Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Scienze Ambientali
Ciclo XXVII
Anno di discussione 2015

A Decision Support System for Probabilistic Ecological Risk Assessment (PERA) of pollutants on aquatic ecosystems

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: CHIM/12
Tesi di Dottorato di Panagiotis Isigonis, matricola 955934

Coordinatore del Dottorato	Tutore del Dottorando
Prof. Gabriele Capodaglio	Prof. Andrea Critto

Αφιερωμένο στους γονείς μου, Όλγα-Μαρία και Μανώλης

Dedicated to my parents, Olga-Maria and Manolis

Summary

Ecological Risk Assessment (ERA) is a process undertaken for estimating the environmental harms caused by human activities. The assessment is based on three components: effect assessment, exposure assessment, and risk characterisation. The latter is a combination of the former two. Various methodologies can be used for performing ERA, which can be categorised into deterministic and probabilistic. Probabilistic techniques have been at the focus of research the last years, due to their elaborated character and the possibilities they offer for more refined risk assessments.

Despite their obvious advantages, probabilistic techniques present also disadvantages and challenges that need to be tackled. In the thesis, the possibility of exploring further the concept of Probabilistic Ecological Risk Assessment (PERA) is addressed. The main motivation of the thesis is identified in the effort to combine various well known concepts and methods for Ecological Risk Assessment, while enhancing them with new features and functionalities to serve the current needs of Risk Assessors. Therefore, providing a complete software package that allows performing efficient Propabilistic ERA (PERA) and offers related functionalities, all gathered in one place. The proposed software is developed as part of the research project AMORE (funded by the National French Research Academy – ANR).

A proposal for a Decision Support System (DSS), named AMORE DSS, supporting Probabilistic ERA is described in detail and validated through the application of the proposed DSS to a case study for assessing the effects and risks posed by the presence of cyanide in a river in north-western France. The AMORE DSS aims at allowing efficient Probabilistic ERA and tackles issues related with PERA and the concept of weighted Species Sensitivity Distributions (SSWD) such as the handling of uncertainty in PERA, the production of reliable SSWD graphs and the assessment of the quality of ecotoxicological data.

The theoretical section of the thesis is split into two main parts. In the first, the concept of Ecological Risk Assessment is introduced and the principal methods of interest are described. It is followed by the presentation of the concepts of Multi-Criteria Decision Analysis (MCDA) and Decision Support Systems (DSS), which are important aspects of the developed research.

The methodological developments of the thesis are based on a proposal for the estimation of the reliability and relevance of ecotoxicological data used in ERA, which is presented in detail and evaluated. The proposed methodology is based on Multi-Criteria Decision Analysis and allows the assessment of ecotoxicological data on the basis of a fixed set of criteria and mathematically stable and robust aggregation techniques. Therefore, the methodology suggests the production of reliable weighted Species Sensitivity Distributions, a vital component of the probabilistic ERA and the calculation of risk probabilities. The proposal allows incorporating in the risk assessment the knowledge gathered from an expert panel and gives significant strength to the risk assessors for the performed assessments, through the use of previously not widely available information and expertise.

The proposed DSS is built on the three components (exposure, effects, risk) of ERA and provides a complete set of functionalities to the risk assessors, enhanced with unique features. The thesis describes in detail the development of the software and the functionalities of each of its modules. The Exposure Assessment module aims at providing to the Decision Maker/Risk Assessor a collection of tools for the statistical analysis of

environmental exposure data, through the concept of Predicted Environmental Concentration. The Effect Assessment module is based on the concept of weighted Species Sensitivity Distributions (SSWD) and incorporates the proposed methodology for the assessment of the reliability and relevance of ecotoxicological data. The Risk Assessment module is based on the concept of Potentially Affected Fraction (PAF) and aims at synthesising the results of the previous two modules for the estimation of risks, in an efficient and easy to present way.

The last part of the thesis is dedicated to the application of the DSS to a real life case study. A Risk Assessment process is performed for estimating the sensitivity of species to the presence of Cyanide (CN) in the environment, for estimating Environmental Quality Criteria (EQC) for the assessed case and for estimating the level of risk posed from Cyanide at the ecosystem. The assessed area is the Selune rivershed in the Manche department of the lower Normandy region in France, where four sampling stations have been identified with records of Cyanide presence for the period 2005-2014. Regarding the ecotoxicological data of the case study, 26 scientific articles on cyanide toxicity, published in the period 1965-2011, have been analysed for the extraction and assessment of 46 toxicological endpoints for the aquatic environment.

The case study is firstly based on all the available ecotoxicological data and secondly based on data split per taxonomic groups (i.e vertebrates, invertebrates) and trophic levels (i.e. primary producers, primary/secondary consumers). Specifically, six (6) sets of SSWD graphs are produced (i.e. All data, Vertebrates, Invertebrates, Primary producers, Primary consumers, Secondary consumers) with the use of two weighting options: (i) the weighting coefficients that are produced with the application of the MCDA based methodology and (ii) equal weighting coefficients for all the data. A comprehensive comparison of the two types of SSWD is performed and discussed in detail for the identification of the appropriateness of the fitting of the SSD curves to the experimental data. Hazardous concentrations (HCx) are estimated and presented for all the taxonomic groups and trophic levels. In addition, in combination with the results of the statistical analysis of the environmental exposure data, the risk is estimated for the assessed stations in the case study area. The results of the case study show that the primary producers are found to be the most sensitive trophic level while Invertebrates are more sensitive as a taxonomic group for low cyanide concentrations and Vertebrates are more sensitive for higher concentrations. Regarding the calculated risk indices, station 3 (L'Yvrande) of the Manche region is the area with the higher estimated risk.

The performed application of the DSS in the cyanide case study demonstrates a complete probabilistic Ecological Risk Assessment process with the use of Species Sensitivity Distributions and the utilisation of Multi-Criteria Decision Analysis. The case study, alongside with the validation of the developed DSS, demonstrates the performance of the proposed MCDA-based WoE framework for the analysis of ecotoxicological data, based on three distinctive Lines of Evidence (Experimental Reliability, Statistical Reliability, Biological Relevance). The framework and the related MCDA methodology constitute an innovative development in the field of quantitative ecotoxicological data assessment frameworks. Furthermore, a robust performance of the DSS has been identified, which allows potential for adoption within the risk assessment research fields.The thesis is concluded with future considerations for the developed DSS, which could provide interesting functionalities and extensions of the capabilities of the software.

Acknowledgements

First I would like to thank Dr. Alex Zabeo, who acted as co-tutor, colleague and friend at the same time. Without his continuous and unrivaled support, precious guidance and professionalism, this thesis would not have been a reality. Alex, you have shown great flexibility and understanding, always a pleasure working with you during those 3,5 years.

I am grateful to Dr. Elena Semenzin for her support and valuable comments for the thesis and to prof. Giove for his contributions in the development of the mathematical models.

I warmly thank Marco Stefan for the excellent contribution in coding the software of my PhD thesis and for coping with my endless to-do lists with patience.

I am thankful to Dr. Philippe Ciffroy, the supervisor of the AMORE research project, and Dr. Bertille Richard, who have offered lavishly their expertise on ecotoxicology in the creation of the DSS and the application of the case study.

I acknowledge and thank my tutor, prof. Andrea Critto, for his guidance and the leader of our research group, prof. Antonio Marcomini, for giving me the possibility to be part of a talented group of researchers during my PhD programme and the financial support.

Nevertheless, I would like to thank each member of the expert panel for their outstanding contribution to the project: Dr. Agerstrand Marlene, Dr. Andres Sandrine, Dr. Beaugelin Karine, Dr. Bisson Michele, Dr. Casas Stellio, Dr. Cauzzi Nicolas, Dr. Emmanouil Christina, Dr. Geoffroy Laure, Dr. Gilbin Rodolphe, Dr. Grasso Paola, Dr. Grote Matthias, Dr. Hayaud Nathalie, Dr. James-Casas Alice, Dr. Marliere Maryse, Dr. Maurau Sylvaine, Dr. Paschke Albrecht, Dr. Pery Alexandre, Dr. Pucheux Nicolas, Dr. Roth Nicolas, Dr. Troise Adrien, Dr. Vincent Jean-Martin and Dr. Vivier Stephanie.

Last but most important of all, I would like to thank my parents, Olga-Maria and Manolis, to whom I simply owe everything. This research is dedicated to them for their continuous and uncontested love and all their hard work to create our family, educate me and my siblings, and offer us everything we ever needed. I am who I am, simply because of you. I love you deeply.

Table of Contents

List of figures

List of tables

1. Introduction

In this chapter, the motivation behind the need for the research into probabilistic ecological risk assessment and the proposed developments for enhanced and more reliable results for the purposes of risk characterisation is discussed. In addition, an outline of the focus of subsequent chapters is provided.

1.1. Motivation & Objectives

Water, as one of the most important natural resources, was, is and will be essential to humans (Aulenbach, 1968). On one hand water consumption and on the other hand industry, urbanisation and human development require large amounts of water to be used by humanity (Meinzen-Dick et al, 2002). Water pollution is an area that has been highly researched by scientists all over the world, in numerous aspects and situations (APHA, 1915; Hellawell, 1986; Olness, 1995; Zhao et al., 2014), though a necessity for continuous research is present due to the importance of the effects. In many cases, natural or anthropogenic factors (e.g. chemicals, heavy metals) cause severe water pollution to fluvial systems (e.g. release of single substances) which, on a larger scale, is interconnected with various other impacts on the environment. The assessment of the risks and impacts created by such factors is at interest, as a case of concern.

Moreover, the assessment of risks of chemicals is a topic of major concern on European regulatory level. Several regulations have been put into force with most important the REACH (Registration, Evaluation, Authorisation and Restriction of Chemical substances) regulation (EP/EC, 2006) that aims to improve the protection of human health and the environment through the better and earlier identification of the intrinsic properties of chemical substances.

Environmental Risk Assessment (ERA) is the procedure to examine the risks resulting from hazards in the environment that threaten ecosystems, plants, animals and humans (Critto et al, 2009). ERA is conducted in phases, with the most common being the 'Exposure', the 'Effect' and the 'Risk' Assessment (EEA, 1998). Major issues in Environmental Risk Assessment are the heterogeneity of the information that needs to be taken into consideration, the present variability in the amounts of information and the different sources of origin (Duboudin et al, 2004; Forbes and Callow, 2002). In addition, the insights and preferences of the stakeholders involved in an assessment is an important aspect of the information that can possibly be taken into consideration.

Standard approaches are based on deterministic comparison of estimated exposure concentrations to the concentrations of the toxicant, below which adverse effects are unlikely to occur for the potentially exposed species (Hickey, 2010). The calculated concentration is known as the 'Predicted No Effect Concentration' (PNEC). Deterministic methods are usually useful for screening purposes and provide only one point estimate of environmental risk, known as 'Risk Quotient' (RQ).

The latest developments in the research field of risk assessment have introduced the use of more sophisticated probabilistic methodologies. Probabilistic ERA is an improvement over the RQ approach and it will likely continue to develop as the entire science of risk assessment

advances (Solomon et al., 2002). Probabilistic risk assessment produces distributions or range of values instead of one fixed value in the exposure and effect assessment sections. The results of the refined risk assessments show the range of possible environmental impacts and therefore provide the risk assessors with flexible tools for making decisions.

Probabilistic approaches to ecological risk assessment (PERA) have been recommended for later tiers in the ERA process (ECOFRAM, 1999). Despite its advantages (i.e. use of all relevant single-species toxicity data, allowance of quantitative estimations of risks when combined with exposure distributions), PERA does have some disadvantages as well: more effects and exposure data are usually needed, it does not address all sources of uncertainty and has not been widely calibrated against field observations (Solomon et. al, 2002).

Various techniques and methods have been used in Europe and the US, the last 40 years, for performing ERA. The most widely adopted approach is the 'Species Sensitivity Distribution – SSD' (EC, 1996). Despite its well known advantages, the SSD approach has received various critisisms regarding issues such as the statistical methods used for the analysis of the data, the differences in the handling of the data between experts and the availability of tools.

Various software exist for individual phases of the risk assessment process, such as 'Simple Box' (Brandes et al., 1996) and 'Focus models' (Focus, 2001) for exposure assessment, 'AQUATOX' (US EPA, 2002) and 'Demetra' (Craciun et al., 2006) for effect assessment and 'Gps1' (Hommen et al., 1993) and 'Prat2' (Solomon, 2000) for risk assessment. Though, none of them combines the functionalities for performing a complete risk assessment. 'Crystalball' (Oracle, 2008) and 'Risk Calc' (Ferson, 2002) are commercial, general risk assessment software but are not dedicated to Ecological Risk Assessment.

ETX 2.0 from RIVM (Van Vlaardingen et al., 2004) is the most well known and freely available software for probabilistic Ecological Risk Assessment, though it does not incorporate the use of weighting coefficients and produces conventional SSD graphs. Busy (Aldenberg, 2007) is a software for probabilistic ERA, though it focuses on Bayesian techniques and is programmed as a package in Mathematica.

In this context, the National French Research Academy (ANR) has funded the European Research project 'AMORE' (Multi-criteria Analysis for the development of Decision Support tools for the prevention of Environmental Risks), which has been supervised and coordinated by EDF (Electricité de France) and aimed at the development of state-of-the-art tools for Ecological Risk Assessment and the support of the development of new methodologies to tackle the major issues in the field of ERA, PERA and SSD, as well as their related available software, by the creation of a complete software package for Probabilistic Ecological Risk Assessment. The proposed software is in the form of a Decision Support System (DSS), named AMORE DSS, and aims at supporting and allowing efficient Probabilistic ERA, while it tackles issues related with PERA and the concept of Species Sensitivity Distributions (SSD).To this end, it has been identified that Multi-Criteria Decision Analysis (MCDA) provides many tools to the research world for the analysis, understanding, assessment and complex presentation of issues, by taking into consideration the opinions of various stakeholders (of completely different working/decision/research groups), the impact of various groups of criteria and the fragile balance between the importance of each characteristic (Giove et al, 2009).

Notable uses of MCDA are the integration into Decision Support Systems (DSS) and into Weight of Evidence (WOE) approaches. A Decision Support System is usually computer based

software that is designed for supporting management decisions (Giove et al, 2009) and there is a wide use for environmental problems and general risk assessment. A Weight of Evidence approach can be defined as a framework for combining individual lines of evidence, using methods that are either qualitative or quantitative, in order to develop conclusions regarding questions concerned with the degree of impairment or risk (Linkov et al, 2009). In this context the use of Multi Criteria Decision Analysis, integrated in a WoE framework, has been considered an excellent approach for the proposed research issues and can be used in order to support the decision making processes.

The objective of the PhD project, within the AMORE project, has been the development of an innovative MCDA-based Decision Support System (DSS) for the probabilistic assessment of environmental risks related with fluvial systems. The DSS aims at performing Ecological Risk Assessment, based on the concept of Species Sensitivity Distributions (SSD), and incorporates a series of improvements from existing methodologies and other software. The DSS is based on the three main concepts of ERA: (1) Exposure Assessment, (2) Effect Assessment and (3) Risk Assessment. Each concept has been implemented in a stand-alone module of the software. Each module incorporates a series of functionalities and capabilities, which are described in detail in chapter 4.2.

The DSS allows the integration of all sorts of available experimental information (related to contaminated aquatic systems) by means of MCDA methods, provides the stakeholders with reliable information on the uncertainties in the performed evaluation and supports the REACH (Registration, Evaluation, Authorisation and Restriction of Chemical substances) regulation implementation.

The proposed research aims at tackling existing issues of the SSD concept and specifically provide a methodology for assessing the quality of input data, by any independent evaluator/assessor, a process that allows performing a transparent and good risk assessment (Forbes and Callow, 2002). In this view, a WoE methodology has been envisioned and proposed for effectively estimating the quality of SSD input data. The methodology is based on Multi-Criteria Decision Analysis techniques, builds up on the concept of ‘weighting coefficients’ for the production of weighted Species Sensitivity Distributions (SSWD) and constitutes an attempt to handle the topic in a quantitative and robust way. The PhD project includes the application of the developed DSS to a case study: the river Selune (France), representing a highly impacted western European river basin with a strong interaction with the coastal zone.

The following subchapter provides a comprehensive outline of the thesis.

1.2. Outline of the thesis

The presented research of the PhD thesis is split into the following three main sections: (A) the theoretical background, (B) the Methodological development and (C) the Application to case study.

The thesis begins with the exploration of the theoretical background information of the key concepts, mainly Ecological Risk Assessment (ERA), Multi-Criteria Decision Analysis (MCDA) and Decision Support Systems (DSS), in Section A and specifically chapter 2. Each of the topics is introduced in greater depth, to support the reader and emphasize the necessary details for the flawless understanding of the proposed research. The necessary definitions and notations, which are important throughout the entire research report, are described.

In particular, Ecological Risk Assessment is introduced and the concepts of tiered risk assessment procedures is shortly discussed in chapter 2.1. Empashis is given to the use of the well known concept of Species Sensitivity Distributions (SSD) in risk assessment and the derivation of Environmental Quality Criteria/Standards (EQC/S).

The remaining two topics (MCDA and DSS) are discussed and presented in chapters 2.2 and 2.3 respectively. These two concepts, are harmonically combined for performing Ecological Risk Assessment in the second part of the PhD thesis and the main instances which are utilised in the methodological development are explained shortly.

The two chapters (chapters 3 and 4) of the methodological development (Section B) are ordered so that they cover two distinct strands of the proposed improvements to the way current ecological risk assessment is understood and performed. The key research topics and chapters are briefly outlined below.

One of the main contributions of this thesis is the MCDA-based Weight of Evidence (WoE) methodology for the estimation of the reliability and relevance of ecotoxicological data, as described in Chapter 3. The importance of analysing ecotoxicological data, which may be used in the risk assessment processes, has been discussed by a number of authors, who have proposed their visions for the assessment frameworks (Klimisch et al (1997), Warne et al (1998), Hobbs et al (2005), Schneider et al (2009), Breton et al (2009), Ågerstrand et al (2011)), yet the implications and the nature of the issue allow refining the assessment and provide space for the proposal and development of new frameworks. The chapter describes our vision for the assessment of ecotoxicological and explains in detail the process that has been followed throught the period of the PhD programme for the development of the WoE framework and the complete mathematical notations, upon which the MCDA methodology is built.

Chapter 4 is dedicated to the description of the developed Decision Support System for probabilistic Ecological Risk Assessment with the use of multiple criteria and the MCDA-based methodology for the assessment of ecotoxicological data. Effectively, the software is a combination of current scientifically accepted probabilistic ecotoxicological methodologies, adapted and enhanced with innovative features in order to take into account more effectively the reliability and relevance of ecotoxicological data and provide a more robust and efficient process for risk assessments. The methodological framework of the DSS, alongside with the description of the DSS modules are presented in detail.

Evidently, every proposed theoretical scientific development has to be validated and tested for its usefulness and robustness. In this view, Section C (Chapter 5) is dedicated to the presentation of the case study and the discussion of the related results. The proposed methodologies and the developed DSS have been applied to a case study for the risk assessment from contamination of cyanide in the Selune rivershed in France. The results of the performed risk assessment are presented in detail with numerous graphs and are shortly discussed.

The research discussed above is evaluated and summarised in Chapter 6. An overview of the performed activities is shortly presented and accompanied by a discussion on the challenges faced and the proposed solutions for tackling them.

Section A: Theoretical background

2. Basic theoretical backgrounds

The following chapters contain the main theoretical backgrounds, which have been considered essential for a smooth understanding of the thesis and provide the main information that a reader needs to be able to follow smoothly the content of sections B and C.

Chapter 2.1 starts with a short introduction to Ecological Risk Assessment (ERA), which is followed by a consice presentation of the concepts of Environmental Quality Criteria (EQC), in which the quality of ecotoxicological data as well as the methods (e.g. Species Sensitivity Distributions (SSD)) used to derive them, play a crucial role. The second chapter is dedicated to the introductions of the well established methods 'Weight of Evidence', 'Multi-Criteria Decision Analysis' and 'Fuzzy logic' (chapter 2.2). The last chapter is devoted to the presentation of Decision Support Systems (DSS – chapter 2.3).

2.1. Ecological Risk Assessment frameworks and methods

Ecological Risk Assessment (ERA) is defined as the estimation of both the magnitude and the probability of environmental harm caused by human activities (Barnthouse et al. 1986). Ecological risk assessment usually focus on the estimation of negative effects on specific ecosystems (Breitholtz et al, 2006) and according to the European Commission (2003) it is completed in four steps: hazard identification, dose-response assessment (effect assessment), exposure assessment, and risk characterisation. Many international organisations have developed frameworks for ERA, such as the US EPA (1998), the WHO (2001), the EC (2003) and others (OECD – Organisation for Economic Co−operation and Development, EPPO – European and Mediterranean Plant Protection Organisation, ECETOC – European Centre for Ecotoxicology and Toxicology of Chemicals). These frameworks have been evaluated, advanced and adapted in order to meet the needs of the assessors in various countries (e.g. European Union, United States, Japan, Canada, South Africa, Australia and New Zealand), as identified by Suter (2006) and Bradbury et al (2004).

In ERA two main tiers can be distinguished: screening ERA and site-specific ERA (Critto and Suter, 2009). While screening risk assessment aims at identifying chemicals and agents that do not pose hazards at the ecosystem under analysis, and thus could be excluded from the assessment process, site-specific risk assessment aims at providing estimations of risks to support decision-making processes (Critto and Suter, 2009).

In screening risk assessment environmental quality criteria (EQC) or standards (EQS) are usually adopted. They are threshold numerical values that indicate a level beyond which there is a significant risk that the associated environmental quality objective has not been achieved and for which the assessors should adopt actions for the preservation of the ecosystems, including the development of a site-specific risk assessment (EPA, 2005). The way environmental standards are derived, and the frameworks within which they are used, differ between countries and regions. In the recent years, various international frameworks and legislation have been developed to tackle important issues regarding the EQC, such as the establishment, the derivation methods and the implementation. In Europe these include i) legislation focusing on a specific environmental compartment such as the Water Framework

Directive (EC, 2000), followed by the Environmental Quality Standards Directive (EC, 2008a) of the European Commission setting EQS for 44 priority pollutants in inland, transitional and coastal waters and the related Technical Guidance Document (TGD-EQS) for Deriving Environmental Quality Standards (EC, 2011), as well as ii) legislation aimed at classifying and regulating the highly produced chemicals (>1 tonnage per year) such as the EU REACH regulation (EC, 2006), in which the standards of the European Chemicals Agency (ECHA, 2008b) are adopted, followed by the CLP (Classification, Labelling and Packaging) regulation (EC, 2008b). Outside Europe two documents can be cited as main references: the Water Quality Standards Regulation of the US Environmental Protection Agency (US EPA, 1983) and the related Water Quality Standards Handbook (US EPA, 1994), developed by US EPA for the aquatic environment but applied worldwide, while no other legislations specifically addressing the highly produced chemicals currently exist apart from the EU REACH.

The EQC can be derived either through deterministic or probabilistic approaches, with the latter being preferred in the recent advances in the field as they allow to take into consideration uncertainty as well as the spatial and temporal variability of the data (Verdonck et al, 2002). The most widely adopted probabilistic approach used for this purposed is the Species Sensitivity Distruibution (SSD), that has received significant attention the last 40 years, both in the US and in Europe. The concept was proposed as an ecotoxicological tool that is useful for the derivation of environmental quality criteria and ecological risk assessment and it has been initiated in the late 1970s in the US and the next decade in Europe (Posthuma et al., 2002), though SSDs have been steadily used ever since as, when used correctly, they allow greater statistical confidence into risk assessment processes when compared to traditional quotient and assessment factor approaches (Wheeler et al., 2002).

Posthuma et al. (2002) describe the SSD as "*a statistical distribution describing the variation among a set of species in toxicity of a certain compound or mixture. The species set may be composed of a species from a specific taxon, a selected species assemblage, or a natural community."*

The basic assumption of the SSD concept is that the sensitivities of a set of species can be described by some distribution, usually a parametric distribution function (e.g. triangular, normal, or logistic distribution) or a nonparametric method. The available ecotoxicological data are seen as a sample from this distribution and are used to estimate the parameters of the SSD (Posthuma et al. 2002).

SSDs can be used in the well known 'forward' and 'inverse' ways (Van Straalen and Denneman, 1989). At the forward use, the risk assessment is performed through the calculation of the Potentially Affected Fraction (PAF), which is calculated based on the SSD and the estimated environmental concentration of a contaminant in the environment, whereas, in the inverse way the SSDs can be used for the derivation of EQS. The derivation is based on the selection of a cutoff percentage *p,* and the calculation of the estimated safe concentration (*HCp*) from the SSD graph, that is protective for the species of the compartment under assessment.

A complete description of SSD is presented in Posthuma et al. (2002) and a detailed critique of SSD is presented in Forbes and Calow (2002) in which the most signicant assumptions made in SSD-theory are reported and appraised. Specifically, Forbes and Calow (2002) raised a number of questions regarding effect of intraspecies variation, proportion of data between the different taxonomic groups and adopted statistical methods in SSD. To tackle these considerations, in 2004 Duboudin and colleagues (Duboudin et al., 2004) have introduced the concept of Species Sensitivity Weighted Distributions (SSWD) in which various statistical

production methods as well as weights for the ecotoxicological data are used in the production of SSDs.

In their study, Duboudin et al. (2004) have proposed a weighting coefficient combining two different criteria that allow taking into account: (1) the intraspecies variation in effect response and (2) the taxonomic groups' abundance. Though, this weighting coefficient is neither related with the quality of the assessed data nor with their reliability and relevance for the ecosystem of concern, elements which are considered highly important for the derivation of robust and reliable EQC/S.

The derivation of robust and reliable EQC/S mainly depends on the availability and quality of relevant ecotoxicological data. Ecotoxicological data can be obtained through many different approaches and conditions, e.g., the protocol can be standardised or not; time duration can vary among experiments, leading to chronic or acute data; different physiological endpoints can be observed, e.g. mortality, growth, reproduction and more; statistics used for interpreting data can differ, leading to e.g. NOEC or ECx and more. It is therefore of high interest the analysis of their reliability and relevance that will allow the derivation of more significant and relevant EQ criteria to be adopted in screeing ERA, as well as more reliable site-specific ERA.

Several frameworks have been proposed in order to address this issue. The most important are presented and analysed for their strengths and limitations in chapter 3.

2.2. Weight of Evidence, Multi-Criteria Decision Analysis (MCDA), Fuzzy Logic and Group Decision Theory

The term 'Weight of Evidence' constitutes neither a scientifically well-defined term nor an agreed formalised concept characterised by defined tools and procedures (Weed, 2005). According to ECHA, an evidence based approach involves an assessment of the relative values/weights of different pieces of the available information that have been retrieved and gathered in previous steps (ECHA, 2010). Therefore, each piece of information of a Weight of Evidence approach should be assigned to a value. This can be performed, either with the use of expert judgement or by applying a formal process to obtain objective values.

Weight of Evidence (WoE) refers to a large family of methods and is applied into various scientific projects, mainly known for the applications into human health and ecological risk assessments. Weed (2005) and Linkov et al. (2009, 2011) have provided comprehensive critical reviews on the concept and the uses of Weight of Evidence, both in an exploratory way as well as in an effort to provide a categorisation of the available qualitative and quantitative WoE methods and their use in environmental assessments. WoE can be defined as a framework for synthesizing individual Lines of Evidence (LoE), which are developed from available data to address a specific question (Linkov et al., 2009).

Multi Criteria Decision Analysis (MCDA) can be defined as a decisional support tool whose main goal concerns the **selection**, **ranking**, **scoring** or **screening,** among a set of admissible alternatives, on the basis of **multiple criteria**, taking into account Decision Makers/stakeholders preferences and Experts' knowledge (Koksalan et al. 2011, Figueira et al. 2005).

Multi-Criteria Decision Analysis includes a wide variety of methods for the evaluation and ranking, or selection, of different alternatives that consider all the aspects of a decision problem involving many actors (Giove et al., 2009).

Using Multi-Criteria Decision Analysis methods in WoE approaches allows to: i) classify available information according to a hierarchical structure based on different 'Lines of Evidence', each of them being subdivided into several levels of criteria; ii) normalise information, i.e. assigning common units to qualitative (e.g. originating from expert judgement), semi-quantitative (e.g. Boolean information) or quantitative information; iii) assign different weights and relations to the selected criteria in order to rank and compare criteria based alternatives through an integrated approach; iv) define decision indices integrating all the selected criteria on the basis of experts' judgements and decision makers' insights.

Features common to almost all the decision making processes include the following items:

- the decision maker (DM). A single person, a group of persons or an entity in charge of finding the best solution for the problem under examination;
- a set A of alternatives, in the finite case: $A = (a_1, \dots, a_m)$, out of which the DM must choose the best solution;
- a countable group of criteria $K = (k_1, \dots, k_n)$. Criteria define the alternatives; they are aspects of the problem that the DM considers crucial. Criteria can be organized into a hierarchical structure, i.e. a decision tree where the root is the objective function whose branches are the first-level criteria, each of them splits again into second-level criteria (sub-criteria), and so on till the last level, whose terminal leaves are the indicators calculated on the basis of the available information (data or judgments);
- the decision maker's preferences for the different evaluations of the criteria.

In case of an infinite set of alternatives the final solution of a MCDA problem is also related to:

- an objective or target function (to be optimised) used to score alternatives, usually an aggregation function;
- an algorithmic tool designed to optimize the objective function, considering all the above information.

The infinite-alternatives based field of MCDA is called Multi Objective Decision Making (MODM), and is the counterpart of the finite-alternatives based branch called Multi Attribute Decision Making (MADM). Typically MADM can be subdivided into three main categories (Vincke, 1992):

1. Multi-Attribute Utility/Value Theory (MAUT/MAVT)
2. Outranking and
3. Interactive methods.

In Multi-Attribute Utility/Value Theory (MAUT/MAVT) criterion values are firstly normalised into a common numerical scale, by means of a suitable transformation function (or Utility/Value Function). Then, the criteria are aggregated by a suitable aggregation operator, a function which satisfies a set of rationality axioms. Using a bottom up approach, this operation is repeated for all the nodes in the decision tree (if the problem is hierarchically structured) for all the alternatives. Each branch or level of the tree may be aggregated to its root by using different aggregation functions on the basis of the relations between the criteria

of concern. At the tree root (the objective) a single numerical value is finally computed, which is the score of the proposed alternatives. The alternatives can then be rated and ranked, since MAUT/MAVT produces a total ordering, and so the best one can be selected.

