Rianne obtained a PhD position at the Mathematical and Statistical Methods Group (Biometris), Wageningen University under the supervision of Cajo ter Braak and Hilko van der Voet. Her research on developing and studying various probabilistic methods for risk assessment of nanoparticles was part of NanoNextNL, a micro and nanotechnology consortium of the Government of the Netherlands and 130 partners. During her PhD, Rianne presented her research at various international conferences and organised a special session on statistics in nanoparticle risk assessment at the 57th South African Statistical Association Conference in 2015.

Currently, Rianne is a postdoctoral researcher at the National Institute for Public Health and the Environment (RIVM). She is developing statistical methods for the identification and tracing of contaminated food products during food-borne disease outbreaks.

LIST OF PUBLICATIONS

Jacobs, R., Meesters, J.A.J., ter Braak, C.J.F., van de Meent, D., van der Voet, H. (2016) Combining exposure and effect modelling into an integrated probabilistic environmental risk assessment for nanoparticles. *Environmental Toxicology and Chemistry*, Accepted Author Manuscript. DOI:10.1002/etc.3476.

Jacobs, R., van der Voet, H., ter Braak, C.J.F. (2015) Integrated probabilistic risk assessment for nanoparticles: the case of nanosilica in food. *Journal of Nanoparticle Research*, **17**(6):251. DOI:10.1007/s11051-015-2911-y.

Jacobs, R., Bekker, A., van der Voet, H., ter Braak, C.J.F. (2015) Parametric estimation of $P(X > Y)$ for normal distributions in the context of probabilistic environmental risk assessment. *PeerJ*, **3**:e1164. DOI:10.7717/peerj.1164.

Jacobs, R., Bekker, A., Human, S. (2012) On the Bivariate Kummer-Beta Type IV Distribution. *Communications in Statistics - Theory and Methods*, **41**(18):3339-3354. DOI:10.1080/03610926.2011.560737.

Jacobs, R., Bekker, A., Human, S. (2011) 'n Nuwe ontwikkeling tot die bivariate betaveld. *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie*, **30**(1), Art. #103, 2 pages. DOI:10.4102/satnt.v30i1.103.

DANKWOORD

Dit proefschrift was er niet geweest zonder de hulp, ondersteuning en motivatie van heel veel bijzondere mensen. Ik wil daarom ook mijn welgemeende dank betuigen aan iedereen die mij tot hulp en steun is geweest tijdens deze lange, doch zeer dankbare, promotie reis.

Allereerst mijn begeleiders, Cajo en Hilko: hartelijk dank voor jullie zeer kundige begeleiding. In het begin kwam ik wat moeizaam op gang, maar met jullie begeleiding en steun heb ik me kunnen ontwikkelen als een onafhankelijke onderzoeker. Cajo, met jouw ideeën en inzichten werd ik gedwongen tot dieper nadenken. We gingen vaak zijweggetjes in die soms doodliepen, maar andere keren weer tot verrassende inzichten en interessante onderzoeksvragen leidde. Hilko, jij was altijd de relativerende factor. Jij zorgde ervoor dat onze besprekingen niet in de lucht bleven hangen, maar concreet en “to the point” bleven. Ook jouw expertise en kennis op het gebied van risicobeoordeling was onmisbaar. Aan het lijstje begeleiders wil ik ook graag Andriëtte Bekker toevoegen: Prof, baie dankie vir prof se aanhoudende motivasie en ondersteuning. Prof se entoesiasme vir statistiek was aansteeklik: vanaf my tweede jaar, toe ek verlief geraak het op statistiek in prof se verskriklike teoretiese statistiek klas tot aan my MSc verhandeling waar my liefde vir navorsing ontstaan het. En selfs toe ek na die verre noorde vertrek, was prof altyd daar om te ondersteun en te bemoedig. Baie dankie.

Dan mijn collega's van Biometris. Hartelijk dank voor een onvergetelijke tijd in Wageningen. Mede door jullie voelde ik mij al gauw thuis in Nederland en kon ik maar moeilijk los komen van Biometris, zelfs na het beginnen aan mijn nieuwe baan bij het RIVM. Joost, bedankt for jouw hulp bij het schrijven van mijn Nederlandse samenvatting. To all my dear PhD colleagues, past and present: without you, my PhD journey would have been much harder. George, thank you for your support and interest in my research. Due to our similar statistical background, we could always discuss our research with each other. Thank you also for you invaluable comments and suggestions on my thesis. Daniela,

thank you for your help and support during the last phase of my PhD journey including the proofreading and the preparations for my defence. Then Nadia, Vincent, Martin, Jeroen, Cassandra, Amber, Sanne dS, Tryntsje, Rianne vB, Guus, Julio, Ricky, Dominique, Gavin, and all the others who have come and gone, thank you for being such fun and inspiring colleagues. Dominique and Gavin, you will now have to keep the South African legend going.