Outranking methods are based on an “outranking relationship" between alternatives stating that one alternative may be dominant, with a certain degree, over another one. These outranking relationships are neither complete nor transitive generating therefore only partial orders. This is due to the fact that outranking methods comprise the existence of non-comparable alternatives.

Interactive methods obviously consist of the iteration of certain procedure steps. At first, a rough solution is proposed to the DM, which can accept or reject it. In the latter case new data are acquired and/or more information is supplied (e.g. extra information concerning a DM's preferences) to the system. Then a new solution based on new data and information is presented to the decision maker. This extraction of preferences and re-computation steps are repeated, creating successive compromise solutions, until the satisfaction of the DM is reached.

The WoE framework and the MCDA methodology of the PhD project, which are described in chapter 3, have been developed based mainly on the notions of the MAVT, as introduced in this paragraph, since the concept of ‘value functions’ is used for the handling of the criteria values and their transformations into certain scales and a ‘decision tree’ has been designed and used as a foundation of the assessment process. The methodology is combining the value theory with elements of two other branches of decision making: fuzzy logic and group decision theory. Fuzzy logic introduces the concepts of ‘partial membership’ and ‘degree of truth’ and elements of group decision theory provide support in the managerial and decisional processes for the development of the WoE and MCDA frameworks.

The following paragraphs are describing in a short but comprehensive way the basic definitions of:

- Value functions
- Fuzzy Logic and
- Group Decision Theory

2.2.1. Value Functions

Multi Attribute Value Theory (MAVT) consists in the creation of a value function (normalisation function) for each criterion, used to normalise all criteria values in a common numerical closed interval (Keeney et al. 1976). The normalised criteria values are aggregated towards the obtainment of a final alternative score value. Normalisation functions are usually monotonic and their co-domain is included in the closed interval $[0,1]$. Given that the assignment of such functions is subjective (even if guided by a suitable software interface) and depends on the user's preference structure or perception about the criterion impact, the normalisation problem must be solved without resorting to any type of data-driven formulas (e.g. subdivision by maximum). This is due to the fact that any data-driven normalisation algorithm is quite sensitive to outliers and may therefore induce distortion in the final scoring. Distortion is also present if the data which have to be normalised are dense around an average value. As a consequence the most feasible solution for normalisation is performing re-scaling of all the available data of criteria into a common closed numerical scale. This solution not only is simpler but also solves the normalisation problem in a better way.

Normalising functions can be divided in two main categories: discrete and continuous. Discrete normalisation functions are mapping the domain into a finite number of alternatives (which may be expressed as fixed numerical values like in Figure 1a but also as lexical labels e.g. "BAD" or "GOOD"). Instead, continuous normalisation functions are continuous, usually monotonic, functions mapping the domain into any value in the co-domain e.g. piecewise linear functions (Figure 1b).

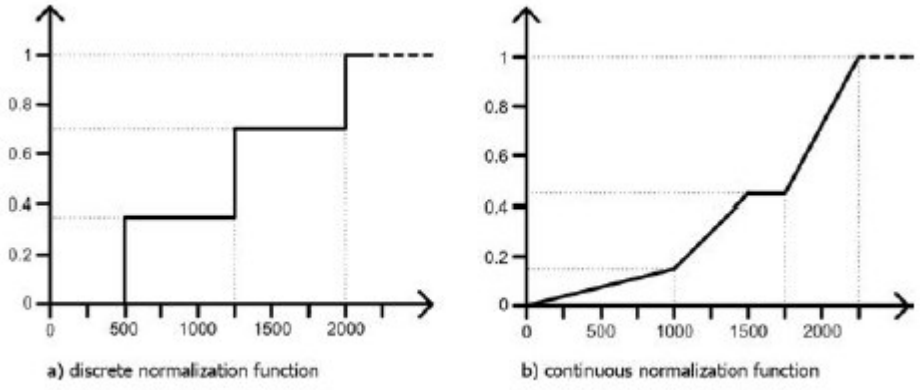


Figure 1: Examples of discrete a) and continuous b) normalisation functions

In crisp logic value functions are dealt with Boolean logic, since the characteristic functions of criteria can take only two values and the evaluation of criteria can be either true or false. Crisp sets are well-defined sets based on those characteristic functions. Though, crisp logic is not sufficient for describing many real life applications, due to the existence of uncertainty and subjectivity. Those cases can be handled with fuzzy logic, an extension of the notion of crisp logic, which is described in the next paragraph.

2.2.2. Fuzzy Logic

The notion of *fuzzy logic* and *fuzzy sets* were introduced by Lotfi Zadeh (Zadeh, 1965) as a formalization of vagueness. The basic idea concerns the fact that a predicate may apply to an object in a non-absolute way, but rather to a certain degree, e.g. who can say if a person is part of the set of tall persons or if a movie is part of the set of interesting movies? These inclusion problems are very likely to happen in real life but are almost untreatable with classical bivariate *crisp* logic. Fuzzy logic is in fact a multi-valued logic (i.e. a logic which admits truth values different from “true" and “false") characterised by a continuous truth degree space, usually corresponding to the whole interval $[0,1]$. Furthermore, when linguistic variables are used, the membership degrees may be managed by specific functions called *membership functions* and usually denoted by the μ symbol. In Figure 2 membership functions for the crisp and fuzzy interpretations of the “tall persons" example are reported.

Fuzzy logic has been applied to many fields, from control theory to artificial intelligence (Mamdani 1977, Klement et al. 1994). Formally, given a set U (i.e. Universe) whose generic elements are denoted by x, a fuzzy set A in U is characterized by a membership function $\mu_A(x)$ which associates with each element in U a real number in $[0,1]$. Then the fuzzy set A is usually denoted by the set of pairs:

$$A = \{x, \mu_A(x), x \in U\}$$

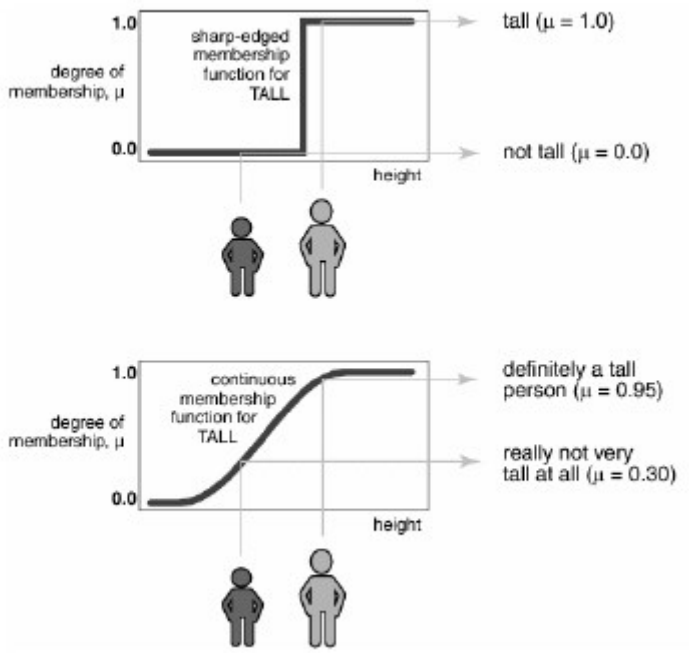


Figure 2: Crisp and fuzzy membership functions for the set of "tall persons"

For a classical crisp set:

$$\mu_A(x) = \begin{cases} 1 & iff\ x\ \in A \\ 0 & iff\ x\ \notin A \end{cases}$$

Other characterisations of classical crisp sets can be translated in the fuzzy environment. For example the formalisations of the concepts of *complement* and *cardinality* of the fuzzy set A are reported below:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x), x\ \in U$$

$$|A| = \sum_{x \in U} \mu_A(x)$$

It is important to note that membership degrees are not probabilities. This can be perceived by noting that the probabilities related to a finite set must sum up to one which is absolutely not true in fuzzy sets theory.

An important role in fuzzy logic is played by set-theoretic operations related to fuzzy sets. The notions of intersection and union can be translated into fuzzy sets as explained by Zadeh (Zadeh, 1965). In his work Zadeh utilises the minimum and maximum aggregation operators to mimic respectively intersection and union. Bellman and Giertz (Bellman et al., 1973) pointed out that intersection and union can be interpreted as the logical "AND" and "OR" operators respectively and gave a formal justification for the use of minimum and maximum by Zadeh.

2.2.3. Group Decision Theory

Group Decision and Negotiation is a field of research that aims in developing and studying methods and tools that provide humans the ability to use formal procedures for reaching collective decisions (Kilgour et al. 2010). Group Decision Making is a multi-person process that is used in complex and ill-structured situations where decisions need to be taken (Jelassi et al. 1990). Jelassi (Jelassi et al. 1990) identifies four types of multi-person decision making situations, based on the way the final decisions are taken:

- Individual decision making in a group setting, where one person (i.e. the DM) utilises the knowledge of a group for taking a decision,
- Hierarchical decision making, where decisions are organised in a hierarchical way, a person is responsible for the top level decision and group members are responsible for lower level decisions,
- Group decision making or one-party decision making, where many members participate in the process and are responsible for the final decision, and
- Multi-party decision making or negotiation, where several decision makers represents different parties with possibly conflicts of interest regarding the decision to be taken.

Already for many years, advances in Information Technologies (IT) have created a growing interest for the development of Group Decision Support Systems and reviews are available (Finlay et al. 1992). A Group Decision Support System can be defined as an interactive, computer-based system which facilitates solution of unstructured problems by a set of decision makers working together as a group (De Sanctis et al. 1985). Group Decision processes are usually complex and limitations may exist for the allocation of resources. Researchers have explored the possible combination of various models and methods, such as MCDA models (Davey et al. 1998) and Multi-Attribute Utility methods (Bose et al. 1997), with Group Decision Making and their inclusion in Group Decision Support Systems. Olson (Olson et al. 1987) reviews and presents techniques for the extraction of expert knowledge and the design of expert systems. Many direct (e.g. interviews, questionnaires, observation of task performance and more) and indirect methods (multidimensional scaling, hierarchical clustering, ordered trees) are identified and presented in the paper of Olson, which are techniques similar with the ones used for the development of the methodologies used in this PhD project (as those are explained in chapter 3.1).

2.3. DSS

Decision Support Systems (DSSs) are tools created with the aim of supporting Decision Makers (DMs) in taking more informed and concrete decisions. They are designed to take into consideration different forms of input and information and are calibrated to fit the special needs of the specific projects they are produced for.

A Decision Support System (DSS) is a computer-based information system that supports business or organisational decision making activities (Burstein et al. 2008). DSSs serve the management, operations and planning levels of an organisation and help to make decisions, which may be rapidly changing and not easily specified in advance (Burstein et al. 2008). Decision Support Systems can be either fully computerised, human or a combination of both.

According to Keen and colleagues (1978), Decision Support Systems (DSSs) are IT-enabled tools that aim to enhance the effectiveness and efficiency of managerial and professional decision making for ill-structured problems. There exists a wide variety of Decision Support

Systems, including passive DSSs that provide the user with compiled information only and active DSSs that provide specific solutions or recommendations (Holsapple, 2008).

Scientific literature provides a rich collection of articles regarding the development of DSSs related with risk assessment and management of various different environmental topics, such as contaminated sites (Marcomini et al., 2009), water supply systems (Baroudy et al., 2006), flood management (Levy et al. 2007), regional forest management (Zambelli et al., 2012) and many more.

2.3.1. Definitions and objectives

There are many different ways of categorising Decision Support Systems, due to their vast fields of applications as well as their different characteristics.

According to Power (2002), a first significant classification separates the DSS into:

1. Model-driven
2. Data-driven
3. Communication-driven
4. Document-driven
5. Knowledge-driven

According to the same author (Power, 2002): 'A Model-driven DSS emphasizes access to and manipulation of a statistical, financial, optimization, or simulation model. A Model-driven DSS use data and parameters provided by DSS users to aid decision makers in analysing a situation, but they are not necessarily data intensive. A Data-driven DSS emphasizes access to and manipulation of a time-series of data. A Communication-driven DSS supports more than one person working on a shared task. A Document-driven DSS manages, retrieves and manipulates unstructured information in a variety of electronic formats. Finally, a Knowledge-driven DSS provides specialised problem-solving expertise stored as facts, rules, or procedures'.

Holsapple (2008) divides the structural definition of Decision Support Systems into four essential components:

1. a language system (LS),
2. a presentation system (PS),
3. a knowledge system (KS) and
4. a problem-processing system (PPS).

A 'language system' consists of all messages the DSS can accept, a 'presentation system' consists of all messages the DSS can emit, a 'knowledge system' consists of all knowledge the DSS has stored and retained, while a 'problem processing system' is the software engine of the DSS (Holsapple, 2008).

On a different note, Marakas (1999) divides the architecture of Decision Support Systems into five components:

1. user(s)
2. the user interface
3. the knowledge engine
4. the data management system
5. the model management system

The data and model management systems relate to the definitions provided by Power (2002) and are responsible for the analysis performed by the DSS on a data or model basis. Users can

have various roles and influence on the DSS and the user interface is the component that defines how the users interact with the system. The knowledge engine is the component that connects the users and the data/model management systems, through the user interface. Regardless, though, of the available classifications and architecture descriptions all DSSs have the same main objective: To assist and systemise processes of decision making for the benefit of their users.

Section B: Methodological development

The following chapters contain all the information regarding the methodological developments that took place during the PhD programme. Initially, a very detailed description of the developed MCDA-based, WoE methodology for the assessment of the reliability and relevance of ecotoxicological data is provided in chapter 3. The chapter contains all the information regarding the procedure that has been followed for the design and development of the methodology (paragraph 3.1). Furthermore, the mathematical foundations of the MCDA-based methodology are described in complete detail in paragraph 3.2 and the description of the theoretical WoE framework and the multiple criteria which are used for the assessment of ecotoxicological data for laboratory biotests are described in paragraph 3.3. Lastly, interesting statistics, which are extracted from the knowledge base of the developed DSS are described in paragraph 3.4.

The AMORE Decision Support System is described in detail in Chapter 4. Paragraph 4.1 is dedicated to the description of the methodological framework and the model development details, whereas paragraph 4.2 contains all the important information of the three modules of the DSS.

3. Weight of Evidence framework and MCDA methodology for the analysis of the reliability and relevance of ecotoxicological data

As described in section (A), ecotoxicological data are used in the Ecological Risk Assessment processes and in the derivation of EQS. During the last decades, various frameworks have been adopted for the assessment and evaluation of ecotoxicological data and the analysis of their reliability. Evaluations of individual ecotoxicity data have been often done on a case-by-case expert judgement. This resulted usually in a poor transparency, reproducibility and predictability of the risk assessment process because different experts may have their own implicit set of criteria and rankings for rejecting or not an ecotoxicity datum.

To improve ecotoxicity data evaluation, several structured frameworks based on lists of pre-defined criteria have been proposed. A first attempt to classify ecotoxicological data, according to a systematic approach, and to harmonise data evaluation processes was proposed by Klimisch et al (1997), who proposed the classification of data into four qualitative reliability categories (i.e. Reliable without restriction, Reliable with restriction, Not reliable and Not assignable). Warne et al (1998) proposed a more detailed scheme for assessing the quality of aquatic ecotoxicological data. It is based on a series of questions and a score is given to the answer of each question; the scores of all questions are then summed in order to obtain a 'total score' for each datum, expressed as a percentage of the maximum possible score. The data are classified as being unacceptable, acceptable or high quality, depending on whether the quality score is <50%, between 51-79% and >80% respectively. Hobbs et al (2005) submitted Warne's scheme to a panel of experts and refined the set of questions in order to modify/clarify ambiguous or poorly written questions, to reduce assessor variation and thus improve the consensus level among experts. Similarly, Schneider et al (2009) developed a tool (called ToxRTool) for assessing reliability of toxicological data (both *in vitro* and *in vivo* data and rather dedicated to human health risk assessment). The process followed by Schneider et al (2009) is similar to those of Hobbs et al (2005), i.e. based on a set of questions, refined after

consultation of a panel of experts. One innovation of Schneider's framework is the introduction of 'red criteria': non-compliance with at least one red criterion leads to the 'Not reliable' category, irrespective of the total score achieved. Breton et al (2009) developed a Quality Assurance system (called eco-QESST) specifically dedicated to three of the most common tests used in ecotoxicology, i.e. the *fish acute toxicity test* (OECD 1992), the *Daphnia acute immobilization and reproduction toxicity test* (OECD 2004) and the *algae growth and inhibition effects test* (OECD 2002). The eco-QESST system is based on a set of questions, most of them being answered as either by 'Yes', 'No', 'Not applicable' or 'Not reported'. A scoring process is included in the eco-QESST system: a 'Yes' answer is given a specific weight, depending on the relative importance of the factor addressed by the question, while a 'No' or a 'Not reported' answer is given a zero weight. The overall study quality score (OSQS) is calculated as a percentage of maximum sum of weights. Finally, Ågerstrand et al (2011) reviewed criteria for reporting and evaluating ecotoxicological tests dedicated to pharmaceuticals. A framework allowing a comparative assessment of standard and non-standard tests was then developed. A main innovation of Ågerstrand's framework was the explicit subdivision of the analysis criteria in reliability and relevance criteria.

The described frameworks are a good starting point for the analysis of ecotoxicological data, because they are based on a priori selected and objective criteria and help to rank the acceptability of individual datasets to the fulfilment of strict specifications. Despite their obvious positive input, existing assessment schemes present however some flaws, e.g.: (i) initial schemes proposed by Klimisch are based on rather poorly written questions that can be interpreted differently by risk assessors, leading to significant variations among experts. More unambiguous questions are needed for improving consensus and reproducibility among experts; (ii) for most of the frameworks, the qualitative 'summary' result appears 'poor' compared to the information collected during the assessment process: data are assigned to three (or four) qualitative categories only. Concretely, the first categories (Reliable without restriction, Reliable) are actually not (or poorly) distinguished in further risk process (e.g. they are generally equally used for SSD construction), while data belonging to the last one (Not reliable) are actually completely ignored. Only the eco-QESST system proposes a quantitative scoring system, but rules for the assignment of data to Klimisch categories were finally defined (from 'Reliable without restriction' if Score>90%, to 'not assignable' if Score <60%) and quantitative scores are hidden; (iii) Schneider et al (2009) identified rightly that "a source of heterogeneity among experts is the degree to which they include elements of relevance and adequacy into their rating and how they weighted those against reliability. The concepts of reliability and relevance and their discrimination need to be discussed more thoroughly". However, except Ågerstrand's scheme, it can be noted that reliability and relevance are not explicitly distinguished, leading to biased weighting process.

To this end, a new framework is proposed for the analysis of the reliability of ecotoxicological data, based on the use of multiple criteria. The proposed framework is a Weight of Evidence framework, which allows the analysis of data based on a hierarchically structured set of criteria and the use of an innovative MCDA-based aggregation methodology.

3.1. Participatory process

One of the main innovative aspects of the proposed methodology is the strong focus that has been given to the participation of experts in the design and creation of the assessment process. A total of 23 experts and senior researchers in the fields of ecotoxicology and chemistry, from prestigious research centres, universities and companies in France (EDF

Energy, Veolia Environmental Services, INERIS – French National Institute for the Industrial Environment and Risks, IRSN – Institute for Radiological Protection and Nuclear Safety) and Europe (KTH Royal Institute of Technology, University of Basel, UNICATT – Catholic University of the Sacred Heart, UFZ – Helmholtz Centre for Environmental Research, BPI – Benaki Phytopathological Institute) have been involved throughout the course of the project, in a strongly interactive participatory process.

The participatory process has been structured in such a way that it would allow the extraction of valuable information regarding the criteria that are used in the assessment of ecotoxicological data, as well as the possible interactions among those criteria and the importance of the elements used in the Weight of Evidence methodology. In addition, it also enabled the research team to gather the necessary information for the creation of the knowledge database of the European-funded AMORE research project, upon which the MCDA methodology has been designed. The process included the use of two different questionnaires and the organisation of a workshop for gathering live the experts and providing them the possibility to interact in real time.

In the next sections, the process that has been followed is described in further detail.

3.1.1. Questionnaire for the evaluation of the criteria hierarchy

The assessment methodology is based on a hierarchical structure, which relates the different aspects of ecotoxicological data in a clear and solid fashion. The hierarchy is based on four levels, starting from the Lines of Evidence (LoE). Each LoE is subdivided into several categories, which are further subdivided in criteria groups, and finally in specific assessment criteria, which are evaluated with the use of detailed questions. Criteria-questions are the lowest level of the hierarchy, the one which must be informed by the user.

As a basis for defining such a hierarchical structure, the frameworks developed by US EPA (1991), Klimisch et al (1997), Warne et al (1998), Hobbs et al (2005), Schneider et al (2009), Breton et al (2009) and Ågerstrand et al (2011) were used. The criteria evaluation questions and weighting rules suggested in these publications were reviewed in detail to provide a first hierarchical structure, which included 24 criteria-questions, organised in 10 categories. This latter was submitted for evaluation to a panel of eight experts in the field of ecotoxicology, who had to answer the following five (5) questions for each criterion/question:

- In your opinion, is this criterion potentially relevant in the quality assessment of eco-toxicological data generated through lab bio-test?
- In your opinion, is the question unambiguous?
- Would you split the question into several questions because you consider that it ambiguously merges several issues?
- How would you change the sentence?
- In your opinion, is this criterion in the right Category / Criteria Group?

In addition, experts were asked if they would add new Categories/Criteria groups/Criteria-Questions.

This set of questions aimed at: (i) defining if the hierarchical structure was well designed (i.e. if criteria-questions are properly placed in the right group/category) and shared by a significant panel of experts; (ii) detecting potential lacking LoEs/Categories/Criteria groups/Criteria-Questions; (iii) guaranteeing that questions were unambiguous.

Based on the results of the first part of the experts' participatory process, the first version of the criteria-questions hierarchy has been significantly expanded to include a total of 57 assessment criteria-questions, organised in 23 criteria groups and 11 categories.

In the thesis, only the final WoE hierarchical structure resulting from this process of expert consultation is presented (see Table 11 in the Annex).

3.1.2. Questionnaire for expert consultation

Once the WoE hierarchical framework has been established and the evaluation basis had been set up, the next step of the participatory process consisted in the design of a process for the exploitation of the aggregation operators to be used in the MCDA methodology. In cooperation with the existing panel of experts, an analysis has been performed in order to identify which parameters and characteristics of the criteria hierarchy are important and should be explored further. The procedure included various steps and involved continuously the members of the expert panel. Initially a draft of the second evaluation questionnaire was designed with main purposes the exploration of the relations among criteria (e.g. identification of criteria which are interconnected and their simultaneous positive/negative evaluation influences the assessment) as well as the identification of the different types of importance of criteria (e.g. prerequisite criteria, very important criteria, not relevant criteria and so on). The questionnaire included 6 main points, as described below, and was presented to 12 members of the panel of experts during the dedicated workshop, with the aim of evaluating the clarity of the questionnaire, testing the procedure and identifying possible improvements that could be implemented in the final version of the questionnaire. Reaching a consensus among experts, when possible, was a priority throughout the procedure.

The first version of the questionnaire included the following points, which were defined as important for the evaluation of the criteria hierarchy:

- Identification of criteria whose evaluation overrules (positively or negatively) other criteria belonging to the same criteria group ('Over' and 'Veto' criteria).
- Identification of synergic or redundant effects of criteria to the evaluation.
- Identification of the importance (ranking) of 'Optimum and Worse evaluation' of each criterion.
- Identification of the effects in data degradation of a criterion being 'Applicable but not reported' for a given ecotoxicological datum.
- Identification of the credibility and plausibility of a criterion.
- Identification of the robustness of the evaluation and the possible existence of disputable conditions.

During the workshop, the complete testing of the process was performed and each expert present had the possibility to express his/her insights on the six (6) aforementioned points/questions for each element of the assessment criteria hierarchy (see Table 11 in the Annex). Based on the fruitful discussions, the provided feedback from the present experts and the outcomes of the workshop, the questionnaire for expert consultation was slightly redesigned and adjusted to fit further the characteristics of ecotoxicological data. Specifically, the questions regarding (1) the 'synergic/redundant effects' of criteria and (2) the credibility of criteria were omitted as not applicable in the context of ecotoxicology and the question regarding the 'importance of criteria' was rephrased to include only the ranking regarding the 'Worse evaluation' of a criterion.

The final version of the 'Questionnaire for expert consultation' includes the following four points, for which the related questions and outputs are reported:

1. Identification of elements whose evaluation overrules (positively or negatively) other elements belonging to the same element group ('Over' and 'Veto' criteria).
   Q: *"Does an optimum (i.e. green answer), or conversely, worst (i.e. red answer) evaluation of one of the following criteria make all/some of the other criteria within the same category irrelevant?*"
   The output of the question can be a set of causal relations, in the form of 'IF-THEN' rules or a null set, in the cases where the expert does not define any relation.
2. Identification of the importance (ranking) of 'Worse evaluation' of each element.
   Q: *"Rank the importance of each criterion by assigning each of them to the appropriate category. Each criterion should be ranked, based on your judgment for its effects on data generation. For example, think about one test where all the criteria are optimum except to the criterion you are considering here. How would this worst answer degrade the test?"*
   The output of the question is a classification of each element to five (5) predefined classes (i.e. Prerequisite, Highly important, Moderately important, Slightly important, Not relevant).
3. Identification of the effects in data degradation of an element being 'Applicable but not reported' for a given ecotoxicological data.
   Q: *"Supposing a criterion is applicable for the type of test under assessment but not reported in the paper or not specified by the person evaluating the test's quality, which action would you take? Each criterion should be assigned, based on your judgment for its effects on data generation, in the right answer."*
   The output of the question is a classification of each element to three (3) predefined classes (i.e. Substituted by optimum, No idea on how to substitute, Substituted by worst).
4. Identification of the robustness of the evaluation and the possible existence of disputable conditions.
   Q: *"Evaluate if the Optimum/Worst answer is disputable (i.e. highly depend on the data assessor) or consensus-based (i.e. based on largely recognized assumptions/desired conditions). Each criterion should be assigned in the right answer, based on your judgment for its effects on data generation."*
   The output of the question is a classification of each element to two (2) predefined classes (i.e. Disputable, Undisputable).

It is important to notice that the final questionnaire design, includes the possibility for a user to skip the four questions for a given node of the criteria hierarchy, in case the user does not possess sufficient information or knowledge for evaluating the elements included in that node. This feature is designed for excluding from the evaluation the possible existing lack of knowledge, up to the highest possible percentage. For the completion of the questionnaire, a bottom-up approach is followed; therefore the user starts the evaluation from the criteria-questions level, continues with the criteria groups, the categories, the ecotoxicological categories and ends with the evaluation of the Lines of Evidence (LoEs).

The questionnaire was made available to the expert panel, through an online web application specifically developed, so that it could be submitted to and answered by any member of the panel. A total of 14 experts have provided their answers to the questionnaire, with the aim of: (i) identifying the possible existing relations among criteria, (ii) identifying the relative importance of each criterion and (iii) identifying the possible inherent uncertainty. The uncertainty could be expressed in the form of: (i) unreported information, (ii) disputable information and (iii) possible lack of knowledge of the experts.

The outputs that were gathered from the criteria ranking process are used as part of the MCDA-based aggregation procedure, with the purpose of identifying and scoring the reliability and relevance of the ecotoxicological data under assessment. The elaborated outputs constitute the knowledge base of the proposed methodology, which is used in the implementation of the methodology for the quantitative scoring of ecotoxicological data. The knowledge base constitutes the driving force of the MCDA based aggregation process, as it is described in the following section (3.2), and allows taking into consideration the insights and opinions of all the involved experts. Furthermore, the methodology allows the addition or removal of experts' input from the knowledge base, as it is designed to be modular, flexible and adjustable to the needs of the user/decision maker. Thus, the knowledge base can be expanded to be even more reliable in the future.

3.2. Introduction to the MCDA based aggregation methodology (Scoring system)

Hierarchically structured criteria allows the decomposition of complex decision making problems into smaller subtasks and is therefore attractive for users (Corrente et al., 2012). A great majority of methods designed for MCDA, assume that all evaluation criteria are considered at the same level, however, it is often the case that a practical application is imposing a hierarchical structure of criteria (Corrente et al., 2012).

Notable research has been performed on the application of MCDA methodologies with the use of hierarchically structured criteria. Recent publications include the application of the Multiple Criteria Hierarchy Process (MCHP) in Robust Ordinal Regression (Corrente et al., 2012), with ELECTRE and PROMETHEE (Corrente et al., 2013a) and for the Choquet Integral (Angilella et al., 2013), where authors deal with cases of decision making problems with indirect elicitation of preference information, outranking relations and interacting criteria respectively.

The proposed MCDA methodology builds up on the concept of MCDA and hierarchically structured criteria and suggests a process that is not based on one of the well-known MCDA methods but combines significant characteristics from various methods and concepts, as explained further below.

The proposed MCDA methodology is a vital part of the WoE framework that connects the different elements of the ecotoxicological data assessment process. It serves as the main connector between the analysed ecotoxicological data and the knowledge base, and thus allows the quantification of all the available information firstly for the analysis of the reliability and relevance of ecotoxicological data and secondly for the ranking of the data. Three figures are used to describe in further detail the proposed methodology. Specifically, they illustrate the information flow and the process that is followed for the implementation of the methodology (Figure 3), the connections between the various elements used in the methodology and the various actors (Figure 4) and the way the hierarchical aggregation of information is performed mathematically (Figure 5). In Figure 3, the blue part represents the process followed for the creation of the WoE assessment framework and the knowledge base, the grey part the process for the assessment of ecotoxicological data by a user, based on the WoE framework, and the light green the process for the application of the MCDA methodology. The background colours used correspond to the ones used also in Figure 4 for allowing the reader to connect the various instances of the methodology and provide a better

understanding of the involvement of various actors in the methodology. On the other hand, Figure 5 shows a graphical representation of the hierarchical aggregation techniques used.

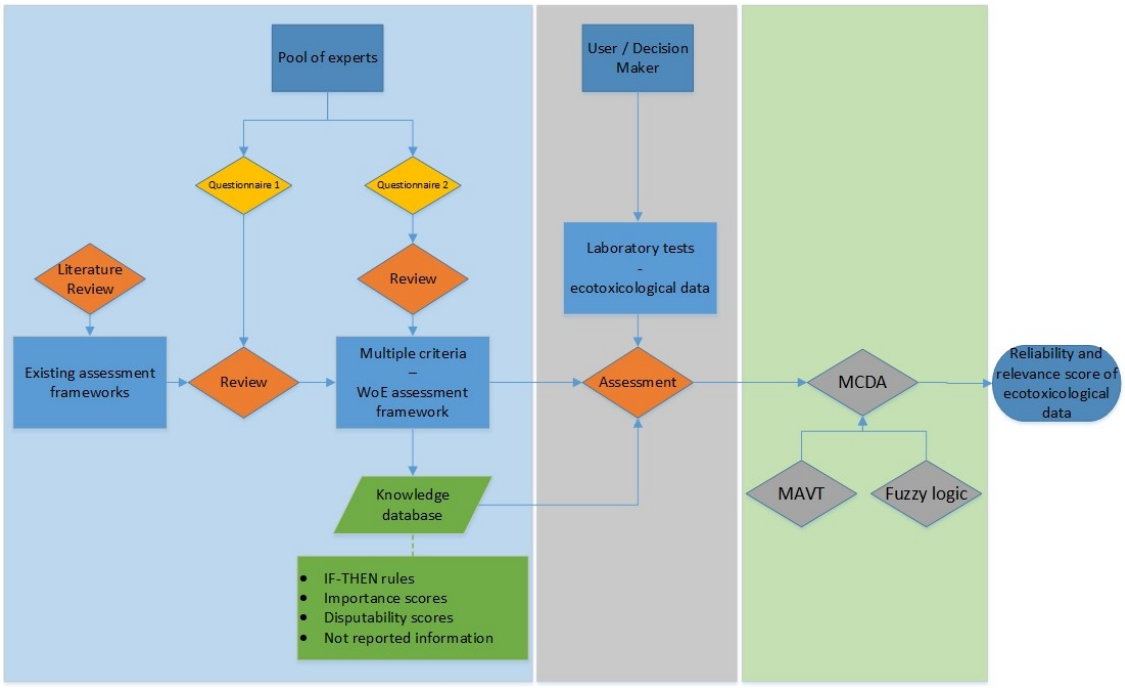


Figure 3 : Information flow and MCDA based aggregation methodology – conceptual representation

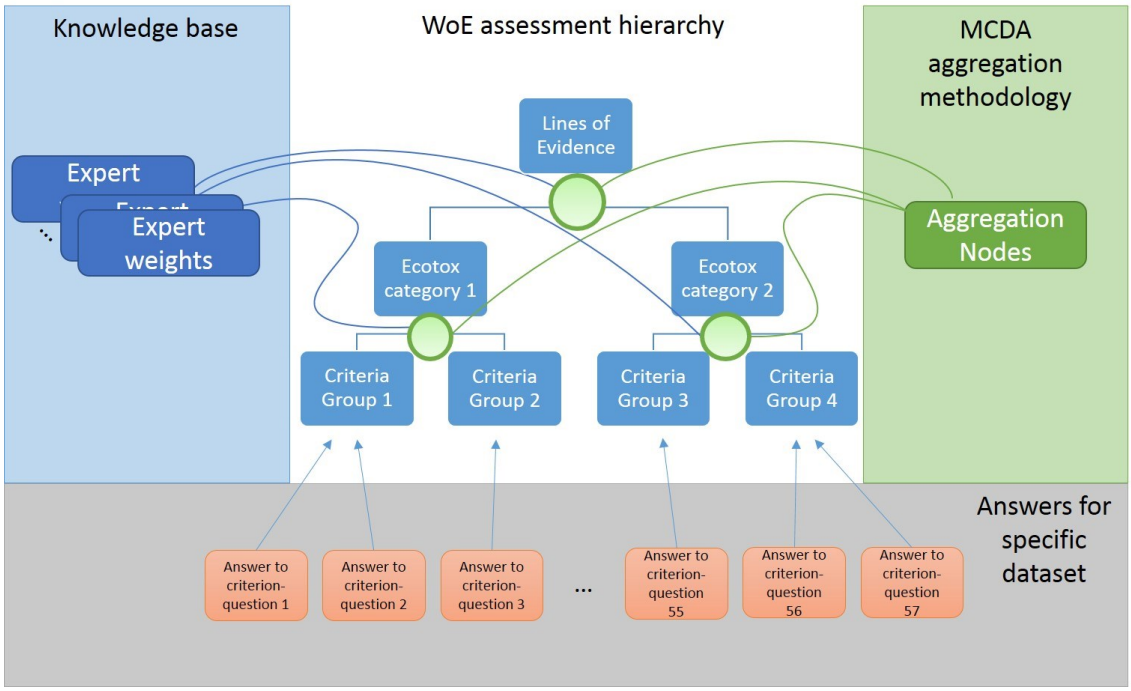


Figure 4: Connections between elements and actors in the application of the MCDA methodology

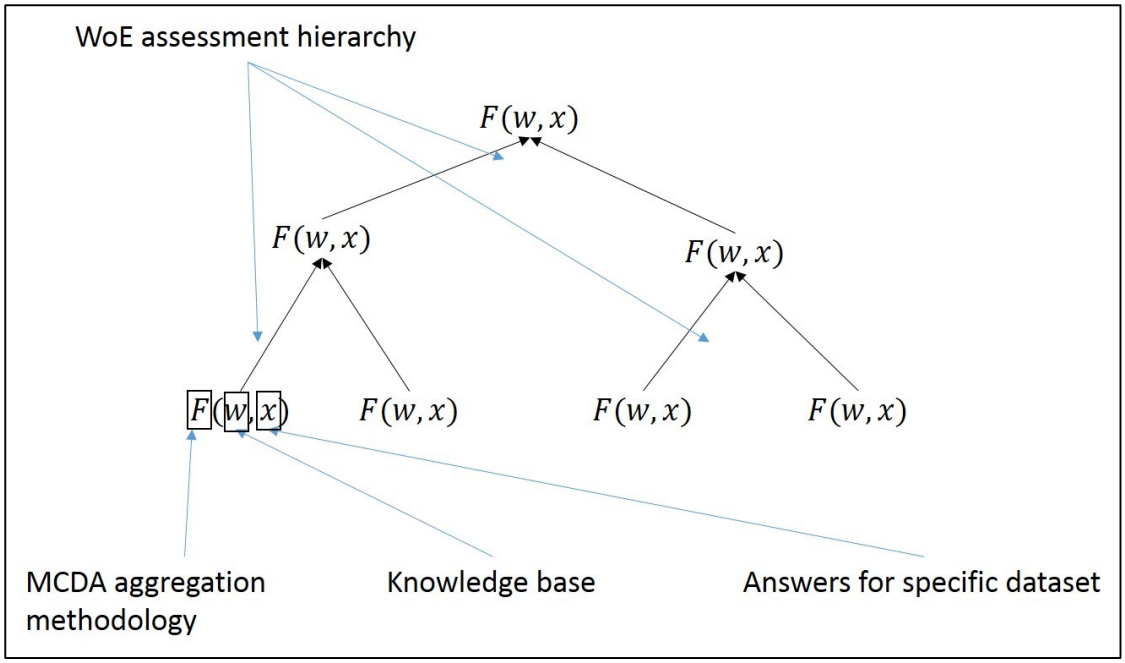


Figure 5: Illustration of the hierarchical mathematical aggregations

The methodology allows performing a systematic analysis of the diverse types of available information and using various complex algorithms for the calculation of a final reliability and relevance index for every ecotoxicological datum under assessment. It is an innovative process that has been designed and tailored to fit the characteristics of the available information in the context of ecotoxicology as well as to fill in the gaps in the state-of-the-art of the analysis of ecotoxicological data, as identified in the 'Introduction'. Specifically, the methodology makes use of all the available information and provides a transparent scoring system based on unambiguous multiple criteria, while the existence of hierarchical aggregations allows the detailed analysis and identification of the elements that influence the ranking of ecotoxicological data in a clear way.

The WoE framework and the proposed MCDA methodology have been developed based on the notions of the Multi-Attribute Value Theory (Keeney et al. 1976), since the concept of 'value functions' is used for the handling of the criteria values, and their transformations into certain scales, and a 'decision tree' has been designed and used as a foundation of the assessment process. The methodology is combining the value theory with elements of fuzzy logic. Fuzzy logic (Zadeh, 1965) introduces the concepts of 'partial membership' and 'degree of truth'. Namely, 'degree of truth' of an element x refers to the associated value in [0,1] of the value of the membership (characteristic) function for element x of a fuzzy set and 'partial membership' refers to the possibility of an element x to partially belong to a fuzzy set.

Various methodologies have been developed to handle multicriteria aggregation problems with the use of integrals (e.g. Choquet, 1954; Sugeno 1974), such as the research performed by Grabisch (1996) for the application of fuzzy integrals for criteria aggregation in decision problems. Due to the nature of the issue under assessment, the characteristics of ecotoxicological data and the absence of synergic/reduntant effects within the WoE criteria hierarchy (see paragraph 3.1.2) it has been deemed necessary to use a MCDA methodology that does not utilise fuzzy integrals and which is tailored to the decision problem at hand.

In order to use the proposed MCDA methodology the assessor has to analyse each ecotoxicological datum, by answering each of the 57 criteria-questions (multiple answer questions) of the WoE framework regarding the conditions and methods under which each ecotoxicological datum was created, as seen in the grey section of Figure 3 and Figure 4. The methodology allows coupling the answers to the criteria-questions with the contents of the knowledge base (as described in section 3.1.2 and seen in Figure 4) by the application of a set of aggregation algorithms and functions for the calculation of the reliability and relevance score as depicted in Figure 4 and Figure 5.

As mentioned in section 3.1.2, each of the questions used in the questionnaire for expert consultation provides specific types of outputs:

1. the relations among elements,
2. the relative importance of elements,
3. the disputability of elements, and
4. information on the handing of unreported data.

Outputs are expressed in different forms:

- relations among elements: a priority ordered set of causal relations, i.e. rules of the 'IF-THEN' form,
- the rest: unique element scores in the interval [0,1].

It is useful to describe the main way each output is incorporated and processed in the aggregation methodology for the analysis of one ecotoxicological datum. In Figure 6, an example of a priority ordered set of causal relations is given and the concepts of a criterion and its evaluation, a rule block, a rule outcome, a rule and a set of rule blocks are presented. For the evaluation of a set of rules, the following steps are followed: Firstly, the causal relations are divided into smaller blocks and evaluated for their validity, based on the answers of each criterion for the given datum. This is done by calculating the degree of truth of each rule block and then combining all the degrees of truth for calculating the degree of truth of each specific rule. Since the set of causal relations is priority ordered, the degree of truth of each rule receives a specific priority in the aggregation procedure.

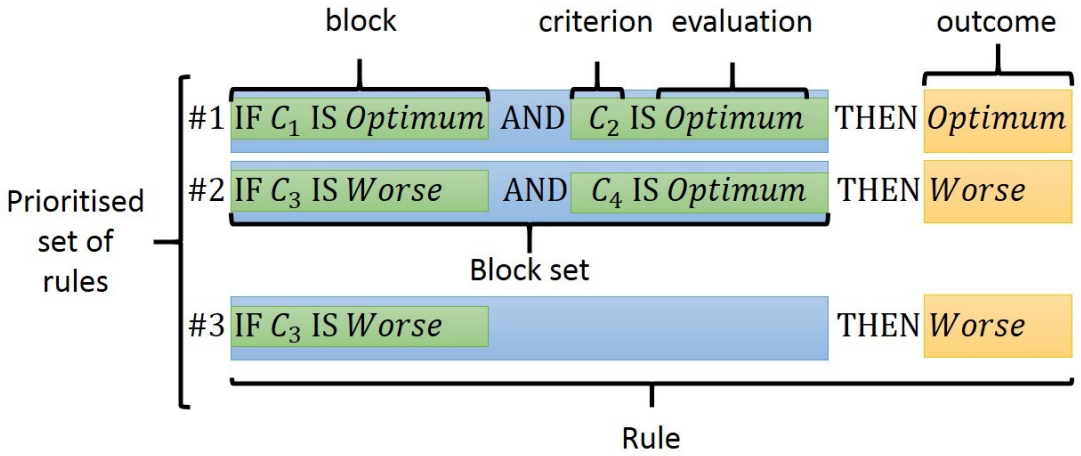


Figure 6: Example of a prioritised set of IF-THEN rules

The outputs of the other three questions are expressed as scores and presented in numerical forms. They are all used for evaluating how much each element evaluation could affect, or not, the production of the ecotoxicological datum and therefore how much it reduces the reliability and relevance of the datum. As they are numerical, no prior-elaboration is necessary.

The following paragraphs provide all the details regarding the way the MCDA based methodology is set up, the types of inputs used, the derived inputs throughout the various steps, the definitions of the aggregation functions and their implementations for the calculation of the overall reliability score of a given analysed ecotoxicological datum.

The MCDA aggregation methodology is based on various types of inputs and 10 distinctive functions, which are used in the methodology for the calculation of the score for one node j $(j = 1{:}\,38)$ of the criteria hierarchy. In an identical way, the scores are calculated for all the nodes of the hierarchy, in a bottom-up approach, that is concluded with the calculation of the total reliability and relevance score for every datum.

Starting from the first level of the hierarchy, each hierarchy criterion-question is denoted with i (i=1:57) and subsequently each Node with j (j=1:38). A Node j refers to a single branch of the hierarchical criteria structure (criteria group, category, LoE), starting from the lowest branch level (criteria groups) and moving upwards to the top of the hierarchy (LoE). For simplification reasons, an element of the aggregation scheme that refers either to a Criterion i or to a Node j is denoted with e (e=1:57).

The replies of the criteria-questions for a given ecotoxicological datum are represented and expressed in the aggregation methodology as $\mathrm{CA_i} \in \{\mathrm{Y, N, NA, NR, DK}\}$, for each criterion-question i $(\mathrm{i} = 1{:}\,57)$. The corresponding answers are Y: 'Yes', N: 'No', NA: 'Not Applicable', NR: 'Applicable but Not Reported', DK: 'I don't know'. These answers are, in subsequent steps of the aggregation, used as an input for the evaluation of the causal relations of the knowledge base and the evaluations of the criteria i.

In the framework, there is a specific correspondence between the answers to each criterion's question $(\mathrm{CA_i} \in \{\mathrm{Y, N, NA, NR, DK}\})$ and the numerical evaluation of each criterion, represented as $\overline{\mathrm{CA_i}} \in [0{,}1] \cup \emptyset$.

The criterion evaluation $(\overline{\mathrm{CA_i}})$ is calculated through a specific function which is described later in this section.

The input term Arrow $(\mathrm{Ar_i} \in \{\uparrow, \downarrow\})$, which is predefined in the framework by the expert panel, provides the correspondence between the answer to the response sheet for a criterion $(\mathrm{CA_i} = \mathrm{Y\ or\ N})$ and the 'Optimum' or 'Worse' status of each criterion in the framework. Formally, a function $\mathrm{F_6} : \mathrm{F_6(CA_i, Ar_i)} \rightarrow \{\mathrm{O, W}\}$ is used to define the criterion correspondence $\mathrm{A_i}$, such that:

$$\mathrm{A_i} = \begin{cases} \mathrm{O}, & (\mathrm{Y}, \uparrow) \vee (\mathrm{N}, \downarrow) \\ \mathrm{W}, & \mathrm{else} \end{cases} \qquad \text{(Eq. 1)}$$

The criterion correspondence $\mathrm{A_i}$ is an input to the membership function $\mathrm{F_5}{:}\,\mathrm{F_5(CA_i, A_i)} \rightarrow [0{,}1] \cup \emptyset$ which calculates the criterion evaluation score $\overline{\mathrm{CA_i}}$ and is defined as:

$$\overline{CA_\iota} = \begin{cases} \emptyset & , CA_i = NA \\ sub_i & , CA_i = NR \\ SC_{DK} & , CA_i = DK \\ SC_O & , CA_i \in \{Y, N\} \wedge A_i = O \\ SC_W & , CA_i \in \{Y, N\} \wedge A_i = W \end{cases} \qquad \text{(Eq. 2)}$$

Where:

- $\mathrm{sub_i}$ is the substitution score input when a criterion is applicable but not reported.

- $\mathrm{SC_{DK}}$ is the score when the response sheet answer for a criterion i is 'I don't know'.

- $\mathrm{SC_O}$ is the score when the criterion correspondence is Optimum and,

- $\mathrm{SC_W}$ is the score when the criterion correspondence is Worse.

Formally:
- $\mathrm{sub_i} \in \{0.1, 0.5, 0.9\}$, depending on the classification of a criterion in the knowledge base by the experts.

- $\mathrm{SC_{DK}} = 0.5$

- $\mathrm{SC_O} = 1$

- $\mathrm{SC_W} = 0$

It is important to first provide the definitions of the disputability and how it is handled in the framework. In detail, the disputability score of an element *e* is denoted and defined as $\mathrm{disp_e} \in \{0,0.3\}$. The score is based on the classification of each element as 'undisputable' or 'disputable', as provided by the expert panel and stored in the knowledge base.

The disputability score is incorporated in the calculation of the disputability of elements' evaluations through two dedicated functions, namely function $\mathrm{F_8}: \mathrm{F_8}(\overline{\mathrm{CA_i}}, \mathrm{disp_i}) \rightarrow [0,1] \cup \emptyset$ for the criteria evaluations $\overline{\mathrm{CA_i}}$ and function $\mathrm{F_9}: \mathrm{F_9}(\mathrm{SC_j}, \mathrm{disp_j}) \rightarrow [0,1] \cup \emptyset$ for the node evaluations $\mathrm{SC_j}$. The disputability of criteria evaluations is denoted with $\overline{\mathrm{CA_i}}^{\mathrm{disp}}$ and the disputability of node evaluations with $\overline{\mathrm{SC_j}}^{\mathrm{disp}}$. It is important to note that the formal definition of a node evaluation $\mathrm{SC_j}$ is deliberately described in the coming paragraphs of the chapter, due to the association of its calculation with the calculation of the relative importance of the criteria-questions of the specific Node.

The formal implementations of functions $\mathrm{F_8}$ and $\mathrm{F_9}$ are, respectively:

$$\overline{CA_\iota}^{disp} = \begin{cases} \emptyset & , \overline{CA_\iota} = \emptyset \\ \overline{CA_\iota} - (\overline{CA_\iota} - 0.5) * disp_i & , \overline{CA_\iota} \geq 0.5 \\ \overline{CA_\iota} + (0.5 - \overline{CA_\iota}) * disp_i & , \overline{CA_\iota} < 0.5 \end{cases} \quad \text{(Eq. 3)}$$

$$\overline{SC_j}^{disp} = \begin{cases} \emptyset & , SC_j = \emptyset \\ SC_j - (SC_j - 0.5) * disp_j & , SC_j \geq 0.5 \\ SC_j + (0.5 - SC_j) * disp_j & , SC_j < 0.5 \end{cases} \quad \text{(Eq. 4)}$$

For each Node *j* the evaluation is based on these disputability aware criteria evaluations $\overline{\mathrm{CA_i}}^{\mathrm{disp}}$ by the application of three steps: *i)* evaluation of priority ordered list of causal relations, in the form of 'IF-THEN' rules; *ii)* evaluation of relative importance of the node components; and *iii)* integration of causal and importance evaluations. In the following sections the definition of the three aforementioned steps are reported for the 2nd-lowest level of the hierarchy, namely criteria groups.

In order to proceed with the theoretical explanation, a general element evaluation $\mathrm{S_e}$ is defined as $\mathrm{S_e} \in \{\mathrm{O}, \mathrm{W}\}$.

3.2.1. Evaluation of priority ordered list of causal relations, in the form of 'IF-THEN' rules

A causal relation, as described in Figure 6, denoted as $R_{j,r}$ and hereby referred as Rule (R), is related with a specific Node j and identified through the subscript $r, \text{for } r \geq 1$. Each Rule defined by the experts consists of a set of blocks $B_{j,r}$, paired with the respective rule evaluation $S_r \in \{O, W\}$ (where O is Optimum and W is Worse).

Specifically,

$$R_{j,r} = \begin{cases} \{B_{j,r}, S_r\} & \text{, for criteria causal relations} \\ \{N_{j,r}, S_r\} & \text{, for node causal relations} \end{cases}$$

Formally a set of blocks $B_{j,r}$ is described as the collection of blocks $B_{j,r,k}$, therefore:

$$B_{j,r} = \{B_{j,r,k} | k \geq 1\} \text{ and } N_{j,r} = \{N_{j,r,l} | l \geq 1\}$$

The definition of a block $B_{j,r,k}$, as part of a Rule $R_{j,r}$ for a given Node j, includes a criterion i and the respective criterion evaluation S_i, therefore $B_{j,r,k} = \{i , S_i\}$ for $i = 1{:}57$, $j = 1{:}38$, $r \geq 1$ and $k \geq 1$.

Similarly, the definition of a node block $N_{j,r,l}$, as part of a Rule $R_{j,r}$ for a given Node j, includes a Node j and the respective Node evaluation S_j, therefore $N_{j,r,l} = \{j, S_j\}$ for $j = 1{:}38$, $r \geq 1$ and $l \geq 1$.

As mentioned above, a Rule $R_{j,k}$, as defined by the expert panel and stored in the knowledge base, for the criteria level is composed by sets of blocks and a respective evaluation ($R_{j,k} = \{B_{j,r}, S_r\}$) and a set of blocks is composed by blocks ($B_{j,r} = \{B_{j,r,k} | k \geq 1\}$). The evaluation of blocks, denoted as $BE_{j,r,k}$, is defined through the function $F_1 : F_1\left(B_{j,r,k}, \overline{CA_i}^{disp}\right) \rightarrow [0,1] \cup \emptyset$, whose inputs are the blocks $B_{j,r,k}$ and the disputability aware criteria evaluations $\overline{CA_i}^{disp}$. The formal implementation is defined by:

$$BE_{j,r,k} = \begin{cases} \emptyset & \text{when } \overline{CA_i}^{disp} = \emptyset \\ \overline{CA_i}^{disp} & \text{when } S_i = O \\ 1 - \overline{CA_i}^{disp} & \text{when } S_i = W \end{cases} \quad \text{(Eq. 5)}$$

It has to be noted that when a criterion is not applicable ($CA_i = NA$) the respective disputability criteria evaluation ($\overline{CA_i}^{disp}$) is represented as the empty set (∅) and therefore the block evaluation is represented equally.

Subsequently, the evaluation of a set of blocks for a given Rule $R_{j,k}$, called Rule evaluation is denoted as $\overline{R}_{j,r}$ and defined through the function $F_3(R_{j,r}) \rightarrow ([0,1], \{0,1\}) \cup \emptyset$.

Formally,

$$\overline{\mathrm{R}}_{\mathrm{j,r}} = \begin{cases} \emptyset & ,\forall\, \mathrm{BE}_{\mathrm{j,r,k}} = \emptyset \\ \left(\min\left(\mathrm{BE}_{\mathrm{j,r,k}}\right), 1\right) & ,\exists \mathrm{BE}_{\mathrm{j,r,k}} \neq \emptyset \wedge \mathrm{S_r} = \mathrm{O} \\ \left(\min\left(\mathrm{BE}_{\mathrm{j,r,k}}\right), 0\right) & ,\exists \mathrm{BE}_{\mathrm{j,r,k}} \neq \emptyset \wedge \mathrm{S_r} = \mathrm{W} \end{cases} \quad \text{(Eq. 6)}$$

The described function is used for the evaluation of all the present causal relations, as $\mathrm{r} \geq 1$.

3.2.2. Evaluation of relative importance of the node components

The second important part of the knowledge base is the info regarding the relative importance of criteria. For the calculation of the relative importance of a criterion, the methodology takes into consideration two variables: The disputability aware criteria evaluation ($\overline{\mathrm{CA}_\mathrm{i}}^{\mathrm{disp}}$) and the importance score of a criterion.

The importance score of an element, $\mathrm{im_e} \in [0,1]$, is an input provided by the expert panel from the knowledge base.

As such, the relative importance of criteria, which are part of a specific Node j, is denoted as $\mathrm{Q_{2,j}}$ and is calculated through the function $\mathrm{F_7}$: $\mathrm{F_7}\left(\overline{\mathrm{CA}_\mathrm{i}}^{\mathrm{disp}}, \mathrm{im_i}\right) \rightarrow [0,1] \cup \emptyset$. Formally:

$$\mathrm{Q_{2,j}} = \begin{cases} \emptyset & ,\forall\, \overline{\mathrm{CA}_\mathrm{i}}^{\mathrm{disp}} = \emptyset \\ \dfrac{\sum_{\mathrm{i=1}}^{\mathrm{n}} \overline{\mathrm{CA}_\mathrm{i}}^{\mathrm{disp}} + \left(1 - \overline{\mathrm{CA}_\mathrm{i}}^{\mathrm{disp}}\right) \cdot (1 - \mathrm{im_i})}{\mathrm{n}} & ,\text{else} \end{cases} \quad \text{(Eq. 7)}$$

3.2.3. Integration of causal and importance evaluations

The Sufficiency level of an expert for a given Node j can take the values 0 or 1 and is defined as

$$\mathrm{Sl_j} \in \{0,1\} \quad ,\mathrm{j} = 1{:}\,38$$

Values $\overline{\mathrm{R}}_{\mathrm{j,r}}$ for each rule and relative importance of criteria $\mathrm{Q_{2,j}}$ have been defined therefore we can proceed with the definition of the function describing the calculation of the node score ($\mathrm{SC_j}$).

Specifically, given the Sufficiency level $\mathrm{Sl_j}$, which is taken from the knowledge base and represents whether an expert has provided feedback on the 'Questionnaire for expert consultation' for the Node j under evaluation or not, the score of a Node j is calculated through the application of function $\mathrm{F_4}$: $\mathrm{F_4}\left(\overline{\mathrm{R}}_{\mathrm{j,r}}, \mathrm{Q_{2,j}}, Sl_j\right) \rightarrow [0,1] \cup \emptyset$ defined as:

$$\mathrm{SC_j} = \begin{cases} \emptyset & \mathrm{Sl_j} = 0 \\ \sum_{i=1}^{n} a_i \cdot \bar{R}_{j,i}^2 + a_{n+1} \cdot Q_{2,j} & \mathrm{Sl_j} = 1 \end{cases} \quad \text{(Eq. 8)}$$

where for each Rule evaluation $\overline{\mathrm{R}}_{\mathrm{j,r}}$, that has an output of the form $([0,1],\{0,1\})$, its two elements are denoted as $\overline{\mathrm{R}}_{\mathrm{j,r}}^1$ and $\overline{\mathrm{R}}_{\mathrm{j,r}}^2$ (the former identifies the degree of thruth of the rule while the latter identifies the associated outcome), $r = 1{:}\,n$, $a_0 = 0$, $\bar{R}_{j,n+1}^1 = 1$ and $a_i = \left[1 - \sum_{k=0}^{i-1} a_k\right] \cdot \bar{R}_{j,i}^1 \quad ,with\ i > 0$. The idea behind this formula is that rules scores and relative importance are aggregated by weighted average, with weights based on degree of truth and decreasing priority with relative importance having the lowest weight.

The process described till now provides the necessary background and the step by step calculation of the score of a Node j of the first (lowest) hierarchy level, the criteria groups. As a natural step in the aggregation methodology, the functions that allow the evaluation of the rest of the framework hierarchy in the bottom-up process can now be defined. Though, the process is identical with the one followed above and only some notations need to be adjusted to correspond to the respective hierarchy levels. Specifically, the functions which are used to define the node block evaluations (F_2), the evaluations of Rules on node level (F_3) and the relative importance of Nodes (F_7) need to be adjusted. Their implementations are described below:

$$NE_{j,r,l} = \begin{cases} \emptyset & when\ \overline{SC_J}^{disp} = \emptyset \\ \overline{SC_J}^{disp} & ,when\ S_j = O \\ 1 - \overline{SC_J}^{disp} & ,when\ S_j = W \end{cases} \quad \text{(Eq. 9)}$$

$$\bar{R}_{j,r} = \begin{cases} \emptyset & ,\forall\ NE_{j,r,l} = \emptyset \\ \left(\min\left(NE_{j,r,l}\right), 1\right) & ,\exists NE_{j,r,l} \neq \emptyset \wedge S_r = O \\ \left(\min\left(NE_{j,r,l}\right), 0\right) & ,\exists NE_{j,r,l} \neq \emptyset \wedge S_r = W \end{cases} \quad \text{(Eq. 10)}$$

$$Q_{2,j} = \begin{cases} \emptyset & ,if\ all\ \overline{SC_J}^{disp} = \emptyset \\ \dfrac{\sum_{j=1}^{n} \overline{SC_J}^{disp} + \left(1 - \overline{SC_J}^{disp}\right) * (1 - im_j)}{n} & ,otherwise \end{cases} \quad \text{(Eq. 11)}$$

Based on the proposed aggregation methodology, it is possible to calculate the overall reliability score of a given ecotoxicological datum, denoted as $\mathrm{SC}_{38} \in [0,1]$. It is important to clarify that this reliability score is associated with a specific member of the expert panel, as the score is extracted by using the input provided by that expert during the creation of the knowledge base. By following the same procedure, it is possible to calculate a specific number of reliability scores for a given ecotoxicological datum, based on the number of members that constitute the expert panel.

For a given number of experts $\mathrm{M} = 1{:}\,\mathrm{m}$, we denote with $\mathrm{X} = \{\mathrm{SC}_{38}^{1}, \mathrm{SC}_{38}^{2}, \dots, \mathrm{SC}_{38}^{\mathrm{m}}\}$ the set of reliability scores of an ecotoxicological datum and $\mathrm{W} = \{\mathrm{w}_1, \mathrm{w}_2, \dots, \mathrm{w}_{\mathrm{m}}\}$ the set of the respective weights, each associated with an expert. Each weight is in $[0,1]$.