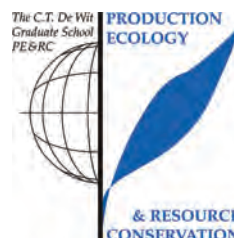
Mijn vrienden en geestelijke familie, hartelijk dank voor jullie vriendschap. Thea, 'n baie groot dankie aan jou vir die pragtige voorblad. Jy het dit perfek reggekry om my idee vorm te gee. Ek skuld jou vir al jou tyd en moeite. Oom Wouter en tante Frieda, ik was als kind aan huis altijd welkom bij u thuis. Met name de eerste periode dat ik in Nederland was, was u voor mij een thuishaven ver van huis. Dank u wel voor uw ondersteuning en het altijd voor mij klaar staan. Frederik en Joyce (en lieve Hanna, natuurlijk), jullie huis stond ook altijd open voor mij. Ik kon bij jullie altijd even tot rust komen als het werk even te veel werd. Joyce, het samen koken, bakken, winkelen en gewoon op de bank zitten kletsen, gaf mij altijd weer moed en kracht om verder te gaan. Dank jullie wel. En dan al de jonge meiden van de gemeente: Femke, Gerry, Petra, Esther, Desi, Hanny, Diane, Anna Laura en Hanna. Bedankt voor jullie vriendschap en de onvergetelijke ritjes naar Duitsland, Zwitserland en Frankrijk. Mogen er nog vele komen. Iedereen van de gemeenten in Well, Waddinxveen en Middelstum, dank jullie wel voor jullie vriendschap. Then the Eurochor and uncle Daniel, thank you all for being a spiritual family to me. Our practice weekends and concert trips were an inspiration for me and gave me renewed spiritual strength to carry me through the hard times. Vielen Dank! Merci beaucoup!

En dan, lieve mammi en pappie. Het was voor jullie niet makkelijk toen ik jullie achter liet om naar Nederland te komen. Maar jullie hebben mij tot op heden altijd onvoorwaardelijk gesteund. Zonder die steun hadden jullie nu niet dit boekje in jullie handen gehad. Dank jullie wel voor jullie gebeden, motivatie en steun die ik tijdens deze promotie reis heb mogen ervaren. Mijn lieve zusjes en broertje, Inge, Esther en Jos, ook jullie bedankt voor jullie gebeden en steun. Zonder een lief en warm gezin om mij heen, was deze reis veel moeilijker geweest. Inge, ik vond het ook heel bijzonder om jou als paranimf te hebben. Dank je wel voor je steun.

Als laatste gaat mijn eeuwige dank uit naar mijn hemelse Vader. Heere, zonder Uw aanwezigheid en leiding zou deze promotie reis een vergeefse onderneming zijn geweest.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review of literature (4.5 ECTS)

- Environmental nano risk assessment: a literature review (2013)

Writing of project proposal (4.5 ECTS)

- Nano risk assessment: a statistical approach (2012)

Post-graduate courses (7.8 ECTS)

- Introduction to R; PE&RC (2012)
- General toxicology; P.E.T. (2013)
- Risk assessment; SENSE (2013)
- RATA Course; NanoNextNL (2013)
- Zero inflated models & GLMM with R; PE&RC (2014)
- Intellectual property and valorisation; NanoNextNL (2015)

Laboratory training and working visits (4.5 ECTS)

- Theoretical approach to deriving risk distribution; University of Pretoria, South Africa (2013)
- Bayesian estimation of $P(X > Y)$ in the context of risk assessment; University of Pretoria, South Africa (2014)

Invited review of (unpublished) journal manuscript (2 ECTS)

- Communication in Statistics - Theory and Methods: distribution theory (2014)
- Trends in Food Science and Technology: nanoparticle risk assessment in food (2014)

Competence strengthening / skills courses (7.6 ECTS)

- Competence assessment: WGS (2012)
- Information literacy and EndNote; WGS (2012)
- Interpersonal communication for PhDs; WGS (2012)

- Effective behaviour in your professional surroundings; WGD (2012)
- Techniques for writing and presenting a scientific paper; WGS (2013)
- Project and time management; WGS (2013)
- Analytical storytelling; NanoNextNL (2014)
- Writing grant proposals; WGS (2015)
- Reviewing a scientific paper; WGS (2015)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.7 ECTS)

- PE&RC First year weekend (2012)
- PE&RC Midterm weekend (2014)
- PE&RC Day (2014, 2015)
- PE&RC Last year weekend (2015)

Discussion groups / local seminars / other scientific meetings (8.5 ECTS)

- MSN Discussion group (2012-2015)
- SENSE Symposium; oral presentation (2013)
- Wageningen PhD symposium; oral presentation (2013)
- NanoCity; poster presentation; Utrecht (2014)
- NanoCity; poster presentation; Amersfoort (2015)

International symposia, workshops and conferences (9 ECTS)

- 8th World Congress in Probability and Statistics; oral presentation; Turkey (2012)
- Seminar lecture at Department of Statistics; University of Pretoria, South Africa (2012, 2014)
- IBS Channel Network Conference; oral presentation; Nijmegen (2015)
- European Meeting of Statisticians; oral presentation; Amsterdam (2015)
- 57th South African Statistical Association Conference; oral presentation; Pretoria, South Africa (2015)

Lecturing / supervision of practicals / Tutorials (6.75 ECTS)

- Statistiek 1 (2012-2014)
- Statistiek 2 (2013-2015)

This thesis is part of NanoNextNL, a micro and nanotechnology innovation consortium of the Government of the Netherlands and 130 partners from academia and industry. More information on www.nanonextnl.nl.

Cover design: Thea Wentzel and Rianne Jacobs

Thesis layout: Based on latex style dissertation.cls from TU Delft

Printed by: GVO drukkers & vormgevers B.V.