The calculation of the total reliability score of an ecotoxicological datum is based on the application of the weighted average and is defined as:

$$\mathrm{T}_{\mathrm{SC}} = \frac{\sum_{\mathrm{M}=1}^{\mathrm{m}} \mathrm{SC}_{38}^{\mathrm{M}} \cdot \mathrm{w}_{\mathrm{M}}}{\sum_{1}^{\mathrm{m}} \mathrm{w}_{\mathrm{M}}} \quad \text{(Eq. 12)}$$

3.3. Criteria Hierarchy of the WoE framework

As described in paragraph 3, the proposed framework is a Weight of Evidence framework, which allows the analysis of ecotoxicological data based on a hierarchically structured set of criteria and the use of the innovative MCDA-based aggregation methodology. The details of the WoE framework and the related assessment criteria are described and shortly discussed below, while the complete hierarchical criteria structure is presented in the Annex, Table 11.

3.3.1. Definition of Lines of Evidence (LoEs)

At the first level of the hierarchical structure, it is proposed to evaluate ecotoxicological data according to three Lines of Evidence:

1. Experimental Reliability LoE. Experimental reliability covers the inherent quality of a test relating to test methodology and the way the performance and results of the test are described.
2. Statistical Reliability LoE. Statistical reliability covers the inherent quality statistical methodology and assumptions chosen for interpreting experimental results and the way the performance and results of the statistical analysis are described.
3. Biological relevance LoE. Biological relevance covers the extent to which a biological endpoint is appropriate for a particular risk assessment. This category includes the extent to which a test is appropriate for a particular substance, given prior knowledge about its mode of action, and for a particular site, given prior knowledge about physico-chemical conditions and biological characteristics (e.g. habitat typology).

These three LoEs are the 'mandatory' block of the Weight-of-Evidence (WoE) framework. They can be informed once and then reused for each assessment made under the same conditions (e.g. if more than one substance is assessed during the same experiment). Some criteria can be informed or not, according to the assessment context (i.e. prior knowledge or not on site conditions). At the lowest level of the hierarchical structure, specific questions were defined and used for the weighting process further developed. The questions could be answered either by 'Yes', 'No', 'Not applicable', 'Applicable but Not reported' or 'I don't know'.

The complete set of LoE, categories, criteria groups and associated criteria-questions is reported in the annex, Table 11, and discussed shortly below.

3.3.1.1. The Experimental Reliability LoE

It is proposed to structure information related to the 'Experimental Reliability' LoE into four categories: 1. 'Quality assurance' category; 2. 'Substance identification and monitoring' category; 3. 'Organisms culture' category; 4. 'Test design' category.

'Quality assurance' category

'Quality assurance' has first to be defined because it highly differs according to frameworks. For example, according to Breton et al (2009), 'Quality Assurance' refers to all the criteria that they included in their system. Most authors defined 'Quality assurance' in a more restricted meaning, i.e. referring to the compliance to standards and/or Good Laboratory Practices

(GLP). Some frameworks do not include 'Quality assurance' in their list of criteria. For example, Schneider et al (2009) indicated that ToxRTool developers preferred to treat all kind of data equally and to base the decision on the data reliability using only information provided in the study report. Thus, the ToxRTool does not specifically consider whether data were obtained in compliance with recent guidelines and under GLP conditions. Similarly, Hobbs' scheme does not refer to guidelines, normative or GLP. Instead, the eco-QESST system (Breton et al, 2009) submits detailed questions related to standards and GLP and weights affected in case of positive answer are the maximum weights of the system, showing that the eco-QESST developers consider this criterion as one of the most important. As stated by Schneider et al (2009), this latter option however introduces bias in reliability assessment because, for old data, guidelines and/or GLP were not necessarily available at the time the experiment was conducted. Considering this (contradictory) background, we proposed to consider two criteria related to Quality Assurance category, related to guidelines/standards and GLP respectively.

'Substance identification and monitoring' category

Many of the schemes analysed (Klimisch et al. 1997; US EPA 1991; Hobbs et al. 2005; Schneider et al. 2009) defined a set of questions related to the substance identification/characterisation. Schneider's scheme is the most detailed one and it is thus proposed to build a similar structure for our methodology. Chemical monitoring during the test period is also mentioned by Hobbs' scheme ('Was the chemical concentration measured?'), but the scoring process is purely dichotomic (i.e. Maximum score if the substance is measured and minimum score if it is not measured), while the relevance of measuring the substance during the test period can also depend on the substance itself, e.g. on its ability to be degraded or lost by any other process (e.g. volatilization). Such interactions between criteria have been considered in the process of extracting the experts' knowledge.

'Organisms' category

The schemes of Schneider (2009), Breton (2009) and Ågerstrand (2011) included significant criteria related with organisms (i.e. species identification and organisms' physiology). Those were proposed to be included in the developed criteria hierarchy together with the criteria related with the culture design, as well as the acclimatation and feeding of organisms.

'Test design' category

'Test design' refers to all the experimental conditions that must be satisfied for guarantying the reliability of the test, i.e. the occurrence of negative and/or positive controls, the monitoring of important physico-chemical characteristics (e.g. temperature), the mode and route of contamination. The proposed framework includes three different criteria groups for test design, i.e. 'Controls', 'Physico-chemical conditions' and 'Exposure conditions' sub-categories.

3.3.1.2. The Statistical Reliability LoE

It is proposed to structure information related to the 'Statistical Reliability' LoE into three categories: 1. 'Test design'; 2. 'Assumptions'; 3. 'Estimation quality'.

'Test design' category

'Concentration design', 'Replicates' and 'Replicates Individual numbers' are the aspects of the test design that are considered important and relevant for the statistical reliability of an ecotoxicological test. Criteria related with the aforementioned characteristics have been identified and included in the assessment framework. In detail, the choice of concentrations, both the number of them and their value, affects the precision of LC/EC or NOEC/LOEC estimates. Guidelines currently often require four or five concentrations that are geometrically spaced, in addition to an untreated control. Even if these recommendations are conditional upon certain aspects of the tests, criteria are proposed to check that data actually respect them. Furthermore, criteria from Breton's (2009) and Ågerstrand's (2011) frameworks regarding the replicates have been identified as important and taken into consideration.

'Assumptions' category

Ecotoxicological data can be analysed either through hypothesis-testing or regression techniques. Assumptions are required in both cases and related criteria are proposed in this category. Information related to power and precision are important for assessing statistical reliability of hypothesis-testing methods. When hypothesis-testing, several assumptions need to be considered: independence of errors, normality of errors, and homogeneity of variance between treatments. Homogeneity of variance is the main assumption which must be satisfied. On the other hand, for the regression techniques, the model selection, the model comparison and the prior assumptions are evaluated and assessed.

'Estimation quality' category

An advantage of the regression approaches is the capability to provide a confidence interval for the calculated summary statistics (e.g. LC50). To be able to check the precision of the calculation, it is necessary that this information is available in the report/publication under analysis. The 'estimation quality' category includes criteria that are selected for assessing if an appropriate model has been selected for adequately fitting the set of data and if the quality of fit and/or comparison of models and/or justification of a given model are useful for assessing the quality of the summary statistics.

3.3.1.3. The Biological Relevance LoE

It is proposed to structure information related to the 'Biological Relevance' LoE into four categories: 1. 'Duration'; 2. 'Test design'; 3. 'Biological endpoint'; 4. 'Organism relevance'.

'Duration' category

According to the REACH guidance published by ECHA (ECHA, 2008a) 'Chronic toxicity' related to waterborne exposure refers to the potential or actual properties of a substance to cause adverse effects to aquatic organisms during exposures which are determined in relation to the life-cycle of the organism. Duration is then a key factor for determining whether a test can be considered as chronic or not and for identifying potential bias. Four criteria were proposed to check whether the test duration is optimal or not, for acute and chronic tests respectively.

'Test design' category

This category combines criteria, which were not included in the existing frameworks found in the literature, regarding the physicochemical conditions of the tests under assessment and an updated version of the relevance criterion of Ågerstrand's (2011) framework.

'Biological endpoint' category

For some biotests, several biological endpoints can be followed (e.g. for fish long term toxicity: hatching success, growth and/or survival; for fish embryo-larval test: several development effects like abnormally coiled tail or flexed tail with reduced length, small and disorganised trunk or inhibition of trunk morphogenesis, absence or malformation of ocellus, etc.; for invertebrates: long term toxicity, reproduction, growth and/or lethality; for macrophytes, fresh weight change, yield of energy conversion at photosystem; etc.). When several endpoints can be estimated, it would be necessary to justify whether the reported endpoint is actually appropriate for the risk assessment or not. In this context, seven criteria were proposed that were related with the summary statistics, the sensitivity of endpoints and the population relevance (as expressed also in Ågerstrand's framework).

'Organism relevance' category

The criteria included in this category answer some questions related to species relevance in case of site-specific assessments. Species that are tested in laboratory are indeed generally selected mainly for their ease of breeding and handling (species should be readily available year-round and tolerate handling and laboratory conditions) and their ability to be used as usual standards (allowing a good reproducibility of biological responses to toxicants). However, such test species do not necessarily relate, both phylogenetically and ecologically, to the organisms that naturally dominate natural ecosystems. Species can be described by a list of biological and ecological traits (e.g. life-cycle patterns, reproductive features, food/feeding habits, habitat preferences, etc.). Those traits are related to exposure (routes and levels) and toxicity mechanisms (contamination and elimination kinetics as well as bioaccumulation) and they can thus be used for evaluating the relevance of lab species for a given risk assessment context.

3.4. Statistics of the knowledge database of the AMORE DSS

The paragraph includes the main statistics and information which have been collected from the analysis of the input provided by the expert panel, regarding the insights of the experts on which criteria are important for the analysis of ecotoxicological data and which elements of the criteria hierarchy are considered the most influential.

An analysis of the input provided by the experts, allows the extraction of interesting statistics and information regarding the insights of the experts on which criteria are important for the analysis of ecotoxicological data and which elements of the criteria hierarchy are considered the most influential. A wealth of information can be extracted from the database, thus in the following paragraphs the most relevant graphs and statistics are presented and shortly discussed.

Firstly, an overview of the provided IF-THEN rules by the experts is presented. In Figure 7 the percentage of experts that have provided rules for each level of the criteria hierarchy is presented, information from which we can identify which nodes of the criteria hierarchy are more influential, either positively or negatively, in the evaluation of ecotoxicological data according to the experts. In detail, the most influential nodes are '1.2.1 – Substance identification' , '1.4.1 – Controls', '1.2 – Substance identification and monitoring' and '1.4 – Test design', followed by nodes '1.3.1 – Organisms identification and physiology', '2.1.2 – Replicates', '1 – Experimental Reliability' and '0 – Laboratory biotests'.

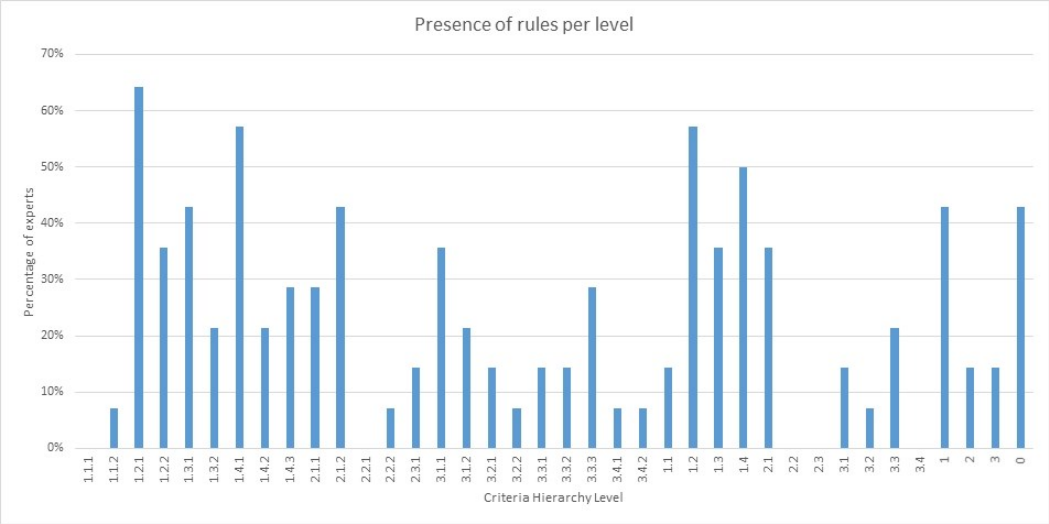


Figure 7: Presence of IF-THEN rules per criteria hierarchy level

Furthermore, in Figure 8, the number of elements per each provided rule is presented. The majority of experts have identified rules with up to two elements (therefore including up to two rule blocks) as the most representative for the performed evaluation as seen in Figure 8, since 104 rules with a single element and 38 with two elements have been provided. The use of rules with up to two elements indicates a slightly conservative approach expressed by the expert panel, due to the fact that single elements can influence significantly the performed evaluation.

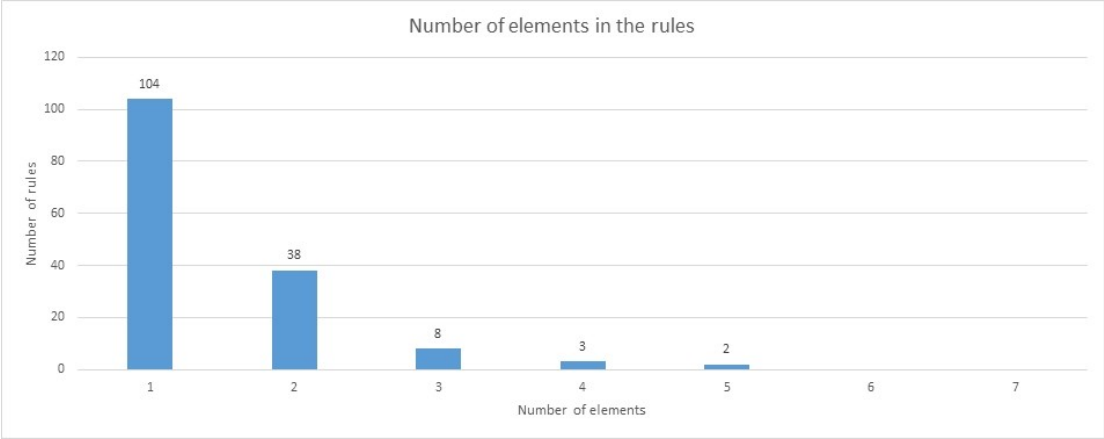


Figure 8 : Number of elements per IF-THEN rule

The following figures (Figure 9, Figure 10 and Figure 11) provide a concise analysis of the influence of specific criteria (and nodes) in the analysis of ecotoxicological data, through their appearance in rules provided by the expert panel. Each graph provides a summary of the number of times each element (i.e. criterion or node) appears in the provided IF-THEN rules, and shows how many times the element has a positive/negative influence, as well as how many times each element has been included in single element rules. Figure 9 shows the statistics for the 'Experimental reliability' LoE, Figure 10 the 'Statistical reliability' LoE and Figure 11 the 'Biological relevance' LoE. Based on these graphs, we identify the following criteria as the most influential ones:

- 1.2.1.1 – Substance identity
- 1.4.1.1 – No toxicant control
- 2.1.2.1 – Replicates
- 2.1.1.2 – Concentration spacing
- 2.1.2.2 – Individual numbers
- 3.1.1.1 – Acute duration
- 3.3.3.1 – Population dynamics

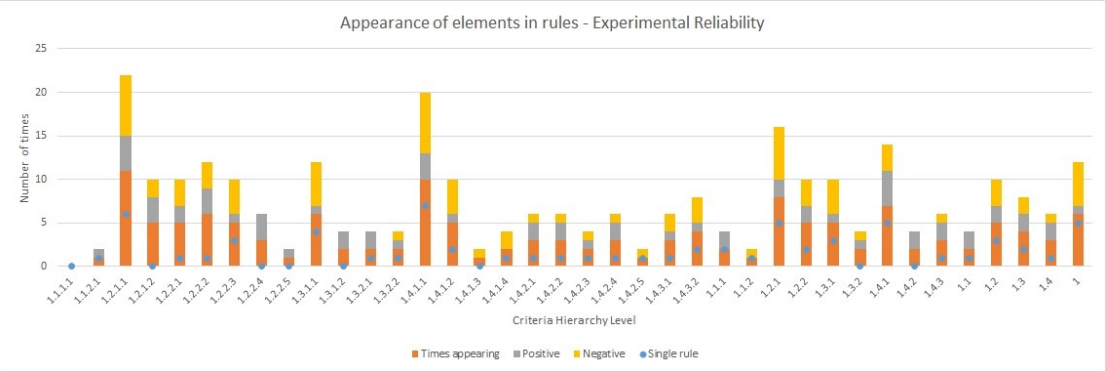


Figure 9 : Experimental reliability – Appearance of elements in IF-THEN rules

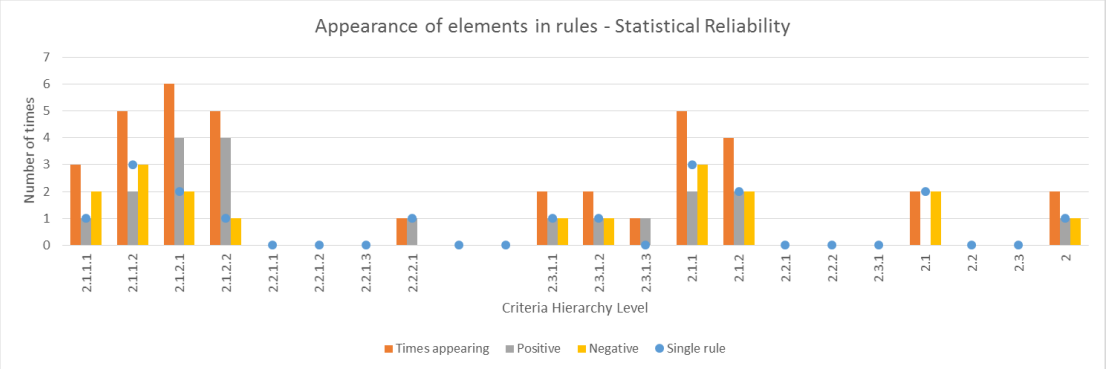


Figure 10 : Statistical reliability – Appearance of elements in IF-THEN rules

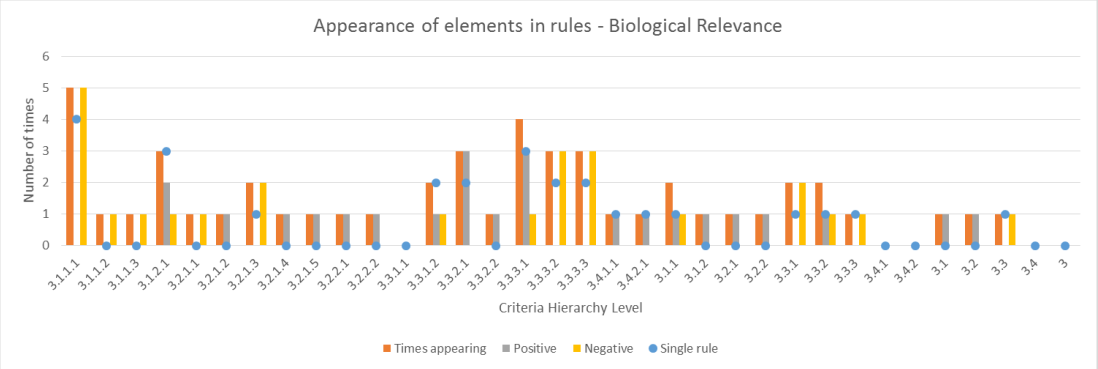


Figure 11 : Biological relevance – Appearance of elements in IF-THEN rules

Secondly, an overview of the importance of the criteria hierarchy elements is provided in Figure 12 (criteria), Figure 13 (criteria groups) and Figure 14 (ecotoxicological categories and LoE). The figures provide an overview of the classification of elements into the five predefined categories (from 'Prerequisite' to 'Not relevant') for every element and the importance a non-optimum evaluation may have in the assessment of ecotoxicological data. As it can be seen, the criteria with the highest number of classifications as 'Prerequisite' (in light red colour) and 'Highly important' (in yellow colour) are '1.2.1.1 – Substance identity', '1.2.2.2 – Concentration monitoring', '1.2.2.3 – Loss acceptability', '1.3.1.1 – Species identity', '1.4.1.1 – No toxicant control', '1.4.1.2 - 'No toxicant control', '1.4.2.1 – Temperature', '1.4.2.2 – pH', '1.4.3.1 – Exposure system', '1.4.3.2 – Exposure route', '2.1.1.1 – Concentration number', '2.1.2.1 – Replicates' and '3.1.1.1 – Acute duration'.

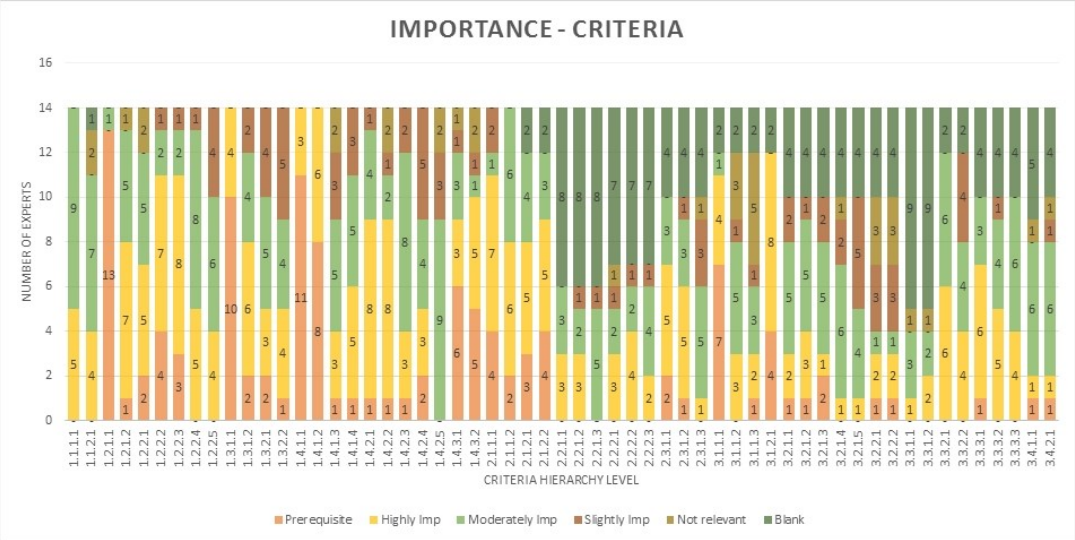


Figure 12: Importance of assessment criteria

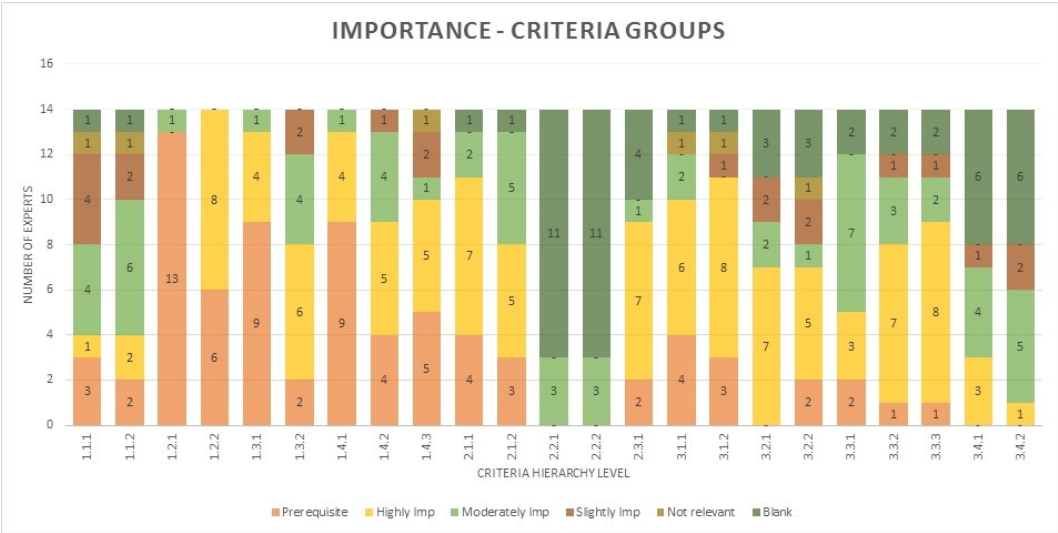


Figure 13 : Importance of assessment criteria groups

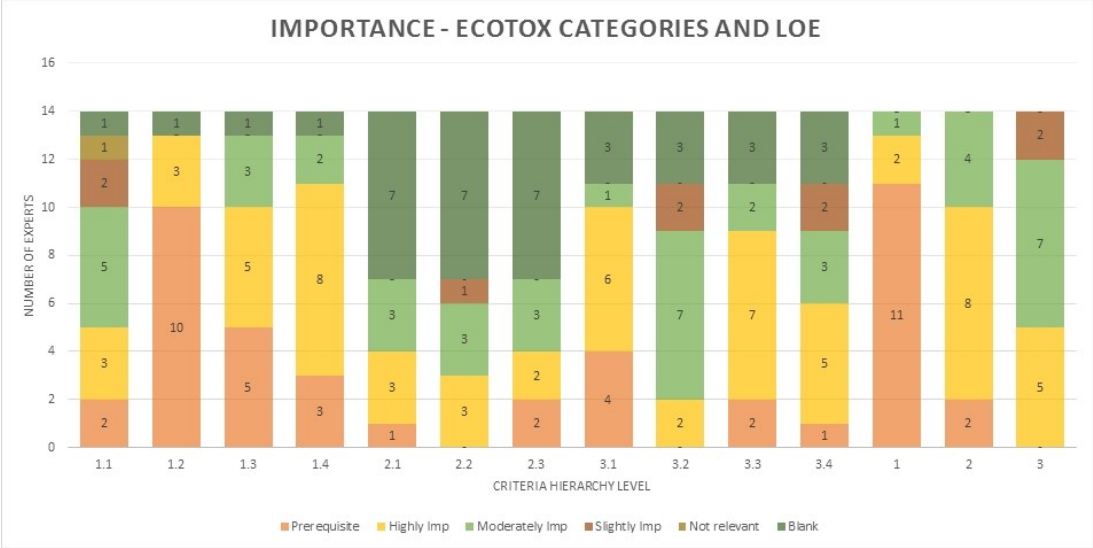


Figure 14 : Importance of assessment ecotoxicological categories and LoE

Similarly, for the criteria groups the most important according to experts' inputs are '1.2.1 – Substance identification', '1.2.2 – Substance loss and monitoring', '1.3.1 – Organisms identification and physiology', '1.4.1 - Controls', '1.4.3 – Exposure conditions', '2.1.1 – Concentration design', '3.1.1 – Acute test relevance' and '3.1.2 – Chronic test relevance'.

Lastly, the most important ecotoxicological categories are '1.2 – Substance identification and monitoring', '1.3 – Organisms', '1.4 – Test design' and '3.1 – Duration' and the most important LoE are '1 –Experimental reliability' and '2 – Statistical reliability'.

4. The AMORE Decision Support System

The AMORE Decision Support System is developed as part of the AMORE research project and consists of three modules which aim in assisting environmental researchers and experts in assessing environmental risks of chemicals in aquatic systems. To this end, it provides a set of tools for analysing and integrating both exposure and effect information (i.e. modelling as well as experimental data). The complexity of the topic outlines the necessity of the development of a DSS that surpasses the single analysing capabilities of humans.

The three modules of the DSS, namely the 'Exposure Assessment', the 'Effect Assessment' and the 'Risk Assessment' modules, are interactive and complete each other. They perform three different processes for evaluating, as final output of the DSS, the risk for species living on a given contaminated aquatic system in terms of Potentially Affected Fraction (PAF, Traas et al. 2002).

The objective of this section is to describe in detail the functionalities and interface of the AMORE DSS. Specifically, chapter 4.1 includes the details of the DSS framework and the model development, while chapter 4.2 includes the details of the organisation of the DSS into three modules and the characteristics of each module.

The DSS is implemented in an Excel spreadsheet environment programmed through Visual Basic for Applications (VBA) and is developed as independent software.

4.1. Framework and model development

The AMORE DSS is built to support a probabilistic risk assessment approach in a reliable, fast and reproducible way by using the rich calculating possibilities that are offered through the use of Information Technologies (i.e. computers). It is a tool aiding in increasing the productivity and efficiency of risk assessment process (where precision and optimality are essential), and allows the integration of various sources of information for aiding the process of structuring decisions.

The DSS consists of three modules, which are interconnected as depicted in Figure 15:

- The Exposure Assessment module
- The Effect Assessment module
- The Risk Assessment module

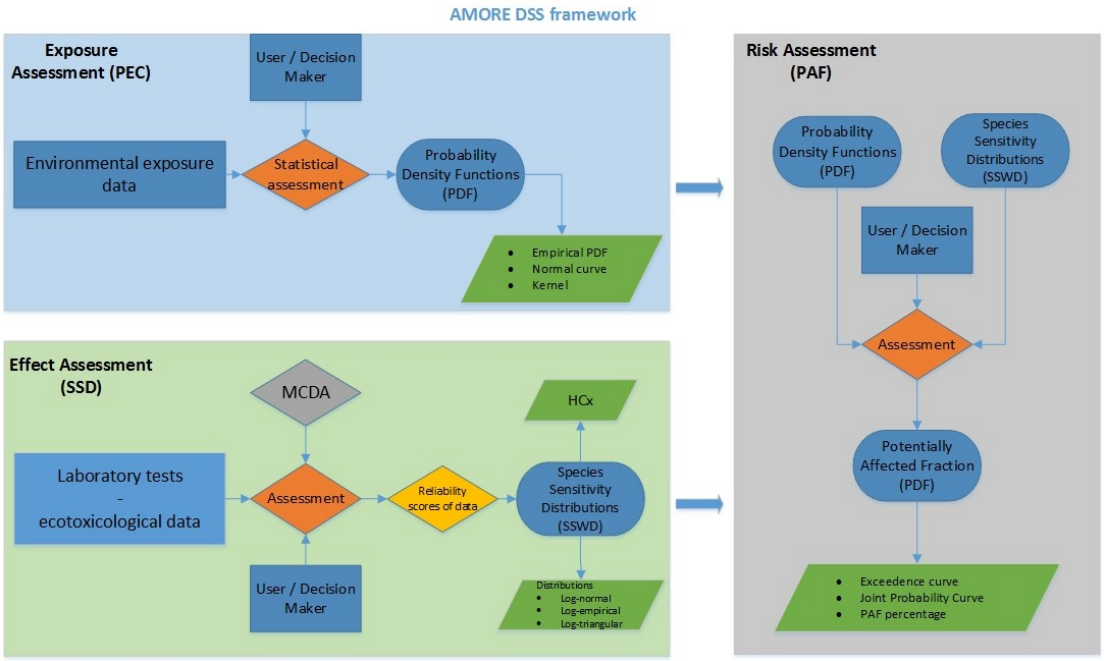


Figure 15: The AMORE DSS framework

The first module is developed with the use of VBA embedded into an EXCEL spreadsheet environment for the 'Exposure Assessment' and the estimation of the Predicted Environmental Concentrations (PEC) of pollutants via the production of their Probability Density Functions (PDF). It is described in detail in paragraph 4.2.1.

The second module is developed with the use of PHP and an EXCEL spreadsheet environment with VBA procedures for the 'Effect Assessment', through the evaluation of the reliability of ecotoxicological data. The reliability is estimated with the combination of use of the 'questionnaire for expert consultation' and the application of the MCDA methodology for the production of weighted Species Sensitivity Distributions (SSD). The module is described in detail in paragraph 4.2.2.

The third module is combining the results of the previous two modules into an EXCEL spreadsheet environment with VBA procedures for the conduction of the final 'Risk Assessment', through the production of the Joint Probability Curve (JPC) of the PEC and SSD graphs and the calculation of the Potentially Affected Fraction (PAF) index, as an estimation of ecological risk. It is described in detail in paragraph 4.2.3. A scheme of the entire software scheme is presented in Figure 16.

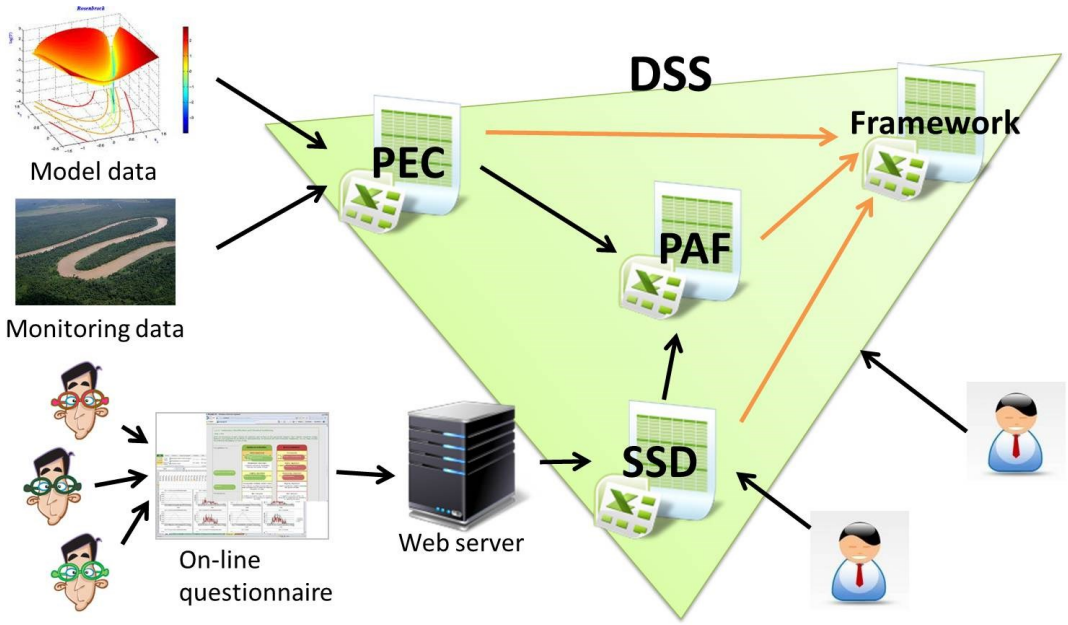


Figure 16: The general scheme of the AMORE DSS module applications

4.2. Modules

4.2.1. Module 1: Exposure Assessment (Predicted Environmental Concentration – PEC)

The design of the 'Exposure Assessment' module is based on the combination of empirical sampling techniques, mathematical modelling and statistical methods. The module is designed for the production of 'Probability Density Functions' (PDF) of contaminants, representing the probability of appearance of various concentration levels, based on sampling and taking into account missing values and their statistical substitutions. The concentration measurements over time are either collected through monitoring sampling or generated with the use of mathematical models. Both methodologies supply measurements at different times, in order to produce a temporal series of data. The time series of contamination measurements are used to calculate the contamination level Probability Density Function (PDF) over time. Obtaining a PDF from a set of measured data is a statistical issue, which can be treated by means of different methods, and can be represented in a graphical way by a curve on a Cartesian plane. The PDF is a useful instrument to estimate the distribution of a given substance in the environment.

Undetected values, which are related with the physical limitations of the analytical instruments, are a known problem of the sampling techniques. The substitution of the missing, undetected values can be treated with various statistical methods such as: deletion,

simple substitution, distributional methods and distributional-based imputation (Baccarelli et al., 2005).

In the AMORE DSS, the following methods are included and implemented in the software:

A. Simple substitution of non-detects with the following options related with the sampling Detection Limit (DL):
    a. 0
    b. $\frac{DL}{2}$
    c. $\frac{DL}{\sqrt{2}}$
    d. DL
B. Distributional methods:
    a. Cohen
    b. Winsorised

The use of simple substitution or a distributional method is mutually exclusive, thus the user can select either a variation of the simple substitution or one of the two available distributional methods.

The software produces a set of four PDF graphs for each analysis and the types depend on the user selected substitution methods. For an assessment using simple substitution the software provides an empirical PDF, an empirical PDF with small intervals, a kernel estimation and a comparison of the previous three PDFs. For an assessment using a distributional method the software provides a normal curve PDF instead of an empirical PDF.

Module 1 interfaces

The following paragraphs present in detail the software implementation and interfaces of module 1. The data samples are simulated by using a random generator, representing the empirical or estimated measurements, and then loaded into the Microsoft Excel environment. A dialog box allows the user to fit parameters for the PDF generation and select the graphs that should be displayed. The scheme of the PEC module application is shown in Figure 17.

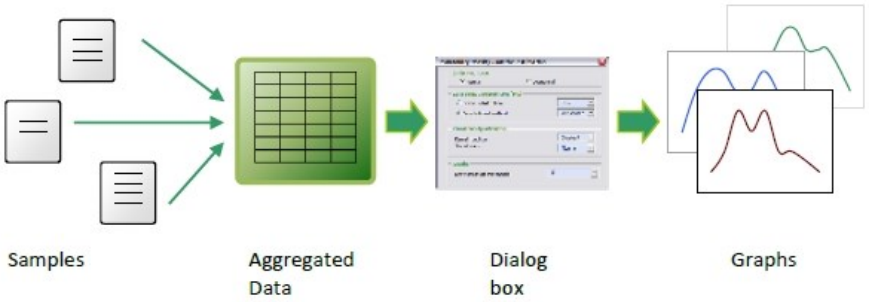


Figure 17: Scheme of PEC module application

The user-interacting dialog box is presented in Figure 18. It allows the user to set up the needed parameters and choices for the production of the PDFs, as presented above.

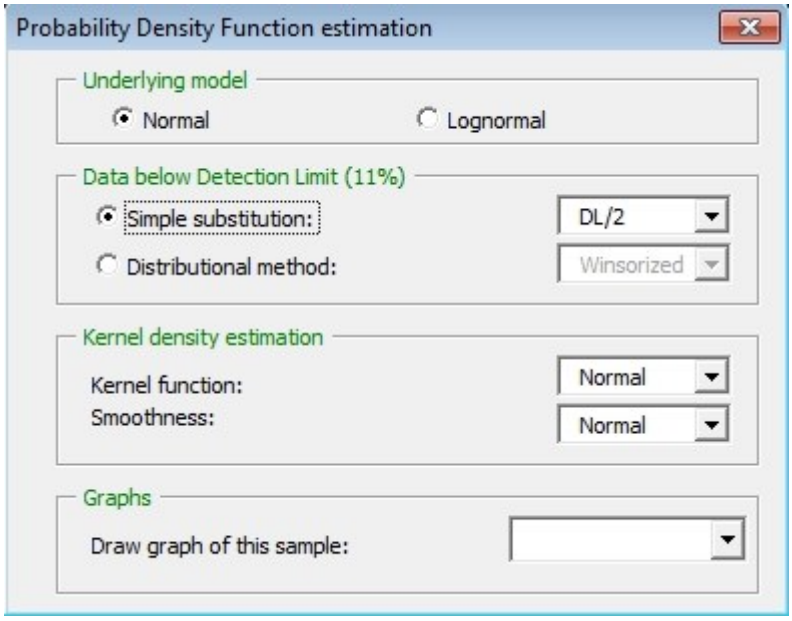


Figure 18: The dialog box of module 1.

Sampling data can be entered manually in a table, where every row represents the series of data measured in the sampling station at each different time. First action the user has to perform is to select the assumed underlying model and whether to use simple substitution or not. If not, the only other choice is a distributional method. For simple substitution, values proposed to user are: $0, \frac{DL}{2}, \frac{DL}{\sqrt{2}}, DL$.

The available distributional methods are *Cohen* and *Winsorized*. In both cases, mean and standard deviation are evaluated from original data without considering missing data. Subsequently, the new data are generated following the behaviour of a Normal variable where mean and standard deviation have been previously computed as parameters.

To support the user in the method decision, a percentage of missing data is calculated first. Then, if the percentage is less than 15%, the option selected by default is *simple substitution*, otherwise a *distributional method* is chosen. In both cases, the user can change this choice. The percentage is also reported in the user interface at the dialog bog, next to the "Data below Detection Limit" box (see Figure 18). Finally the user selects which site graphs he/she wants to be displayed.

Results presentation

Once all parameters have been inserted, four graphs are displayed to user, showing the curve of the probability density function estimation computed in different ways.

1. Empirical PDF or normal curve

If a simple substitution was chosen, the first graph shows the correspondent empirical PDF, made by counting the occurrences of values in large intervals. If distributional methods were selected, the correspondent Normal distribution function is shown.

2. Empirical PDF with small intervals

Simple substitution and the distributional methods produce the same type of empirical PDF when plotted by using smaller value intervals: The graph is created by plotting the data of the empirical PDF but subdividing the range of values into intervals of less width than the original empirical PDF graph of the simple substitution that is described in point 1 of the results presentation paragraph.

3. Kernel estimation

Both simple substitution and distributional method produce the same kind of graph for the kernel density estimation. Kernel density estimation's graph is obtained by placing a kernel function on each data-point. Their sum forms the function estimation, which could be influenced by the kernel shape and the bandwidth. In case of simple substitution the estimation takes into account substituted values (in fact there is a peak in correspondence of the substitution value) while in case of a distributional method it discards missing data. As a consequence, the entire curve is slightly higher if compared with the kernel graph of the simple substitution.

4. Comparison

The last graph contains comparisons of all previous mentioned graphs. Since they have different points on the horizontal axis and different expansions on the vertical axis, data have to be normalised to a common scale. An example of the obtained graph is shown in Figure 19: on the left an example in case of simple substitution, on the right an example in case of a distributional method.

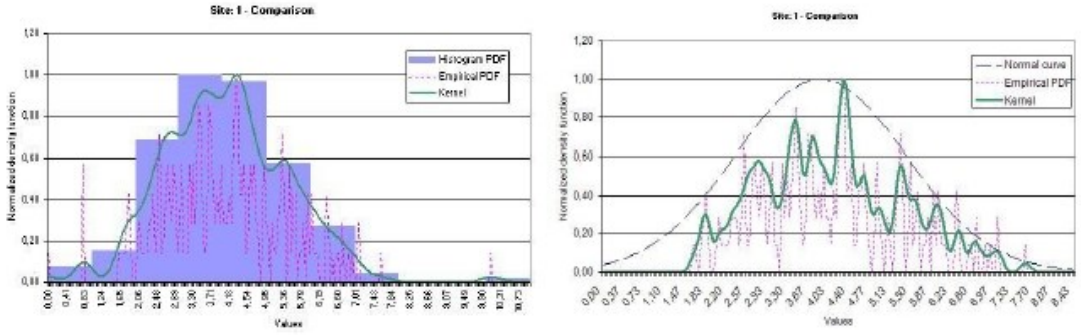


Figure 19: Final comparison graph of the three PDFs produced by module 1.

A general screenshot of the PEC module is presented in Figure 20. The figure represents an example of execution. The charts have been drawn using the values on the top of the sheet as data sample. The user dialog box can be seen on the upper right corner.

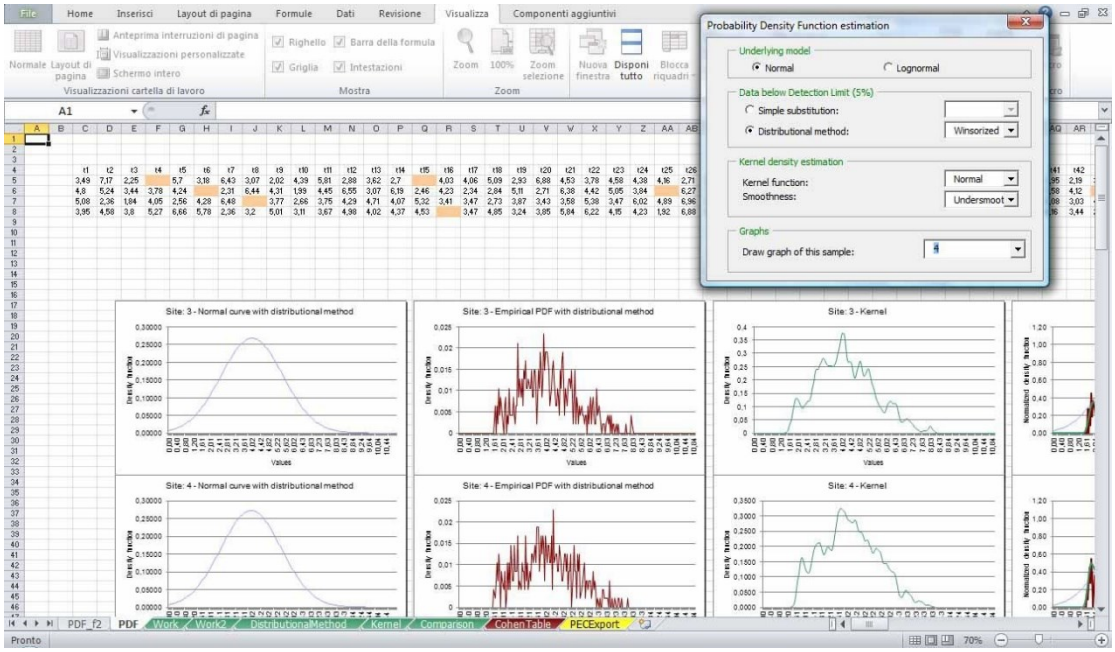


Figure 20: Screenshot of PEC module spreadsheet.

4.2.2. Module 2: Effect Assessment (Species Sensitivity Distribution – SSD)

The design of the 'Effect Assessment' module is based on the Weight of Evidence (WoE) methodology, which incorporates Multicriteria Decision Analysis (MCDA) methods, as described in chapter 3.

The initial step for enabling the reliability evaluation in the DSS is to import the Experts' knowledge base, which is created through the 'Questionnaire for expert consultation' as described in paragraph 3.1.2. The knowledge base is stored in text files in a separate folder. The user must select the folder containing these files and the software is capable of incorporating them into a single external EXCEL spreadsheet. In this way, once created, the expert knowledge file can be reused in multiple different assessments. Once a knowledge base file is available it has to be imported into the current assessment just by selecting its location.

Once the expert knowledge file is linked to the present assessment file, the following step for a user in the implementation of the software is to compile the response sheet, which is shown on Figure 21 and contains the 57 criteria questions of the AMORE WoE framework for given ecotoxicological data. The response sheet is used to extract the reliability score of the ecotoxicological data.

In order for the DSS to generate the SSD, the user has to fill in data in a specific form in the EXCEL spread-sheet, sorted in columns with headers. Columns to be filled in are: (1) the name of the originating test (must be the same name of the correspondent response sheet); (2) the name of the species (or the genus) used in the ecotoxicology test; (3) the trophic level (or the taxonomic group) of the species; (4) the toxicological test results (concentration data). All the data can belong to the same trophic level (or taxonomic group), even to the same species. The number of taxonomic groups is not limited. The 'Worksheet' names should not contain any spaces. Once all other information is supplied the reliability scores are automatically generated by the system based on the response sheets related to the involved tests. The data related to test results and reliabilities are then used to create the SSWD curves. An example can be seen below in Figure 22.

Criterion	Question	YES	NO	NOT APPLICABLE	APPLICABLE BUT NOT REPORTED	I DON'T KNOW
1.1.1.1. Guideline criterion	Did the test strictly follow an international guideline and/or national guideline?					
1.1.2.1. GPL criterion	Did the test strictly follow Good Laboratory Practices?					
1.2.1.1. Substance identity criterion	Was the test substance identified by unambiguous information?					
1.2.1.2. Substance purity criterion	Was the purity of the substance given?					
1.2.2.1. Loss ability criterion	Was the concentration of the test substance potentially able to change during the test period (by e.g. degradation, volatilization, formation of by-products, sorption onto container surface)?					
1.2.2.2. Concentration monitoring criterion	Was the concentration of the tested substance measured during the test period ?					
1.2.2.3. Loss acceptability criterion	Was the concentration loss of the tested substance during the test period acceptable?					
1.2.2.4. Speciation criterion	Does the chemical monitoring allow to discriminate the different chemical species when relevant (dissolved vs particulate forms, ionic vs neutral forms etc)?					
1.2.2.5. By-products criterion	Was the additional ecotoxicity of by-products formed during the test period evaluated?					
1.3.1.1. Species identity criterion	Was the tested species variety and/or strain and/or isolate given?					
1.3.1.2. Organisms physiology criterion	Was the age/length/weight/gender of the tested organisms given and suitable for running the test?					
1.3.2.1. Acclimatation criterion	Is the acclimatation duration (i.e. contact of the organisms with the investigated media during a given time period before introduction of the toxicant) adapted?					
1.3.2.2. Organism feeding criterion	Do you consider that feeding of organisms over the test period is well stated and adapted?					
1.4.1.1. No toxicant control criterion	Were appropriate 'No toxicant controls' (i.e. test under the same conditions without the investigated substance) used?					
1.4.1.2. No toxicant control results criterion	Do the results obtained for no toxicant controls meet the validity criteria of the test?					
1.4.1.3. Reference substance criterion	For the tested organism and effect, does a dose-response relationship exist for a reference substance?					
1.4.1.4. Reference substance results criterion	Do the results obtained for the reference substance meet the expected toxicity range?					
1.4.2.1. Temperature criterion	Was the test temperature suitable for tested organisms?					
1.4.2.2. pH criterion	Was the test pH suitable for tested organisms?					
1.4.2.3. Light criterion	Was the light intensity and photoperiod stated and suitable for tested organisms?					
1.4.2.4. Dissolved oxygen criterion	Was Dissolved oxygen concentration suitable for tested organisms?					
1.4.2.5. Other physic-chemical parameters criterion	Were the other physico-chemical parameters (e.g. organic matter, hardness) known as potentially influent vs toxicity suitable for tested organisms?					
1.4.3.1. Exposure system criterion	Do you consider that conducted exposure system (e.g. static; semistatic; flow through) can generate bias?					
1.4.3.2. Exposure route criterion	Is the route of exposure reliable for the species?					
2.1.1.1. Concentration number criterion	Was the number of tested concentrations acceptable?					
2.1.1.2. Concentration spacing criterion	Was the concentration spacing acceptable?					
2.1.2.1. Replicates criterion	Was the number of replicates per concentration in agreement with the requirement of the statistical method applied?					
2.1.2.2. Individuals number criterion	Was the number of individuals per replicate in agreement with the requirement of the statistical method applied?					
2.2.1.1. Precision criterion	In case of Hypothesis-testing, were Precision (e.g. MSD or PMSD) criteria explicitly given and acceptable?					
2.2.1.2. Power criterion	In case of Hypothesis-testing, were Power (type II error rate) and type I error rate explicitly given and acceptable?					
2.3.2.1. Variance homogeneity criterion	In case of Hypothesis-testing, was the homogeneity of variance between treatments explicitly verified and acceptable?					
2.2.2.1. Regression model selection criterion	Was there some mechanistic support to the choice of regression model?					
2.2.2.2. Regression model comparison criterion	Were several regression models tested and was the selected model the best-fitting?					
2.2.2.3. Prior assumption criterion	In case of Bayesian NEC, was the prior justified and acceptable?					
2.3.1.1. Confidence interval criterion	Were the confidence interval for summary statistics indicated and was the confidence interval acceptable?					
2.3.1.2. Goodness of fit criterion	Was the fit acceptable, with a justification of goodness of fit?					
2.3.1.3. Comparative robustness criterion	Was the robustness of the summary statistics tested by comparison with alternative assumptions or approaches?					
3.1.1.1. Acute duration criterion	In case of acute test, was the test duration acceptable?					
3.1.1.2. Deterministic ACR criterion	In case of acute test, is there a robust deterministic Acute-to-Chronic Ratio able to convert Acute summary statistics to Chronic summary statistics?					
3.1.1.3. Probabilistic ACR criterion	In case of acute test, is there a robust probabilistic Acute-to-Chronic Ratio able to convert Acute summary statistics to Chronic summary statistics?					
3.1.2.1. Life-cycle criterion	In case of chronic test, did the test cover an acceptable range of the species life-cycle?					
3.2.1.1. Temperature relevance criterion	Was temperature used for the lab test suitable for the natural system targeted by the risk assessment?					
3.2.1.2. pH relevance criterion	Was pH used for the lab test suitable for the natural system targeted by the risk assessment?					
3.2.1.3. Salinity relevance criterion	Was salinity used for the lab test suitable for the natural system targeted by the risk assessment?					
3.2.1.4. Dissolved oxygen relevance criterion	Was dissolved oxygen used for the lab test suitable for the natural system targeted by the risk assessment?					
3.2.1.5. Other physico-chemical parameters relevance criterion	Were there some other physico-chemical parameters used for the lab test that are not suitable for the natural system targeted by the risk assessment?					
3.2.2.1. Background criterion	Were the tested concentrations significantly higher than the measured or predicted natural geochemical background concentration?					
3.2.2.1. Ambient concentration criterion	Were the tested concentrations significantly higher than the measured or predicted ambient concentration upstream of the investigated site?					
3.3.1.1. Hypothesis-testing relevance	If the summary statistics is derived from hypothesis-testing, could it be substituted by a more robust summary statistics derived from an alternative approach?					
3.3.1.2. Effect bound relevance	Was the summary statistics a lower bound of effect on population?					
3.3.2.1. Most sensitive endpoint criterion	If different endpoints were measured for the investigated biotest, is the reported endpoint the most sensitive one?					
3.3.2.2. Mode of Action criterion	If information related to Mode of Action is available, is there any justification that the reported endpoint is among the most sensitive ones?					
3.3.3.1. Population dynamics criterion	Can the measured effect be considered to be an adverse effect directly affecting population dynamics?					
3.3.3.2. Behavioral effect criterion	Can the measured effect be considered to be a behavioral effect not directly affecting population dynamics?					
3.3.3.3. Molecular/Cellular effect criterion	Did the biotest measure effect at molecular/cellular scale not directly affecting population dynamics?					
3.4.1.1. Habitat criterion	For the tested lab species, are the ecological traits related to habitat suitable for the natural system targeted by the risk assessment?					
3.4.1.2. Food criterion	For the tested lab species, are the biological traits related to food preferences and feeding habits suitable for the natural system targeted by the risk assessment?					

Figure 21: AMORE user response sheet

chronic / acute	Species	Taxonomic or trophic	Concentration values	Test name	First data quality
C	Asellus communis	INV	27.92	Oseid and Smith 1979	0.4215202
C	Champia parvula	ALG	3.9	Steele_1983_Champia	0.4739313
C	Chlamydomonas	ALG	10	Cairns_1978_chlamy	0.4987536
C	Chlamys asperrimus	INV	5	Pablo_1997_chlamys_	0.5429249
C	Gammarus pseudolimnaeu	INV	3.9	Oseid and Smith 1979	0.4215202
C	Hydra viridissima	INV	200	Rippon et al 1992	0.4071754
C	Moinodaphnia macleayi	INV	5.8	Rippon et al 1992	0.4071754
C	Nitzschia closterium	ALG	10	Pablo chronic_Nitzsch	0.5692871
C	Oncorhynchus mykiss	VE	10	Dixon_1981_salmo	0.3722197
C	Cyprinus carpio	VE	73	Jee and Kang 1999	0.4301761
C	Lemna gibba	PLAN	3.58	Lemna test	0.6569323
C	Chironomus riparius	INV	5	Chironomids Chronic t	0.6665303
A	Chlamys asperrimus	INV	2.24	Pablo_1997_chlamys_	0.6359432
A	Cancer magister	INV	5.1	Brix_2000_Cancer spp	0.563509
A	Nitzschia closterium	ALG	5.487	Pablo acute_Nitzschia	0.568967
A	Cancer productus	INV	10.7	Brix_2000_Cancer spp	0.563509
A	Cancer oregonensis	INV	13.1	Brix_2000_Cancer spp	0.563509
A	Gasterosteus aculeatus	VE	13.1	Broderius 2 Gast	0.5915882
A	Cancer gracilis	INV	14.4	Brix_2000_Cancer spp	0.563509
A	Gammarus pseudolimnaeu	INV	16.269	Oseid and Smith 1979	0.4215202
A	Asellus communis	INV	221.1	Oseid and Smith 1979	0.4215202
A	Gammarus fasciatus	INV	8.39	Smith 1979 2 gammar	0.5598878
A	Spirostomum ambiguum	INV	118	Nalecz-Jawecki_1998_	0.3088132
A	Moina micrura	INV	1546	Bhunia_2000_moina	0.446109
A	Branchiura sowerbyi	INV	16688	Bhunia_2000_branchiu	0.5518175

Figure 22: An example of ecotoxicological data, sorted in columns for use in module 2.

Module 2 interfaces

Similarly to module 1, a dialog box (see Figure 23) provides the user the possibility to select the information to be analysed and the parameters to be used by the software for producing the SSWD graph. An explanation of the three parts (i.e. data, weighting procedures and statistical options) composing the dialog box is reported in the following paragraphs.

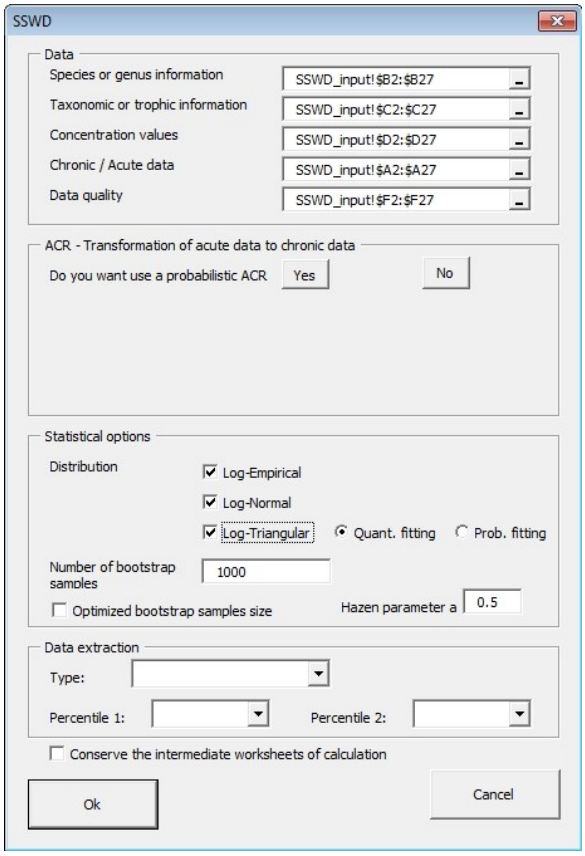


Figure 23: The dialog box of module 2.

Data box

On the 'Species or genus information' field the user should select the range of data or the column containing the name of the species of each ecotoxicological test or the genus of the species of each test, with the heading in the first line of the selection. If a column is selected, the first line of the column must contain the heading. The program stops reading the data at the first empty cell.

On the 'Taxonomic or trophic information' field the user should select the range of data or the column (heading included) containing the information about the taxonomic group or the trophic level.

On the 'Concentration values' field the user should select the range of data or the column (heading included) containing the ecotoxicological test results.

On the 'Chronic / Acute data' field the user should select the range of data or the column (heading included) containing the information about the chronic or acute data.

Similarly, on the 'Data Quality' field the user should select the range of data or column containing the reliability scores of ecotoxicological data.

Acute to Chronic data transformation

The user has the possibility to select the use of an Assessment Factor (AF) for the transformation of acute data to chronic.

If the user selects YES, the software provides the possibility to use either a Normal or a Uniform distribution for the transformation and the user has to input the numerical details for each option (Min-Max, Mean).

Statistical options

The user selects the type of 'Distribution' to be applied to the weighted data points and three available options are included in the module: log-empirical, log-normal and log-triangular, which can all be used at the same time. The parameters of the log-triangular distribution are obtained by fitting the theoretical distribution to the empirical weighted data points. Two possibilities are offered: fitting the quantiles (Quant. fitting); fitting the cumulative empirical probabilities of the data (Prob. fitting).

The confidence limits of the hazardous concentration are estimated by bootstrap. The user should enter the 'Number of bootstrap samples' generated to calculate the confidence limits. The number of data points drawn in each sample of the bootstrap approach (the samples size) is by default equal to the number of used data. If option 'Optimized bootstrap samples size' is selected, the number of data points drawn in each sample of the bootstrap is optimized regarding the proportions of data and the weight of each taxonomic group or trophic level. The user can modify the 'Hazen method parameter α', from the default value 0.5.

Results presentation

The numerical results are presented in a separate worksheet and the SSWD graphics are displayed directly in the same worksheet. An example is shown on Figure 24.

1	Options															
2	Species=	Scoring process														
3	Nb bootstrap samples		1000													
4	Nb data		45													
5	Hazen parameter a		0.5													
6	ACR Distribution : yes	Uniform Distribution	min : 10	max : 10	ACR moyen : 10											
7																
8																
9	Distribution: Weighted Empirical + Confidence limits by weighted bootstrap															
10	HC	2.5%	5%	10%	15%	20%	30%	50%								
11	Best-Estimate	0.0191	0.0994	0.2813	0.4896	0.7170	1.1461	3.2347								
12	Geo. Stand. Deviation	9.9318	3.8937	1.8407	1.6815	1.4731	1.3545	1.5821								
13	Centile 2.5%	0.0003	0.0003	0.1160	0.1438	0.2504	0.7248	1.3100								
14	Centile 5%	0.0003	0.0262	0.1160	0.2240	0.3348	0.7300	1.4400								
15	Centile 95%	0.2012	0.4386	0.7300	1.0100	1.2600	1.7000	5.0000								
16	Centile 97.5%	0.2575	0.5100	0.7860	1.0674	1.2848	2.6000	5.0000								
17																
18																
19	Distribution: Weighted Normal + Confidence limits by weighted bootstrap and Bias Correction															
20	HC	2.5%	5%	10%	15%	20%	30%	50%	70%	80%	85%	90%	95%	97.5%		GWM
21	Best-Estimate	0.0293	0.0640	0.1574	0.2889	0.4682	1.0276	3.7690	13.8242	30.3431	49.1739	90.2721	222.1123	484.9792	Best-Estimate	3.7690
22	Geo. Stand. Deviation	2.4448	2.1964	1.9524	1.8128	1.7165	1.5896	1.4772	1.5109	1.5992	1.6747	1.7898	1.9986	2.2155	Geo. Stand. Deviation	1.4772
23	Centile 2.5%	0.0042	0.0113	0.0364	0.0810	0.1473	0.3908	1.7017	6.1068	11.9574	18.4314	29.6679	58.4772	104.2218	Centile 2.5%	1.7017
24	Centile 5%	0.0065	0.0165	0.0483	0.0982	0.1758	0.4449	1.9362	6.9230	14.0684	20.8808	34.8584	71.5826	128.9113	Centile 5%	1.9362
25	Centile 95%	0.1081	0.1997	0.4013	0.6676	1.0216	2.0544	7.0584	26.6890	63.3036	112.4742	232.4532	708.9746	1857.6759	Centile 95%	7.0584
26	Centile 97.5%	0.1275	0.2303	0.4657	0.7616	1.1400	2.2768	7.9628	29.5778	75.3980	135.8486	279.8506	843.2485	2368.2902	Centile 97.5%	7.9628

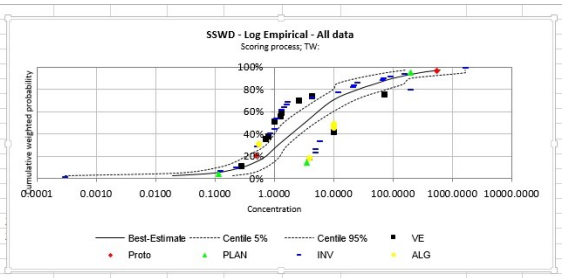


Figure 24: Screenshot of SSD module spreadsheet

The software delivers in the spreadsheet all the details regarding the input information that the user provided, as well as the statistical and numerical results (such as weights, the weighted cumulative probabilities, distribution parameters etc.) that are used for the production of the SSWD.

On the SSWD graphics the weighted data points, the empirical or the theoretical distribution and the 90% confidence limits of this distribution are displayed. The colour of the data points depends on the taxonomic group or the trophic level of the considered species. For the theoretical distribution, the multiple R-square coefficient (R_2) between theoretical and empirical distribution and the Kolmogorov-Smirnov goodness of fit test are displayed. In addition, the user finds the parameters of the distribution calculated on the log of the data. For the log-normal distribution, it is the mean (*wm.lg*) and the standard deviation. For the log-triangular distribution, it is the min, the max and the mode (*wmin.lg*, *wmax.lg* and *wmode.lg*). An example is shown in Figure 25.

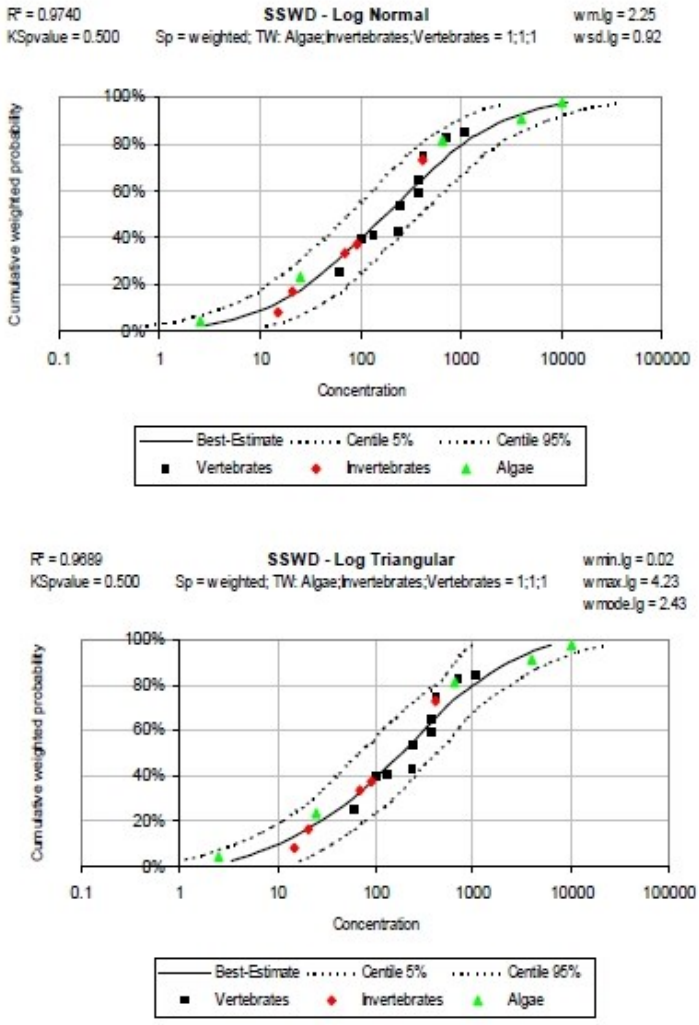


Figure 25: Produced SSWD for Log normal and Log Triangular distributions.

In these graphics, the concentration values are directly put in the unity of the data supplied by the user, but with a log-scale. These graphics are constructed with the worksheet tables and can be modified by the user. In particular, the 90% confidence limits can be replaced by the 95% confidence limits.

4.2.3. Module 3: Risk Assessment (Potentially Affected Fraction – PAF)

The design of module 3 of the AMORE DSS is based on the theoretical developments regarding the estimation of risk based on the notions of the Joint Probability Curve and the Potentially Affected Fraction graphs. The 'Potentially Affected Fraction' graph, representing the percentage of species affected by specific concentration levels, is derived from the combination of the PEC and the SSD cumulative distributions that are produced in modules 1 and 2. PAF is considered an index for ecological risk and allows comparisons between substances, species groups, sites, and regions.

The Joint Probability Curve (JPC) is a method for presenting the joint probability of PEC and SSD in the form of an exceedance curve, which describes the probability of exceeding the concentration associated with a certain degree of effect. Each point on the curve represents both the probability that the chosen proportion of species will be affected and the frequency with which that level of effect would be exceeded. At each point on the curve holds that, under the given conditions, x % of species will be affected and that such proportion of species would be affected by y % of the current observations (Solomon et al., 2000). Depending on the type of exposure data collected such observations could refer to time or locations (Posthuma et al., 2002). Since the exposure profile given by PEC (indicating a probability of occurrence) and the toxicity profile given by SSD (indicating a magnitude of effect) have the same horizontal axis (the concentration), the two curves can be combined into one, which is the JPC (Hendley and Giddings, 1999). The y-axis varies from 0 to 1 (or to 0 to 100 if it's expressed by percentage) because it represents a probability, both for PEC and SSD. The JPC is a decreasing curve, which starts from the coordinates (1,1) - or, alternatively, at (100%;100%) - and ends at (0,0) - or (0%;0%).

How to obtain the JPC

For each point representing a cumulative frequency on SSD, the corresponding concentration on the x-axis is calculated. Then, in the correspondence of such point, the cumulative frequency of PEC is achieved. Such two cumulative frequency values represent, respectively, the coordinates of a point in the JPC graph. In Figure 26, the process of JPC derivation is reported in graphical way. Note that, in the first graph, the vertical axis on the left has the reversed scale respect to the vertical axis on the right: this is because the value, which has to be plotted to form the JPC, corresponds to 'one minus the predicted environmental concentration'.

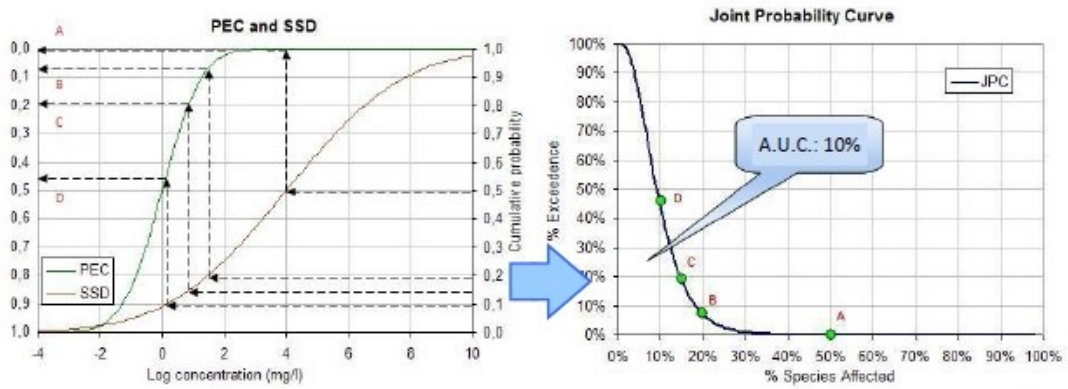


Figure 26: Derivation of a Joint Probability Curve from PEC and SSD distributions.

A different shape for the JPC can be obtained by plotting the cumulative SSD on the vertical axis and the cumulative PEC (not the exceedance profile as before) on the horizontal axis. In this case the curve runs from the origin to the point (1, 1).

Module 3 interfaces

The interface of the module that has been developed in order to obtain the PAF is described in this paragraph. The basic scheme of the module is reported in Figure 27.

Figure 27: General scheme of module 3.

The user has the possibility to select between (i) running the PAF module with previously calculated data or (ii) running the PAF module with completely new data. A dialog box is used to set up the needed parameters and user choices. In the first case, the user can use values calculated in the previous phases of the risk assessment process (with the use of the PEC and SSD modules, as seen in Figure 28) , whereas in the second case the user can insert the mean and standard deviation of the two distributions representing PEC and SSD (as seen in Figure 29). In the first case, the user can choose which PEC sample should be used for the calculations (either a specific station or all the available stations). The SSWD module always provides one parameter's evaluation, so the same action is not necessary. Mean and standard deviation can be related to a Normal or a Log-normal distribution: for both PEC and SSD the user can select which of the two functions to use.

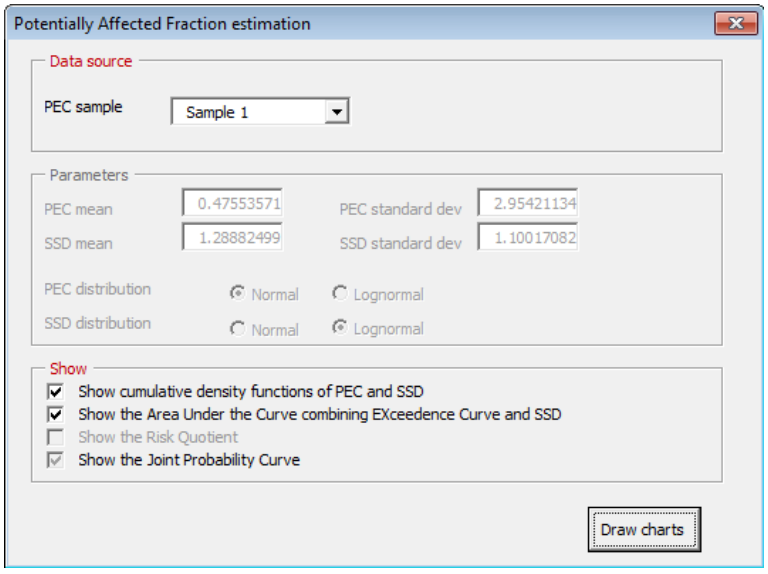


Figure 28: User dialog box of module 3, data previously calculated in modules 1 and 2.

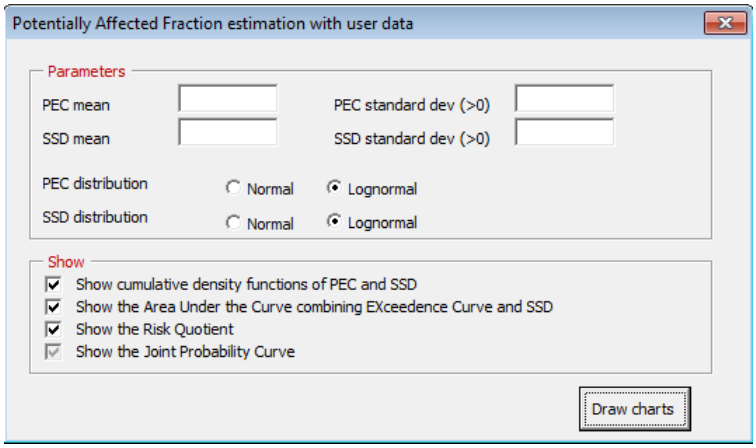


Figure 29: User dialog box of module 3, new data

Results presentation

The first graph the user can select presents the PEC and SSD cumulative density functions to have a rapid overview of the submitted distributions. If the two curves are not overlapping, the graph gives a first idea of the risk: if the SSD is on the right of the PEC, the risk is expected to be low otherwise it is expected to be high. If the two curves overlap other graphs are necessary to better understand the characteristics of risk.

The second chart shows the 'Area Under the Curve', which is obtained as the integral of the exceedance profile (one minus the PEC) and the probability density function of species sensitivity.

The third one is the Risk Quotient, which gives a different visual representation of the generalised risk and is available only in case of two Normal or two Log-normal distributions for PEC and SSD options.

The last graph displays the Joint Probability Curve, which is the most expressive of the four representations. The user can re-iterate the entire procedure. The numerical risk is calculated with the trapezoidal rule and is reported to the user on the title of the JPC chart.

A general screenshot of the module is presented in Figure 30 and represents an example of execution.

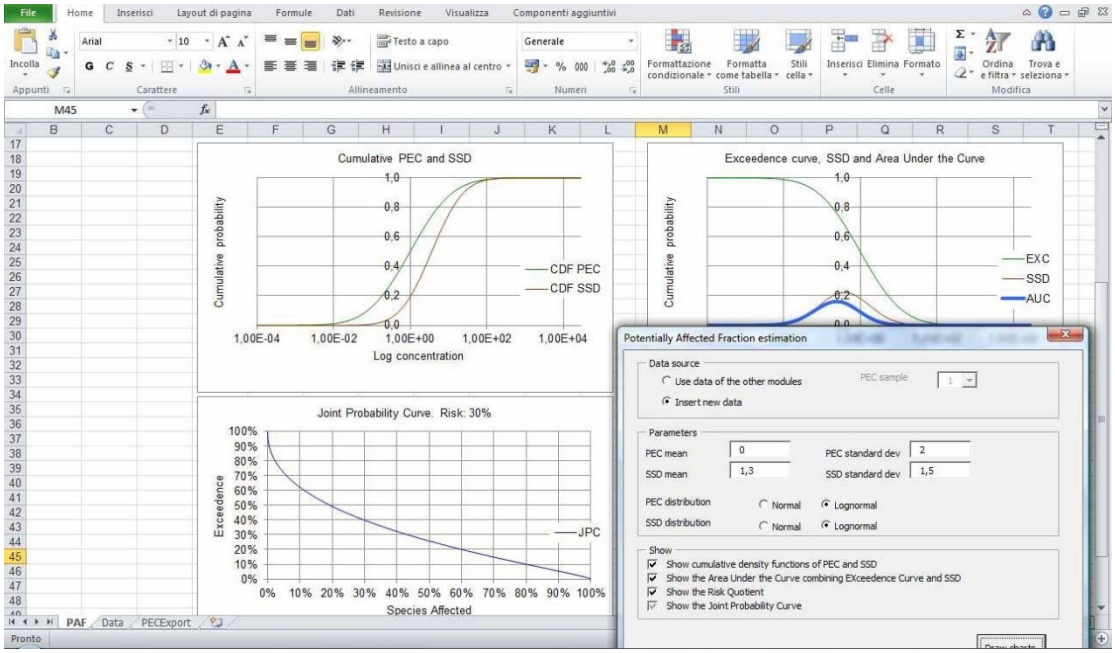


Figure 30: screenshot of PAF module spreadsheet.

Section C: Application to case study

5. Application to cyanide case study

The following chapters contain the full description of the application of the AMORE Decision Support System (DSS) to the case study on cyanide, which has been performed for the validation both of the MCDA based, Weight of Evidence methodology (chapter 3) and of the complete DSS (chapter 4).

Paragraph 5.1 contains the presentation of the case study and includes the information on the assessment region, the theoretical information on cyanide and the basic information of the performed risk assessment process.

The results of the case study are described in detail and discussed in paragraphs 5.2-5.3. Paragraph 5.2 includes the results of the assessment of ecotoxicological data, the application of the MCDA based methodology for the analysis of the reliability of the ecotoxicological data used in the case study, the production of the Species Sensitivity Distributions (SSDs) and the related graphs. Paragraph 5.3 includes the application of the PEC and PAF modules of the DSS for the risk assessment of cyanide in the proposed assessment area.

5.1. Presentation of the case study

A case study application has been performed for the analysis of the ecological risk from the presence of cyanides in the Sélune watershed (Figure 32), at the Manche department of the Lower Normandy (Figure 31) region in the north-west part of France. Environmental exposure data of cyanide (CN) have been collected from the web portal of the Water Agency of 'Seine-Normandie' (http://www.eau-seine-normandie.fr/index.php?id=1628) and used in the Exposure Assessment module, while ecotoxicological data for cyanide toxicity gathered from peer-reviewed publications have been analysed with the use of the MCDA based methodology, in the Effect Assessment module. The ecological risk assessment process was concluded with the calculation of the risk indices in the last module (PAF) of the DSS.

Twelve water quality measurement stations have been identified in the region and included in the query. Cyanide concentrations have been observed at four of those stations and environmental exposure data have been collected for the period 01/01/2005 – 20/08/2014. The four stations are located along the Selune river, in spots close to the cities Isigny-le-buat, Saint-aubin-de-terregatte, Les loges-marchis and Romagny, and can be seen in Figure 32.

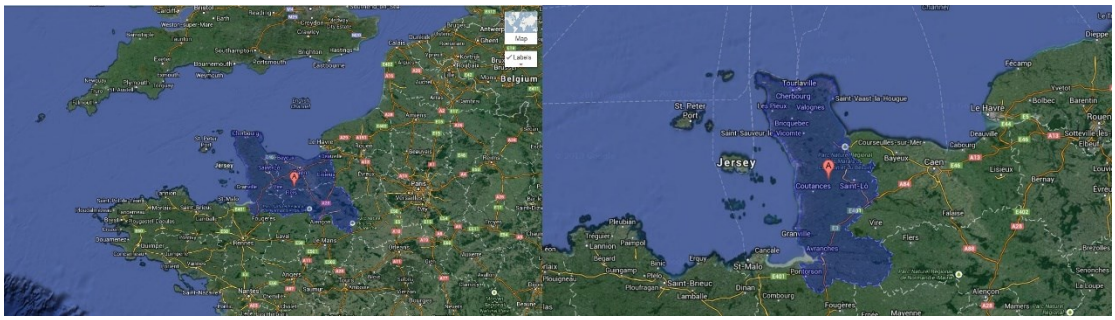


Figure 31: The Lower Normandy region of north-west France (left) and the Manche department (right)

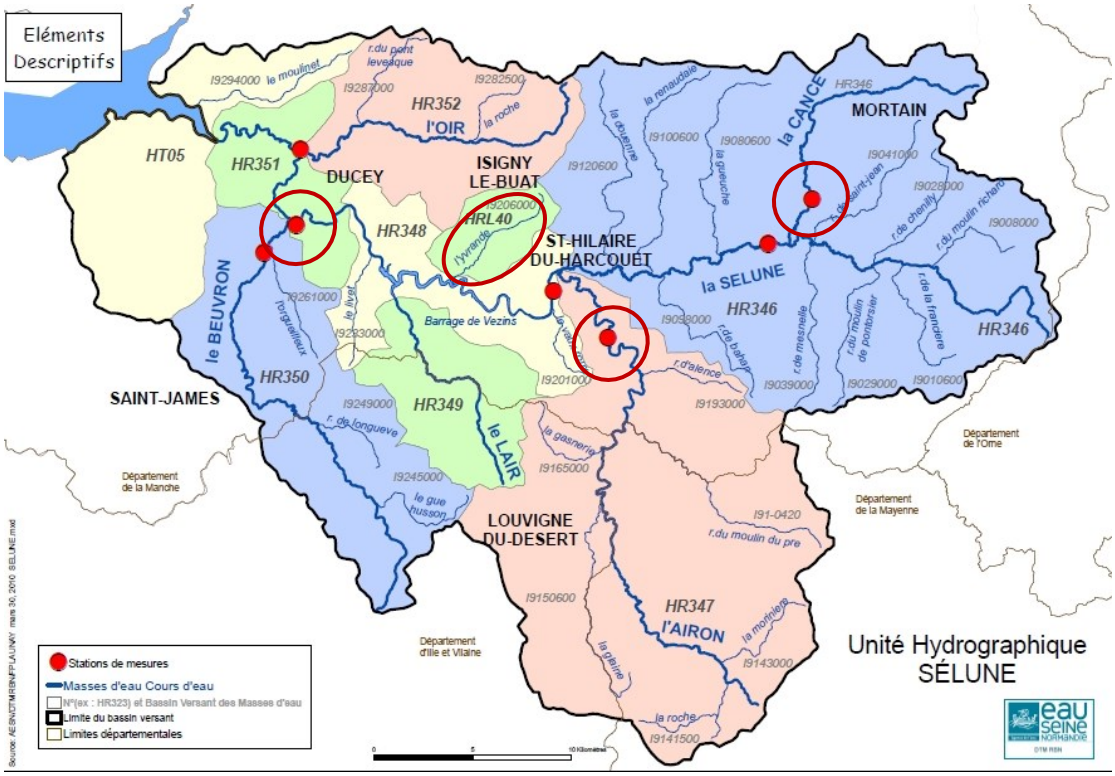


Figure 32: The Selune watershed in France (source: http://eau-seine-normandie.fr/). In circles the locations of the four stations where cyanide concentrations have been measured.

5.1.1. Cyanide[1]

Cyanides are a group of chemical compounds characterised by occurrence of the association of a carbon and a nitrogen atom (CN). These compounds occur in the environment as free cyanides (HCN and CN-), simple associations to cyanides as NaCN and KCN, metal-cyanides complexes (e.g. iron and cyanides) and organic complexes (nitriles and glucosides). Cyanides mainly occur in waters as free cyanides, mostly hydrogen cyanide (HCN) which represents the main toxic form with cyanide ion CN- (US-EPA, 1984).

Hydrogen cyanide (HCN) is a colorless or pale blue liquid or gas with a faint bitter almond-like odor, while sodium cyanide (NaCN) and potassium cyanide (KCN) are white crystalline powders. HCN is a weak acid with a pKa of 9.2; therefore, HCN and CN- can interconvert based on pH and temperature. In solution under physiological conditions, the majority of HCN is present in the undissociated form. The simple cyanide salts, KCN and NaCN, are very soluble in water and mildly soluble in ethanol. These compounds readily dissociate in water, and therefore, exposure to any of these compounds in aqueous media results in exposure to CN-.

[1] Based on information from US EPA (Toxicological review of hydrogen cyanide and cyanide salts, 2010) and WFD-UK Technical Advisory Group (Proposed EQS for Water Framework Directive Annex VIII substances: cyanide (free), 2008)

Cyanides are extensively used in industry and are also emitted from car exhaust fumes. They also occur naturally in the environment and are found in a range of aquatic organisms such as arthropods, macrophytes, fungi and bacteria.

Cyanide originates primarily from anthropogenic sources in the environment, but cyanide is also released from biomass burning, volcanoes, and natural biogenic processes from higher plants, bacteria, and fungi (Agency for Toxic Substances and Disease Registry [ATSDR], 2006). Cyanide is also a component of tobacco smoke and can be present at high concentrations in structural fires (Steinmaus et al., 2007; Brauer et al., 2006; Tsuge et al., 2000; Brandt-Rauf et al.,1988). Cyanide compounds are used in a number of industrial processes, including mining, metallurgy, manufacturing, and photography, due to their ability to form stable complexes with a range of metals. Cyanide has been employed extensively in electroplating, in which a solid metal object is immersed in a plating bath containing a solution of another metal with which it is to be coated, in order to improve the durability, electrical resistance, and/or conductivity of the solid. HCN has also been used in gas chamber executions and in chemical warfare. NaCN and KCN are also used as rodenticides.

Cyanide or cyanogenic compounds are found in many foods. Cyanide compounds occur naturally as part of sugars or other naturally occurring compounds in certain plant-derived foods, including almonds, millet sprouts, lima beans, soy, spinach, bamboo shoots, sorghum, and cassava roots.

Volatilisation and biodegradation are important transformation processes for cyanide in ambient waters. Hydrogen cyanide can be biodegraded by acclimated microbial cultures, but is usually toxic to unacclimated microbial systems at high concentrations.

Cyanides are readily soluble in water, where they exist in the free state (CN- and HCN), as simple cyanides (e.g. NaCN), complex cyanides (organic or metal complexes) or total cyanide (all available species). Hydrogen cyanide (HCN) dissociates in water to give the free ion (CN-) under alkaline conditions (50 per cent of both forms at pH 9.36). The CN- ion has a half-life of 15 days in water; HCN has a tendency to volatilise from water, with a half-life measured from hours to a few days. Simple cyanides readily dissociate, as do some metal complexes (e.g. zinc and cadmium) releasing free CN-. Other metal complexes containing cyanide are very stable with limited dissociation.

Cyanide acts as a respiratory depressant and can inhibit aerobic metabolism. Free cyanide ions can also pass though the gill membranes. Undissociated HCN is primarily used to determine toxicity, with HCN being more toxic than CN-. However, CN- contributes to toxicity due to formation of HCN at pH values up to around 8. Simple cyanides readily dissociate and hydrolyse to form HCN and CN- and, therefore, have the same toxicity as free cyanide.

5.1.2. Exposure data

As mentioned earlier, the Environmental exposure data of cyanide (CN) have been collected from the Water Agency of ‘Seine-Normandie’. The 91km Selune River and its rivershed have been selected as the area of interest and interesting spots along the river have been identified. A total of 12 stations (Table 1) have been examined for the period 01/01/1990-20/08/2014, though contamination from cyanide has been observed and measured for only 4 of them (Table 2) for the period 01/01/2005 – 20/08/2014.

The selected parameter was free cyanide (HCN and CN-) in natural water and the acqueous phase of water (filtered, centrifugal) and was measured in µg(CN)/L.

The actual exposure values and their statistical analysis for the estimation of the Predicted Environmental Concentration (PEC), are presented in paragraph 5.3.

Station Code	Station name	Municipality	County-region	Spot name
03271415	LA CANCE A ROMAGNY 1	ROMAGNY	MANCHE	La Cance
03271515	LA GUEUCHE A MILLY 1	MILLY	MANCHE	La Gueuche
03271437	LA SÉLUNE A NOTRE-DAME-DU-TOUCHET 1	NOTRE-DAME-DU-TOUCHET	MANCHE	La Selune
03272685	LA SÉLUNE A SAINT-AUBIN-DE-TERREGATTE 1	SAINT-AUBIN-DE-TERREGATTE	MANCHE	La Selune
03272040	LA SÉLUNE A SAINT-HILAIRE-DU-HARCOUET 2	SAINT-HILAIRE-DU-HARCOUET	MANCHE	La Selune
03271965	L'AIRON A LES LOGES-MARCHIS 2	LES LOGES-MARCHIS	MANCHE	L'Airon
03274000	LE BEUVRON A MONTJOIE-SAINT-MARTIN 1	MONTJOIE-SAINT-MARTIN	MANCHE	Le Beuvron
03274420	LE BEUVRON A SAINT-AUBIN-DE-TERREGATTE 1	SAINT-AUBIN-DE-TERREGATTE	MANCHE	Le Beuvron
03272400	LE LAIR A HAMELIN 1	HAMELIN	MANCHE	Le Lair
03271840	LE RUISSEAU DU MOULIN DU PRÉ A BUAIS 1	BUAIS	MANCHE	Ruisseau du Moulin du Pre
03273345	L'OIR A DUCEY 2	DUCEY	MANCHE	L'Oir
03272235	L'YVRANDE A ISIGNY-LE-BUAT 3	ISIGNY-LE-BUAT	MANCHE	L'Yvrande

Table 1: The 12 stations of the Selune watershed that have been included in the case study

Station Code	Station name	Municipality	County-region	Spot name
03271415	LA CANCE A ROMAGNY 1	ROMAGNY	MANCHE	La Cance
03271965	L'AIRON A LES LOGES-MARCHIS 2	LES LOGES-MARCHIS	MANCHE	L'Airon
03272235	L'YVRANDE A ISIGNY-LE-BUAT 3	ISIGNY-LE-BUAT	MANCHE	L'Yvrande
03272685	LA SÉLUNE A SAINT-AUBIN-DE-TERREGATTE 1	SAINT-AUBIN-DE-TERREGATTE	MANCHE	La Selune

Table 2: The 4 station of the Selune watershed, where cyanide contamination has been observed and measured

5.1.3. Ecotoxicological data

The ecotoxicological data have been collected from various scientific articles. In detail, 26 articles regarding cyanide toxicity, published in the period 1965-2011 have been analysed. The analysis resulted in the extraction of 46 toxicological endpoints for the aquatic environment, related to the 5 different taxonomic groups: Protozoa, Plants, Algae, Invertebrates and Vertebrates.

The collected endpoints included: no observed effect concentration (NOEC), lowest observed effect concentration (LOEC), effect concentration of x% of species (ECx), median lethal time of x% of species (LTx) and lethal concentration of x% of species (LCx).

The acute ecotoxicological data have been processed (i.e. converted from acute to chronic and from effect to no effect values) in order to obtain the input data (called "calculated NOEC") required to build the SSWD curves. Two assessment factors have been applied, according to the REACH guidance (ECHA, 2008a,b): The first one accounts for the difference between acute and chronic toxicity (factor 10 to extrapolate short-to-long term effect) and the second one to extrapolate from various cyanide effect's endpoint (e.g. EC_x) the no observed effect concentration (NOEC) required to build the SSWD curves. For the second assessment factor the value 10 has been applied to convert the median lethal and effect concentration (LC_{50}, EC_{50}) and median lethal time (LT_{50}) into no observed effect concentration, while for the effect concentration (EC_{25}) a factor 2 has been applied.

The NOEC for the chronic data are reported in Table 3 and the calculated NOECs obtained by the application of the assessment factors for the acute data, as described above, are reported in Table 4.

Chronic

Species	Taxonomic group	NOEC (µg/L)	Reference	Score (Disp 0.3)	Score (Disp 0.1)
Asellus communis	INV	27.92	Oseid & Smith, 1979	0.42	0.50
Champia parvula	ALG	3.9	Steele and Thursby, 1983	0.47	0.56
Chlamydomonas	ALG	10	Cairns et al, 1978	0.50	0.60
Chlamys asperrimus	INV	5	Pablo et al, 1997b	0.54	0.66
Gammarus pseudolimnaeus	INV	3.9	Oseid & Smith, 1979	0.42	0.50
Hydra viridissima	INV	200	Rippon et al, 1992	0.41	0.47
Moinodaphnia macleayi	INV	5.8	Rippon et al, 1992	0.41	0.47
Nitzschia closterium	ALG	10	Pablo et al, 1997a	0.57	0.69
Oncorhynchus mykiss	VE	10	Dixon & Leduc 1981	0.37	0.40
Cyprinus carpio	VE	73	Jee & Kang 1999	0.43	0.47
Lemna gibba	PLAN	3.58	Bertow, 2011b	0.66	0.79
Chironomus riparius	INV	5	Bertow, 2011c	0.67	0.80

Table 3: Chronic ecotoxicological data

Acute

Species	Taxonomic group	EC50/LC50 (µg/L)	Calculated NOEC (µg/L)	Reference	Score (Disp 0.3)	Score (Disp 0.1)
Chlamys asperrimus	INV	22.4	0.224	Pablo et al, 1997b	0.64	0.79
Cancer magister	INV	51	0.51	Brix et al, 2000	0.56	0.71
Nitzschia closterium	ALG	54.87	0.5487	Pablo et al., 1997a	0.57	0.68
Cancer productus	INV	107	1.07	Brix et al, 2000	0.56	0.71
Cancer oregonensis	INV	131	1.31	Brix et al, 2000	0.56	0.71
Gasterosteus aculeatus	VE	131	1.31	Broderius, 1973	0.59	0.73
Cancer gracilis	INV	144	1.44	Brix et al, 2000	0.56	0.71
Gammarus pseudolimnaeus	INV	162.69	1.6269	Oseid ans Smith, 1979	0.42	0.50
Asellus communis	INV	2211	22.11	Oseid and Smith, 1979	0.42	0.50
Gammarus fasciatus	INV	83.9	0.839	Smith et al., 1979	0.56	0.69
Spirostomum ambiguum	INV	1180	11.8	Nalecz-Jawecki and Sawicki, 1998	0.31	0.34
Moina micrura	INV	15460	154.6	Bhunia et al., 2000	0.45	0.55
Branchiura sowerbyi	INV	166880	1668.8	Bhunia et al., 2000	0.55	0.67
Daphnia pulex	INV	100	1	Cairns et al., 1978	0.40	0.47
Rana temporaria	VE	260	2.6	Costa et al, 1965a	0.36	0.40
Gammarus pseudolimnaeus	INV	170	1.7	Smith et al, 1979	0.56	0.69
Artemia salina	INV	6720	67.2	Calleja et al., 1994	0.53	0.61
Streptocephalus proboscideus	INV	2140	21.4	Calleja et al., 1994	0.54	0.62
Patiriella calcar	INV	0.03	0.0003	Mahadevan, 1986	0.38	0.43
Pomoxis nigromaculatus	VE	101	1.01	Smith et al, 1979	0.56	0.68
Pimephales promelas	VE	78.6	0.786	Smith et al., 1978	0.56	0.69
Perca flavescens	VE	73	0.73	Smith et al., 1978	0.56	0.69
Oncorhynchus mykiss	VE	28	0.28	Carballo et al., 1995	0.45	0.55
Tanytarsus dissimilis	INV	2490	24.9	Call et al., 1983	0.51	0.62
Myriophyllum spicatum	PLAN	20000	200	Stanley 1974	0.28	0.28
Physella heterostropha	INV	432	4.32	Cairns and Scheier, 1958	0.48	0.56
Anculosa sp	INV	7000	70	Cairns et al, 1978	0.50	0.56
Aeolosoma headleyi	INV	9000	90	Cairns et al, 1978	0.39	0.49
Tetrahymena pyriformis	Proto	50	0.5	Slabbert and Maree, 1986	0.50	0.64
Philodina acuticornis	Proto	54000	540	Cairns et al, 1978	0.41	0.51
Cichlasoma bimaculatum	VE	87	4.35	Leduc, 1966	0.42	0.50
Lemna gibba	PLAN	11.6	0.116	Bertow, 2011b	0.66	0.79
Chironomus riparius	INV	12.4	0.124	Bertow, 2011a	0.65	0.78
Lepomis macrochirus	VE	126	1.26	Broderius, 1973	0.59	0.73

Table 4: Acute ecotoxicological data with assessment factors applied

5.2. Analysis of ecotoxicological data and weighted Species Sensitivity Distributions (SSWD)

Ecotoxicological data have been analysed for the production of the weighted Species Sensitivity Distribution (SSWD) graphs. As described in paragraph 3.2, the proposed MCDA based aggregation methodology has been used for the analysis of the reliability and relevance of the ecotoxicological data.

A total of 26 peer-reviewed articles has been analysed by members of the expert panel and the various included endpoints have been assessed, based on the multiple criteria of the WoE framework. The 2nd module of the Decision Support System and the incorporated MCDA-based aggregation methodology have been used for the analysis of the 46 available toxicological endpoints (mortality, growth, reproduction and more), of the various assessed species. The assessed tests included both chronic and acute toxicity data. The species belonged to five taxonomic groups (Algae, Plants, Protozoa, Invertebrates and Vertebrates).

Specifically the following number of endpoints has been analysed for each taxonomic group:

- Algae – 4 endpoints
- Plants – 3 endpoints
- Protozoa – 2 endpoints
- Invertebrates – 27 endpoints
- Vertebrates – 10 endpoints

In the analysis, the following aspects of the MCDA methodology have been taken into consideration:

- The complete set of available experts' answers has been used (14 experts).
- All experts have been considered of equal importance for the evaluation and thus equal weights have been assigned to them.
- Two disputability scores have been used. The default value (0.3) of the DSS has been used in the first analysis, while in the second analysis the disputability score has been set as 0.1.

As explained in section (B), the MCDA methodology is taking into consideration the possible existing uncertainty in the evaluation of ecotoxicological data and specifically, the presence of disputable conditions, as described in paragraph 3.2. This element is evaluated in the methodology based on the feedback of the expert panel. The disputability scores are user defined and therefore can be adjusted according to the individual preferences of the decision maker. For the case study, two analysis have been performed, though the second analysis has been included only for giving the reader a demonstration on how different uncertainty factors influence the calculated scores and is not discussed further. The assessment of the ecotoxicological data has been performed under conditions with inherent uncertainty regarding the way the criteria-questions have been interpreted by the assessors and the default disputability score has been deemed as representative for the given conditions. Therefore, only the test scores for the default disputability value have been taken into consideration for the rest of the case study (modules 2 and 3 of the DSS).

The results are presented in Table 5, sorted in descending order, where the first column presents the scores for the disputability scores $\mathrm{disp_e} = 0.3$ and the second column the scores for disputability scores $\mathrm{disp_e} = 0.1$.

Test	Scores	
	D0.3	D0.1
Chironomids Chronic toxicity	0.67	0.80
Lemna test	0.66	0.79
Chironomids Acute toxicity	0.65	0.79
Pablo_1997_chlamys_EC50	0.64	0.78
Broderius 1 Lep	0.59	0.73
Broderius 2 Gast	0.59	0.73
Pablo chronic_Nitzschia	0.57	0.71
Pablo acute_Nitzschia	0.57	0.69
Brix_2000_Cancer spp	0.56	0.69
Smith 1978 2 Perca	0.56	0.69
Smith 1978 1 Pim	0.56	0.69
Smith 1979 2 gammarus	0.56	0.68
Smith 1979 1 pomoxis	0.56	0.68
Bhunia_2000_branchiura	0.55	0.67
Pablo_1997_chlamys_NOEC	0.54	0.66
Calleja 2 Streptoxkit	0.54	0.64
Calleja 1 Artoxkit	0.53	0.62
Call_1983_Tanytarus	0.51	0.62
Cairns_1978_chlamy	0.50	0.61
Slabbert_1986_tetrahymena	0.50	0.60
Cairns_1978_anculosa	0.50	0.56
Cairns_Physa	0.48	0.56
Steele_1983_Champia	0.47	0.56
Carballo_1995_O.mykiss	0.45	0.55
Bhunia_2000_moina	0.45	0.55
Jee and Kang 1999	0.43	0.51
Leduc 1966	0.42	0.50
Oseid and Smith 1979	0.42	0.50
Cairns_1978_philodina	0.41	0.49
Rippon et al 1992	0.41	0.47
Cairns_1978_daphnia	0.40	0.47
Cairns_1978_aelosoma	0.39	0.47
Mahadevan_1986_patriella	0.38	0.43
Dixon_1981_salmo	0.37	0.40
Costa 1965a	0.36	0.40
Nalecz-Jawecki_1998_Spirotomum	0.31	0.34
Stanley_1974_myrio	0.28	0.28

Table 5: Reliability and relevance scores of the toxicological data based on different disputability scores. (a) First column disp=0.3, (b) Second column disp=0.1

Default disputability (disp=0.3)

The minimum score (0.28) is observed for the test on '*Myriophyllum spicatum'* from Stanley (1974) while the maximum (0.67) is observed for the test on '*Chironomus riparius'* from Bertow (2011). The tests have an average score of 0.5. Most tests, whose conditions varied significantly between optimum and non-optimum, received a middle level reliability and relevance score (0.40-0,60).

The MCDA methodology allows the analysis of each test, based on the preferences of each involved expert. The variability that is observed among experts for the scores of an individual test are based on the different input provided by each expert, regarding the importance of the assessment criteria and the relations among criteria, as those are expressed in the questionnaire for expert consultation and stored in the knowledge base. In practice, the variability is one of the main cores of the methodology, as it allows to evaluate and assess ecotoxicological data on the basis of different experts' opinions based on the same criteria hierarchy.

The methodology allows extracting further information regarding the analysis of each specific test for every single node of the criteria hierarchy. As an example, we can see in Table 6 how well the different tests perform for the 'Experimental Reliability', 'Statistical Reliability' and 'Biological Relevance' Lines of Evidence (LoEs) test respectively.

Most of the analysed tests, perform quite well in the 'Experimental Reliability' LoE, with scores in the range 0.50-0.80. Four test have received the maximum score (0.84 - test on '*Chlamys asperrimus*' from Pablo 1997, test on '*Lemna gibba*' from Bertow 2011, tests on 'Chironomus riparius' from Bertow 2011), while the tests on '*Myriophyllum spicatum*' from Stanley 1974 and on '*Spirostomum ambiguum*' from Nalecz-Jawecki 1998 have the lowest performance (0.38 and 0.41 respectively). Most of the tests do not perform well on the 'Statistical Reliability' LoE, as it observed that most of the scores are in the range 0-0.5, with significant number of tests below 0.3. Two tests have score 0 (test on '*Chlamydomonas*' from Cairns 1978 and 'Patiriella calcar' from Mahadevan 1986), which indicates complete unreliability for the statistical aspects of the performed tests. Similarly with the 1st LoE, the scores of the tests for the third LoE (Biological Relevance) are fairly good and in the range 0.49-0.72. Only one test has received a score of 0.38 (test on 'Oncorhynchus mykiss' from Dixon 1981).

In a similar way, we can analyse and compare every assessed test and ecotoxicological datum based on every criteria group, ecotoxicological category and line of evidence of the WoE framework. In this way, apart from the overall reliability score, different tests can be quantitavely evaluated, based on their various other different characteristics.

	Experimental Reliability	Statistical Reliability	Biological Relevance
Calleja 1 Artoxkit	0.73	0.34	0.49
Calleja 2 Streptoxkit	0.76	0.34	0.49
Pablo acute_Nitzschia	0.72	0.48	0.65
Pablo chronic_Nitzschia	0.72	0.47	0.68
Cairns_Physa	0.64	0.03	0.65
Cairns_1978_anculosa	0.64	0.31	0.66
Smith 1978 1 Pim	0.71	0.46	0.65
Smith 1978 2 Perca	0.71	0.46	0.65
Smith 1979 1 pomoxis	0.71	0.43	0.65
Smith 1979 2 gammarus	0.71	0.46	0.65
Broderius 1 Lep	0.75	0.59	0.65
Broderius 2 Gast	0.75	0.59	0.65
Brix_2000_Cancer spp	0.72	0.43	0.72
Call_1983_Tanytarus	0.68	0.38	0.61
Carballo_1995_O.mykiss	0.60	0.15	0.70
Nalecz-Jawecki_1998_Spirotomum	0.41	0.38	0.49
Steele_1983_Champia	0.63	0.11	0.71
Pablo_1997_chlamys_EC50	0.84	0.45	0.63
Pablo_1997_chlamys_NOEC	0.74	0.24	0.71
Slabbert_1986_tetrahymena	0.58	0.63	0.63
Cairns_1978_chlamy	0.70	0.00	0.72
Cairns_1978_aelosoma	0.50	0.35	0.63
Cairns_1978_philodina	0.52	0.35	0.63
Cairns_1978_daphnia	0.50	0.33	0.62
Bhunia_2000_moina	0.55	0.57	0.57
Mahadevan_1986_patriella	0.58	0.00	0.47
Dixon_1981_salmo	0.54	0.11	0.38
Stanley_1974_myrio	0.38	0.13	0.51
Rippon et al 1992	0.51	0.25	0.59
Lemna test	0.84	0.49	0.66
Chironomids Chronic toxicity	0.84	0.49	0.69
Chironomids Acute toxicity	0.84	0.46	0.67
Leduc 1966	0.52	0.38	0.63
Oseid and Smith 1979	0.46	0.42	0.64
Costa 1965a	0.45	0.25	0.55
Jee and Kang 1999	0.61	0.27	0.45
Bhunia_2000_branchiura	0.72	0.46	0.57

Table 6: Scores of the tests for the three Lines of Evidence. Columns: (a) Experimental reliability, (b) Statistical reliability, (c) Biological relevance

5.2.1. SSWD graphs

In this study the proposed MCDA methodology has been used for the assessment of the available ecotoxicological data and the SSWD method was applied to the dataset of ecotoxicological data on cyanide, which has been described in the paragraph 5.1.3. Moreover, for comparison purposes, the conventional SSWD (all data equally weighted) was applied to the same set of data.

The weighting coefficients used for the production of the SSWD are the reliability and relevance scores that have been calculated with the use of the MCDA methodology and are reported in Table 5, while for the conventional SSWD the weighting coefficients were set equal to 1 for all the ecotoxicological data.

The produced SSWD graphs are used in the following two ways, as reported by Posthuma et al. (2002):

1. The forward use
2. The inverse use

The forward use, ecological risk assessment, requires estimation of the ambient concentration of a compound at a contaminated site, in our case the PEC module results (paragraph 5.3.1), or the concentration predicted to result from a proposed use (X-axis). The potentially affected fraction (PAF) at that concentration can then be estimated using the SSWD. The various SSWD graphs for each taxonomic group (i.e. All data, Vertebrates, Invertebrates) and the different trophic levels (i.e. Primary producers, Primary consumers, Secondary consumers) are reported in the following sub-paragraphs and specifically in Figure 33 for all the ecotoxicological data, in Figure 34 for the Invertebrates, in Figure 35 for the Vertebrates, in Figure 36 for the primary producers, in Figure 37 for primary consumers and in Figure 38 for the secondary consumers. Each figure consists of a set of 4 SSWD graphs, which report the log-empirical and the log-normal SSWD which are created with the use of the MCDA methodology (graphs (a) and (b), while the log-empirical and log-normal graphs of the conventional SSWD (all weights equal to 1) are reported in graphs (c) and (d). The results of the forward use of the SSWDs and the risk estimations for the four stations in France are reported in paragraph 5.3.2.

For the inverse use, such as the derivation of environmental quality criteria, a cutoff percentage p is chosen (to protect 1-p percent of species, Y-axis), and the desired “safe” concentration (HCp) is calculated as a result. The 5th percentile of a chronic toxicity distribution has been chosen in the earliest methods as a concentration that is protective for most species in a community, but the value of p is a policy decision, not science. In this study, the cutoff percentages p=5 and p=50 are reported, and the HC_5 and HC_{50} values are presented in Table 7 and Table 8 for the different taxonomic groups (i.e. All data, Vertebrates, Invertebrates) and the different trophic levels (i.e. Primary producers, Primary consumers, Secondary consumers). This approach is used to derive ecological quality criteria (EQCs). This process allowed the identification of the most sensitive trophic level and taxonomic group for the environmental compartments of concern.

Weighted SSD curves

For all the ecotoxicological data, it can be observed that both log-empirical and log-normal SSWD curves have the same range of distribution for the best estimate curves, which spans from 0.01 to 1000 µg/L (Figure 33, a and b).

The log-normal curve has a high R^2 value (0.949), which indicates a good fitting of the ecotoxicological data, while at the same time the KSp value is 0.5 (pvalue of the Kolmogorov-Smirnov goodness of fit test, with Dallal-Wilkinson approximation) is greater than 0.1 and indicates a good fitting of the data as well.

For the Invertebrates, we observe that both log-empirical and log-normal SSWD curves have almost the same range of distribution for the best estimate curves, which spans from 0.01 to 1000 μg/L (Figure 34, a and b).

The log-normal curve has a high R^2 value (0.944), and a good KS pvalue is 0.5.

For the Vertebrates, it can be observed that the log-normal SSWD curve has larger range of distribution, which spans from 0.1 to 30 μg/L from the log-empirical curve, which spans from 0.4 to 30 μg/L (Figure 35, a and b).

The log-normal curve has a R^2 value which is lower than 0.9 (0.832), therefore the fitting of the data is not optimal, while the KS pvalue is good (0.5).

For the Primary producers, it can be observed that the log-normal SSWD curve has larger range of distribution, which spans from 0.03 to 360 μg/L, from the log-empirical curve, which spans from 0.17 to 170 μg/L (Figure 36, a and b).

The log-normal curve has a high R^2 value (0.932), and a good KS pvalue (0.5).

For the Primary consumers, it can be observed that both log-empirical and log-normal SSWD curves have the same range of distribution for the best estimate curves, which spans from 0.1 to 1000 μg/L (Figure 37, a and b).

The log-normal curve has a very high R^2 value (0.983), and a good KS pvalue (0.5).

For the Secondary consumers, it can be observed that both log-empirical and log-normal SSWD curves have the same range of distribution for the best estimate curves, which spans from 0.01 to 100 μg/L (Figure 38, a and b).

The log-normal curve does not have a high R^2 value (0.749), neither a good KS pvalue (0.05). In this case, the curve does not represent well the experimental data, therefore the results are deemed unreliable.

No specific pattern has been observed for the sensitivity of different species, based on the various SSWD graphs (All data, primary producers, primary consumers, secondary consumers), since the different species are distributed alongside the SSWD curves.

Conventional SSWD curves (W=1)

For all the ecotoxicological data, both log-empirical and log-normal SSWD curves have the same range of distribution for the best estimate curves, which spans from 0,01 to 1000 μg/L (Figure 33, c and d).

The log-normal curve has a high R^2 value (0.936), which indicates a good fitting of the ecotoxicological data, while at the same time the KSp value is 0.5 and indicates a good fitting of the data as well.

For the Invertebrates, it can be observed that the log-empirical SSWD curve has a slightly different range of distribution, which goes from a bit below 0.01 to 1000 μg/L, while the log-normal curve has range of distribution, which spans from 0.01 to 1000 μg/L (Figure 34, c and d).

The log-normal curve has a high R^2 value (0.912), and a good KS pvalue is 0.5.

For the Vertebrates, it can be observed that the log-normal SSWD curve has larger range of distribution, which spans from 0. 1 to 30 μg/L, from the log-empirical curve, which spans from 0.4 to 30 μg/L (Figure 35, c and d).

The log-normal curve has a good R^2 value which is almost 0.9 (0.893) but it is still acceptable, while the KS pvalue is good (0.5).

For the Primary producers, it can be observed that the log-normal SSWD curve has larger range of distribution, which spans from 0.03 to 800 μg/L, from the log-empirical curve, which spans from 0.2 to 280 μg/L (Figure 36, c and d).

The log-normal curve has a high R^2 value (0.95), and a good KS pvalue (0.5).

For the Primary consumers, it can be observed that both log-empirical and log-normal SSWD curves have the same range of distribution for the best estimate curves, which spans from 0.1 to 1000 μg/L (Figure 37, c and d).

The log-normal curve has a very high R^2 value (0.982), and a good KS pvalue (0.5).

For the Secondary consumers, it can be observed that the log-empirical and log-normal SSWD curves do not have the same range of distribution for the best estimate curves, as the log-empirical distribution from 0.003 to 100 μg/L, while the log-normal curve spans from 0.008 to 230 μg/L (Figure 38, c and d).

The log-normal curve does not have a high R^2 value (0.774), neither a good KS pvalue (0.06). In this case, the curve does not represent well the experimental data, therefore the results are deemed unreliable.

No specific pattern has been observed for the sensitivity of different species, based on the various SSWD graphs (All data, primary producers, primary consumers, secondary consumers), since the different species are distributed alongside the SSWD curves.

Comparison of SSWD and conventional SSWD

For all ecotoxicological data, all curves have the same range of distribution, while both log-normal curves have good fitting of the data, as indicated by the two coefficients (R^2 and KSp value). The R^2 is slightly higher for the SSWD curve, which shows a slightly better fitting of the data with the use of the MCDA methodology for the assessment of the ecotoxicological data.

For invertebrates, three curves have the same range of distribution and the log-empirical curve has a slightly greater range, while both log-normal curves have good fitting of the data, as indicated by the two coefficients (R^2 and KSp value). The R^2 is again slightly higher for the SSWD curve.

For vertebrates, both log-empirical curves have the same range of distribution, which is smaller than the range of the log-normal curves. The log-normal curve of the conventional SSWD has a better fitting of the data, as indicated by the two coefficients (R^2 and KSp value). In the case of vertebrates, the R^2 is higher for the conventional SSWD curve.

For the primary producers, it can be observed that the log-empirical and log-normal curves of the conventional SSWD have greater range of distribution in comparison with the corresponding curves of the SSWD. Both log-normal curves have good fitting of the data, as indicated by the two coefficients (R^2 and KSp value). In the case of primary producers, the R^2 is slightly higher for the conventional SSWD curve.

For the primary consumers, all curves have the same range of distribution, while both log-normal curves have very good fitting of the data, as indicated by the two coefficients (R^2 and KSp value).

For the secondary producers, it can be observed that the log-empirical and log-normal curves of the conventional SSWD have slightly greater range of distribution in comparison with the corresponding curves of the SSWD. The log-normal curves do not have a very good fitting of the data, as indicated by the two coefficients (R^2 and KSp value). In the case of secondary producers, the R^2 is slightly higher for the conventional SSWD curve. In both cases, the results are deemed unreliable and are not taken into consideration for further comparisons with other taxonomic groups or trophic levels.

The weighted Species Sensitivity Distributions have been produced in two different ways and the results have been analysed, evaluated and compared. A conservative behaviour is expected with the production and use of the MCDA based SSWD graphs in the risk assessment process, especially for higher concentrations, in comparison with the use of the conventional SSWD graphs.

5.2.1.1. SSWD - All data

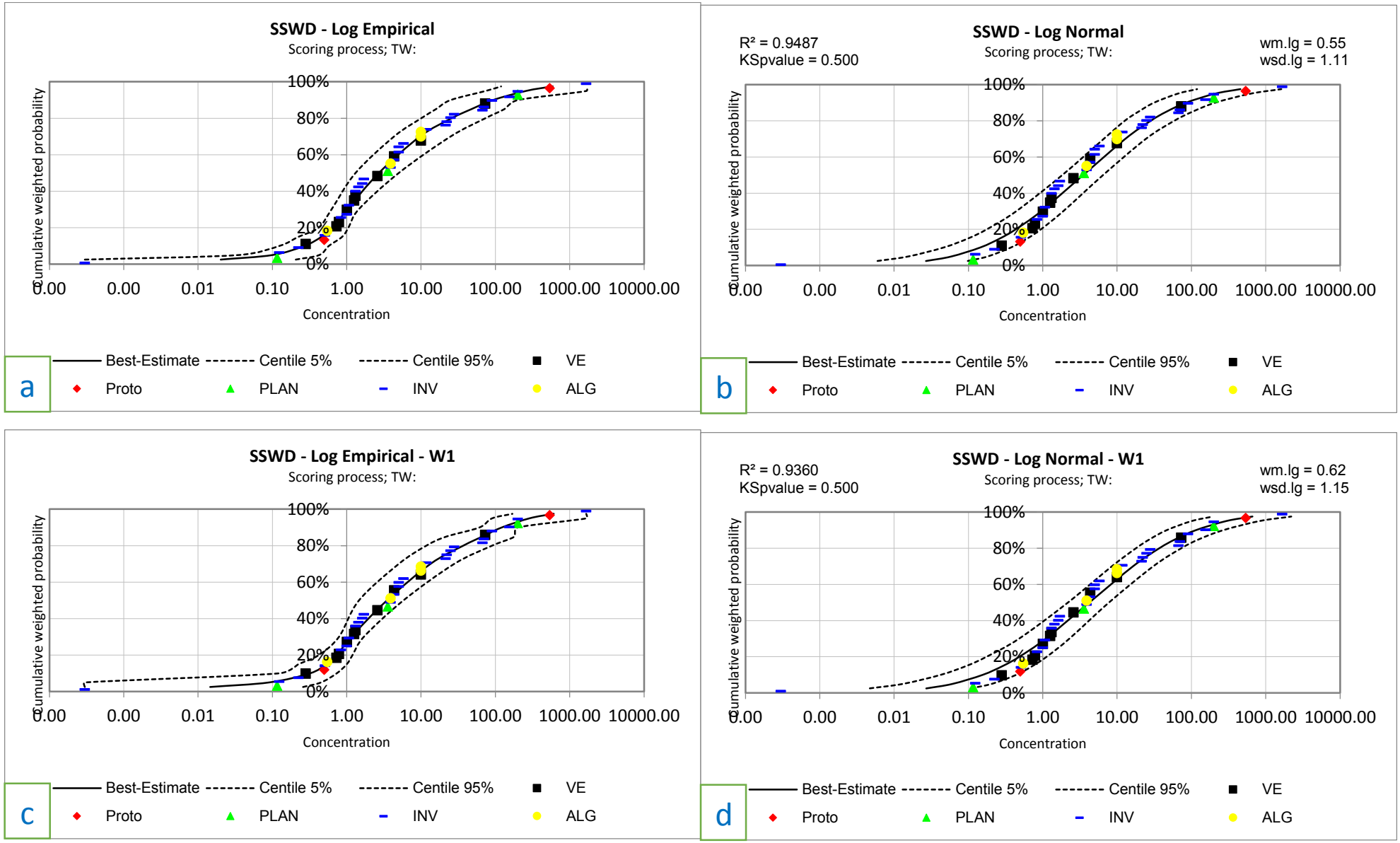


Figure 33: Log-empirical (a, c) & log-normal (b, d) SSWD curves for all data. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.1.2. SSWD – Invertebrates

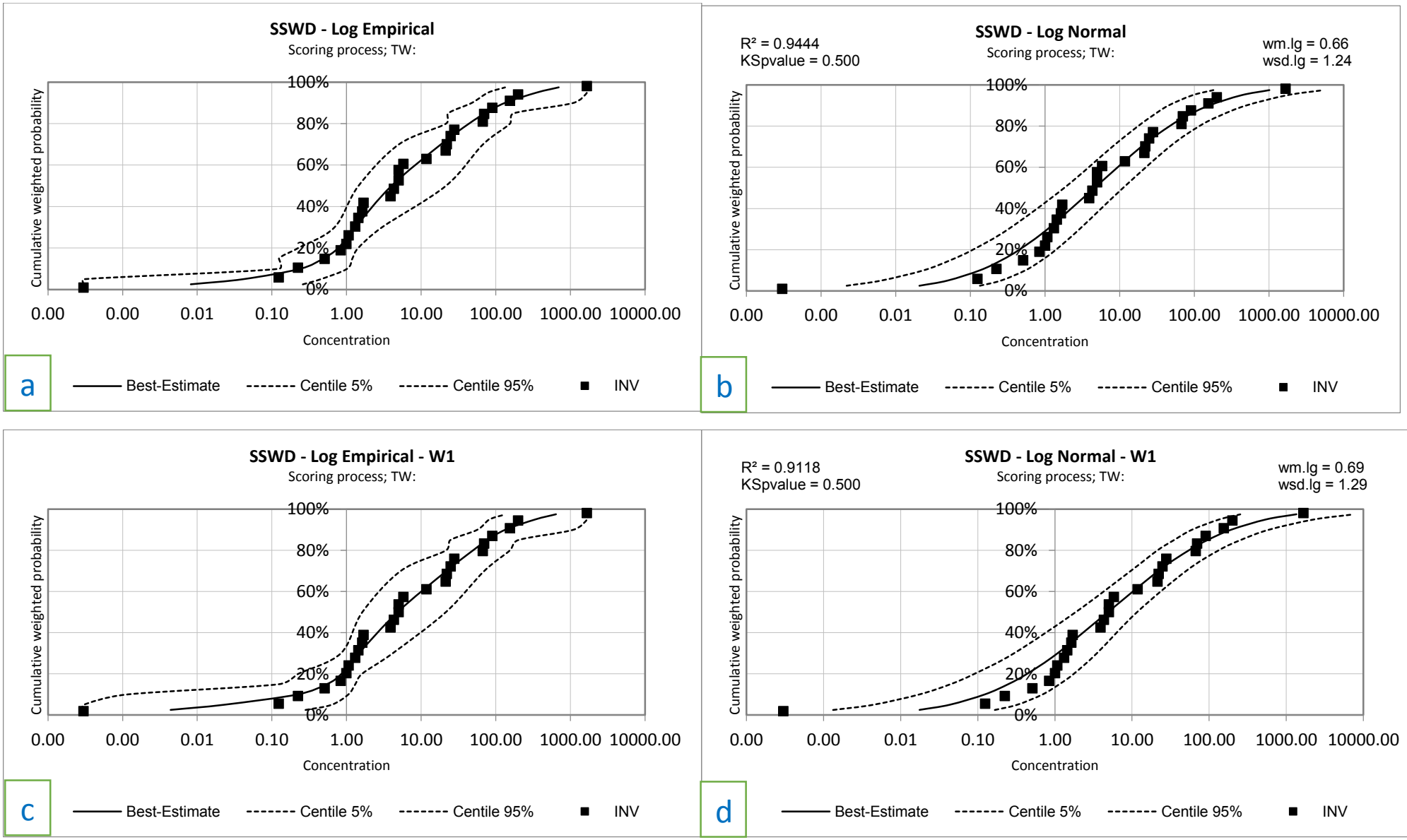


Figure 34: Log-empirical (a, c) & log-normal (b, d) SSWD curves for invertebrates. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.1.3. SSWD – Vertebrates

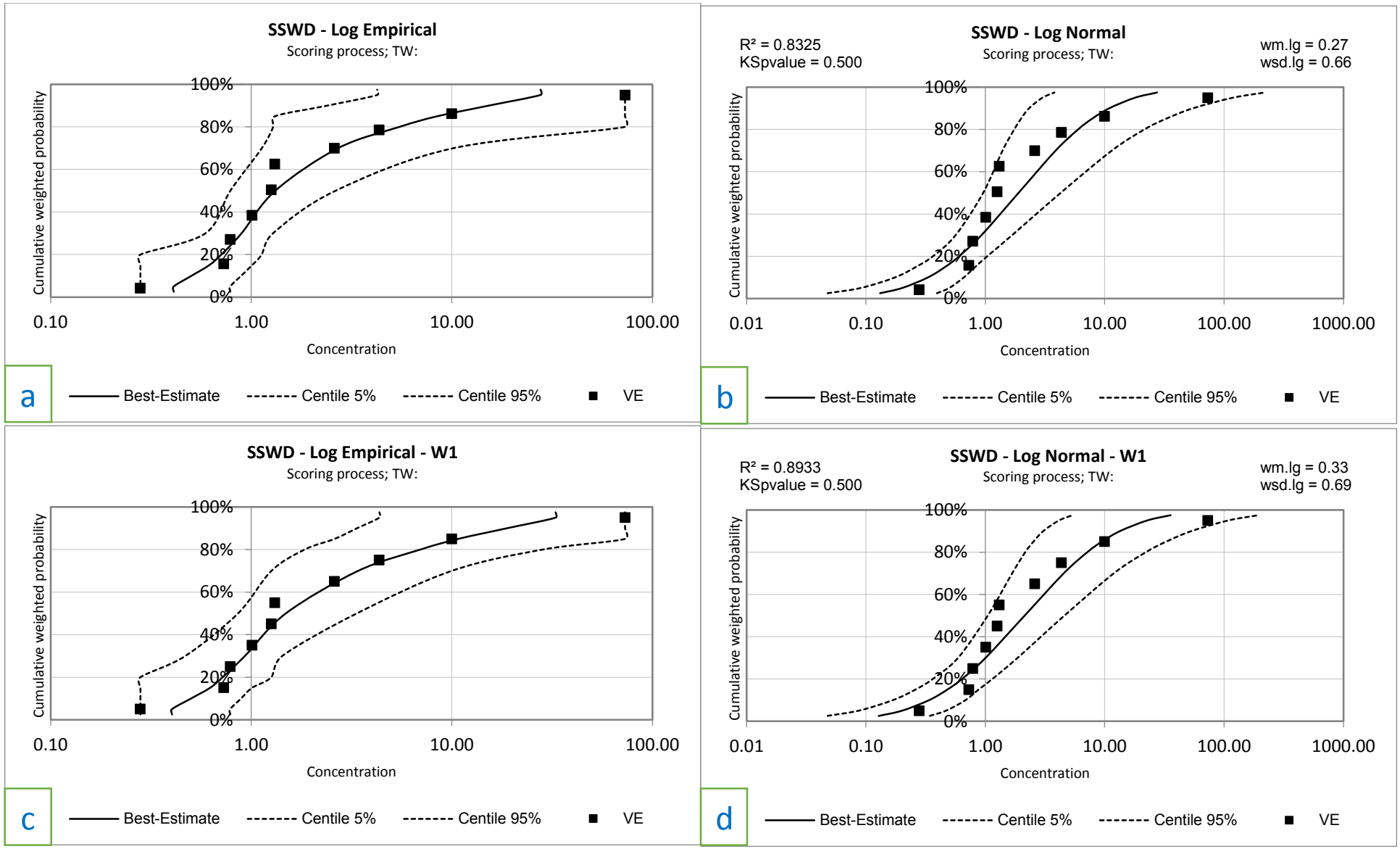


Figure 35: Log-empirical (a, c) & log-normal (b, d) SSWD curves for vertebrates. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.1.4. SSWD – Primary producers

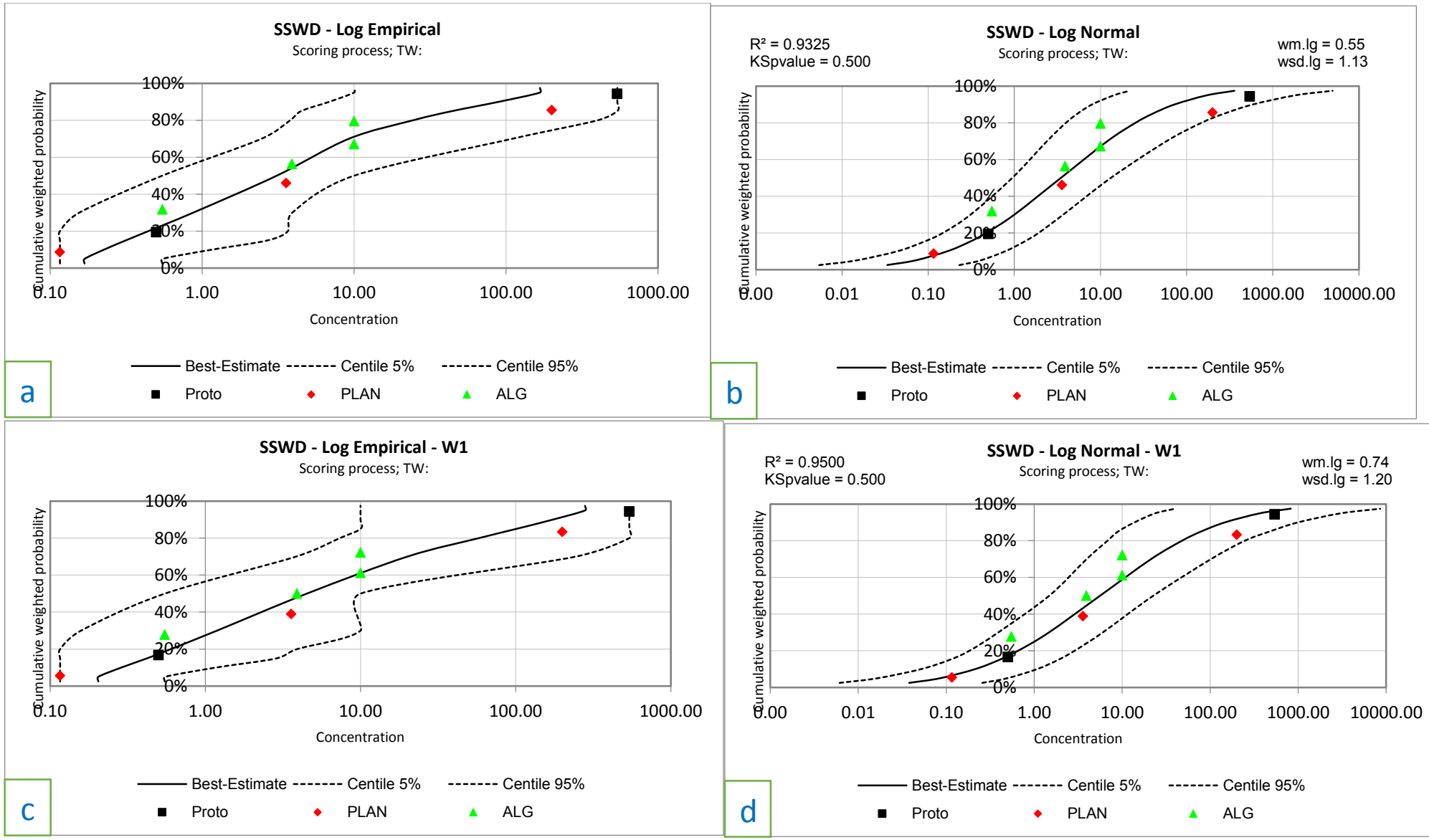


Figure 36: Log-empirical (a, c) & log-normal (b, d) SSWD curves for primary producers. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.1.5. SSWD – Primary consumers

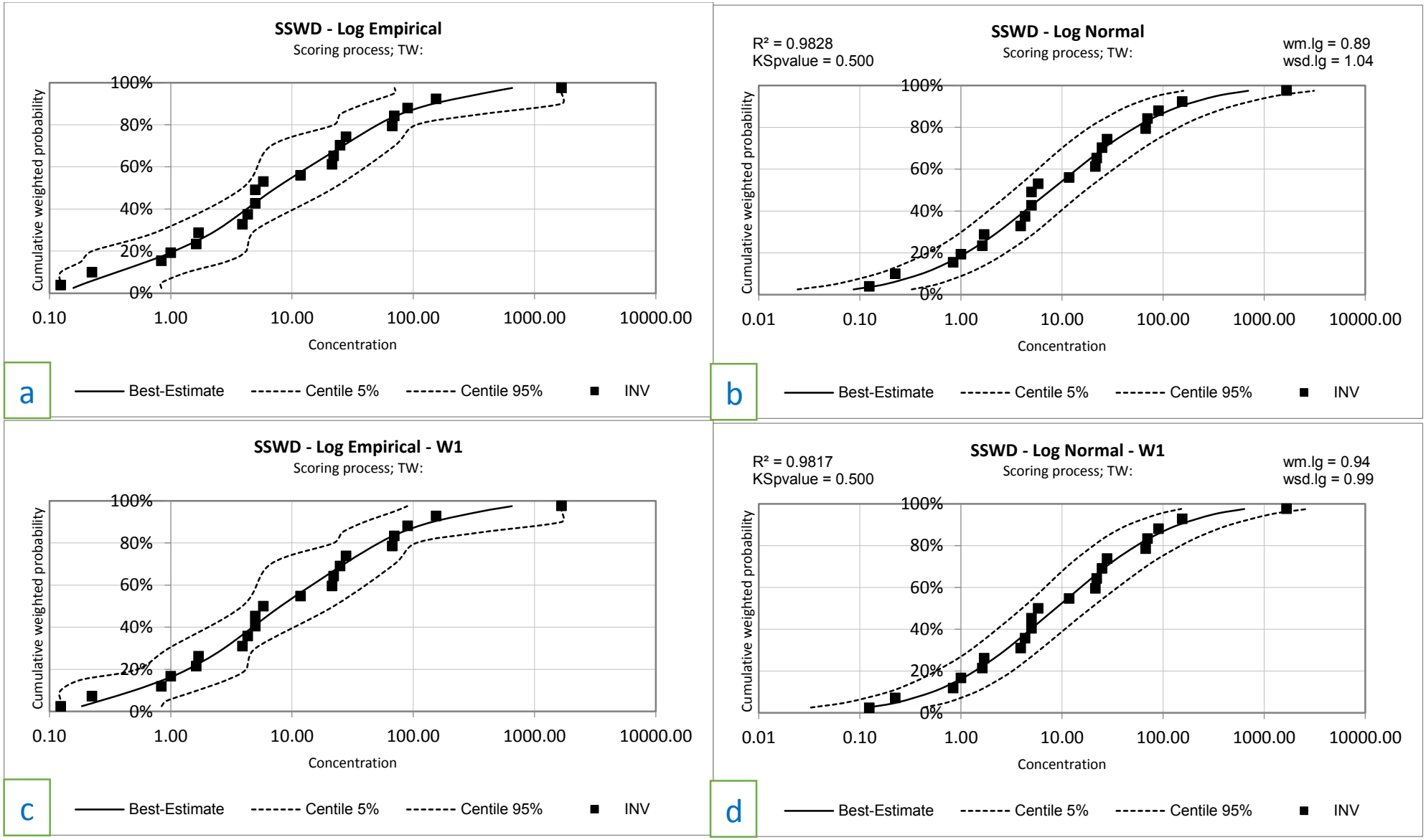


Figure 37: Log-empirical (a, c) & log-normal (b, d) SSWD curves for primary consumers. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.1.6. SSWD – Secondary consumers

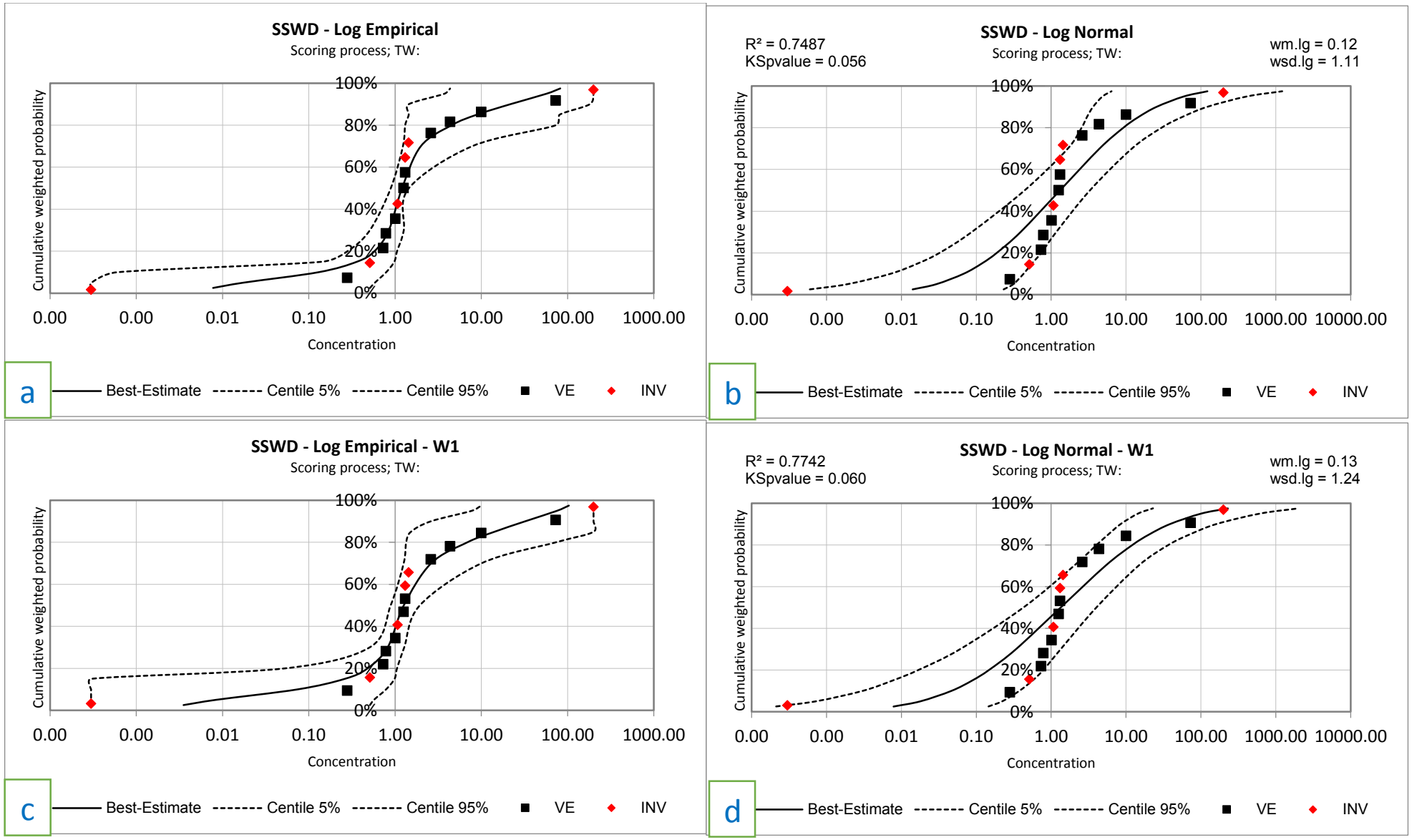


Figure 38: Log-empirical (a, c) & log-normal (b, d) SSWD curves for secondary consumers. MCDA based SSWDs shown in graphs (a) and (b), conventional SSWD are shown in graphs (c) and (d).

5.2.2. Hazardous Concentrations (HCx)

The summary of the calculated Hazardous Concentrations (HC_5 and HC_{50}, best estimate value with 50% confidence interval) for each of the produced SSWD are reported in Table 7 for (a) All data, (b) Vertebrates and (c) Invertebrates and in Table 8 for (a) Primary Producers, (b) Primary Consumers and (c) Secondary Consumers.

It can be observed that the MCDA based SSWD curve present an evident conservative behaviour for the HC_{50} values in comparison with the conventional SSWD curves, since the HC_{50} values for all the log-normal and log-empirical curves of the taxonomic groups and trophic levels are lower for the MCDA based SSWD curves. The differences vary from 0.03-0.06 for the secondary consumers trophic level, up to 1.42-2.13 for the primary producers trophic level, while the range is 0.17-0.91 for the primary consumers, the vertebrates, the invertebrates and the complete set of data.

On the other hand, the HC_5 values vary. In detail, the HC_5 values are slightly lower for the MCDA based SSWD curves for the log-normal and log-empirical curves of the primary producers (0.07 and 0.17 μg/L) in comparison with the respective values of the conventional SSWD curves (0.08 and 0.20 μg/L). Similarly, they are slightly lower also for the MCDA based SSWD curves for the log-normal and log-empirical curves of the primary consumers (0.18 and 0.20 μg/L) in comparison with the respective values of the conventional SSWD curves (0.24 and 0.26 μg/L). The HC_5 values of the MCDA based SSWD curves for all the ecotoxicological data, the vertebrates, the invertebrates and the secondary consumers are slightly higher (0.01-0.02) from the respective values of the conventional SSWD.

According to the obtained HC_5 values it is observed that primary producers are found to be the most sensitive trophic level according to both log-normal and log-empirical curves, while Invertebrates are the most sensitive taxonomic group. The analysis of the HC_{50} values, confirms the sensitivity of the primary producers in comparison with the primary consumers whereas vertebrates are found to be more sensitive than invertebrates for higher cyanide concentrations. In practice, for low cyanide concentrations (around 0.03-0.07 μg/L) the first trophic level to be affected is primary producers and the first taxonomic group to be affected is Invertebrates, while for high concentrations the most endangered trophic level is primary producers and the most sensitive taxonomic group is Vertebrates (for which 50% of species is affected for concentrations below 2 μg/L).

		All data		Vertebrates		Invertebrates	
	HC_x (mg/L)	5%	50%	5%	50%	5%	50%
Log-normal	Best-Estimate (C.I. 50%)	0.06	3.50	0.20	1.90	0.05	4.59
Log-empirical		0.10	2.71	0.41	1.32	0.04	4.01
		All data (W1)		Vertebrates (W1)		Invertebrates (W1)	
	HC_x (mg/L)	5%	50%	5%	50%	5%	50%
Log-normal	Best-Estimate (C.I. 50%)	0.06	4.24	0.20	2.14	0.04	4.88
Log-empirical		0.09	3.48	0.40	1.49	0.02	4.83

Table 7: HC_5 and HC_{50} values in (mg/L) for (a) All data, (b) Vertebrates and (c) Invertebrates, reported for the MCDA based SSWD and the conventional SSWD (W1)

		PRIMARY PRODUCERS		PRIMARY CONSUMERS		SECONDARY CONSUMERS	
	HC_x (mg/L)	5%	50%	5%	50%	5%	50%
Log-normal	Best-Estimate (C.I. 50%)	0.07	3.47	0.18	7.77	0.03	1.31
Log-empirical		0.17	3.09	0.20	7.38	0.02	1.21
		PRIMARY PRODUCERS (W1)		PRIMARY CONSUMERS (W1)		SECONDARY CONSUMERS (W1)	
	HC_x (mg/L)	5%	50%	5%	50%	5%	50%
Log-normal	Best-Estimate (C.I. 50%)	0.08	5.59	0.24	8.67	0.02	1.34
Log-empirical		0.20	4.51	0.26	8.08	0.01	1.26

Table 8: HC_5 and HC_{50} values in (mg/L) for (a) Primary producers, (b) Primary consumers and (c) Secondary consumers, reported for the MCDA based SSWD and the conventional SSWD (W1)

5.3. Predicted Environmental Concentrations (PEC) and Potentially Affected Fraction (PAF)

Following the calculations of the reliability and relevance scores of the ecotoxicological data used in the case study and the production of the SSWD graphs, the next steps in the probabilistic ecological risk assessment are: (1) the calculation of the Predicted Environmental Concentrations (PEC – paragraph 5.3.1) for each station of the case study, through the production of the Probability Density Functions of cyanide and (2) the calculation of the Potentially Affected Fraction (PAF – paragraph 5.3.2) through the production of the Joint Probability Curves from the cumulative distributions of the calculated PECs for each station and the produced SSWDs for the sensitivity of species on cyanide. The application of the DSS and the related results are described in the next paragraphs.

5.3.1. Predicted Environmental Concentrations (PEC)

The collected measurements of cyanide for the four stations have been inserted in the 1st module of the DSS for the statistical analysis of the measured concentration and the production of the 'Probability Density Functions' (PDF) of cyanide for the case study.

The data have been considered as a continuous set of measurements, without taking into consideration their temporal information. In order to estimate the distribution of the observed concentrations and their probabilities of appearance, the 'Winsorized' distributional method has been selected, as it has been considered an appropriate and interesting statistical technique for estimating the PDF graphs. This is due to the fact that mean and standard deviation have been evaluated from the original data without taking into consideration undetected (missing) values. New data have been generated, following the behaviour of a Normal variable with mean and standard deviation previously computed as parameters.

The measurements of cyanide in µg(CN)/L for the period 01/01/2005 – 20/08/2014, for the four stations of the case study, are presented in Table 9. The selected underlying model has been selected as 'Normal' and the Kernel parameters (kernel function and smoothness) have both been selected as 'Normal' respectively due to their convenient mathematical properties and the best estimated fit to the available dataset. The results of the application of the Exposure Assessment module of the DSS to the available exposure data are presented for each station separately, with the 4 graphs that are produced by the software.

The graphs report the predicted probabilities of appearance of the various concentration levels of cyanide through a normal curve, an empirical PDF and a kernel density estimation. In addition, a comparison of the three graphs is provided. The four graphs for each station can be seen in Figure 39 (La Cance), Figure 40 (L'Airon), Figure 41 (L'Yvrande) and Figure 42 (La Selune).

As expected from the available data, a higher probability of appearance of cyanide concentration is predicted for station 3 (03272235, L'Yvrande) in comparison with stations 1, 2 and 4 (La Cance, L'Airon and La Selune respectively). The PEC graphs show a significant higher predicted concentration for station 3, where there is probability of appearance of concentrations up to 10 µg(CN)/L and higher, whereas the other three stations are restricted to lower concentrations, around 1.33 µg(CN)/L, therefore an expected difference by 1 order of magnitude.

2005-2014	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28
3271415	0.71	0.61	0.61	0.64	1.09	0.42	0.48	0.46	0.71	0.42	0.82	0.63	0.75	0.42	0.62	0.25	0.55	0.14	0.14	0.20	0.30	0.51	0.90	0.30	0.41	1.15	0.17	0.55
3271965	0.41	0.62	1.28	0.62	0.65	0.83	0.94	0.39	0.68	0.18	0.35	0.57	0.66	0.46	0.50	0.40	0.79	0.14	0.20	0.30	0.80	0.32	0.34	0.50	0.50	0.14	0.50	0.94
3272235	0.76	0.74	1.41	5.86	3.66	4.17	10.80	2.36	4.45	5.71	3.71	0.20	3.81	5.62	0.90	0.22	0.63	3.00	1.00	4.89	8.29	0.60	1.20	4.95	3.00	0.74	1.00	5.43
3272685	0.48	0.42	0.78	0.51	0.54	0.55	0.70	0.66	0.80	0.60	0.72	0.81	0.58	0.35	0.44	0.55	0.86	0.14	0.30	0.38	0.21	0.40	0.65	0.25	0.46	0.90	0.14	0.20

2005-2014	m29	m30	m31	m32	m33	m34	m35	m36	m37	m38	m39	m40	m41	m42	m43	m44	m45	m46	m47	m48	m49	m50	m51	m52	m53	m54	m55	m56
3272235	0.80	0.14	0.18	0.40	0.14	0.40	0.55	0.14	0.20	0.40	0.86	0.14	0.20	0.43	0.50	0.14	0.21	0.30	0.40	0.14	0.50	0.69	0.30	0.60	0.61	0.16	0.48	0.80
3272685	0.30	0.46	0.77	0.14	0.14	0.40	0.14	0.14	0.20	0.40	0.30	0.38	0.83	0.14	0.67	0.70	0.16	0.45	0.90	0.14	0.14	0.50	0.20	0.40	0.44	0.14	0.30	0.99
3271965	3.00	3.99	7.66																									
3271415	0.47	0.14	0.20	0.40	0.77	0.20	0.20	0.26	0.31	0.14	0.30	0.42	0.50	0.16	0.40	0.68	0.42	0.50	0.68	0.30	0.47	0.69	0.21	0.40	0.44	0.20	0.33	0.83

Table 9: Values of Cyanide in µg(CN)/L for measurements 1-56 for the period 01/01/2005 – 20/08/2014.

5.3.1.1. PEC graphs – Winsorized distributional method

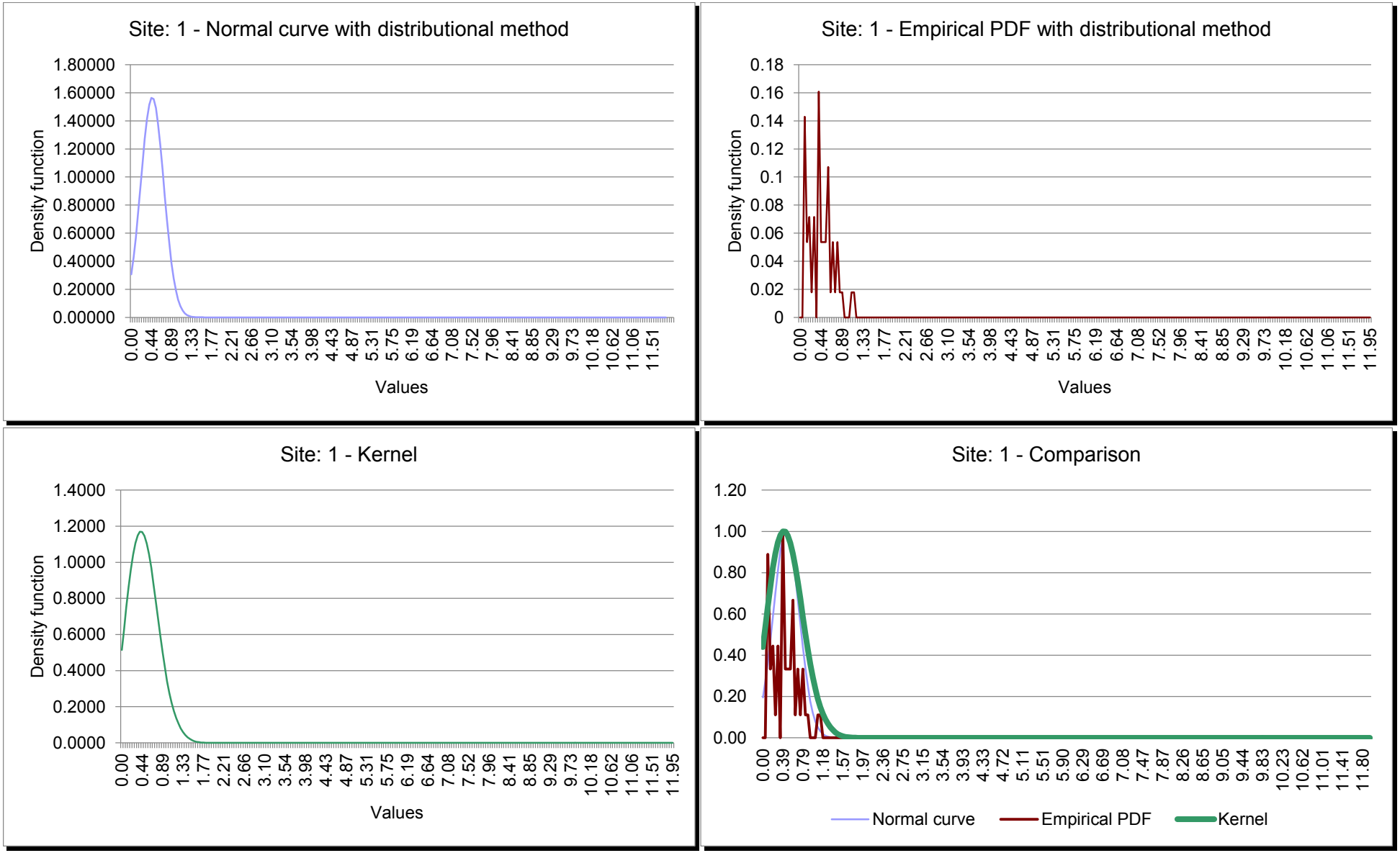


Figure 39: PEC graphs for station 1 (La Cance)

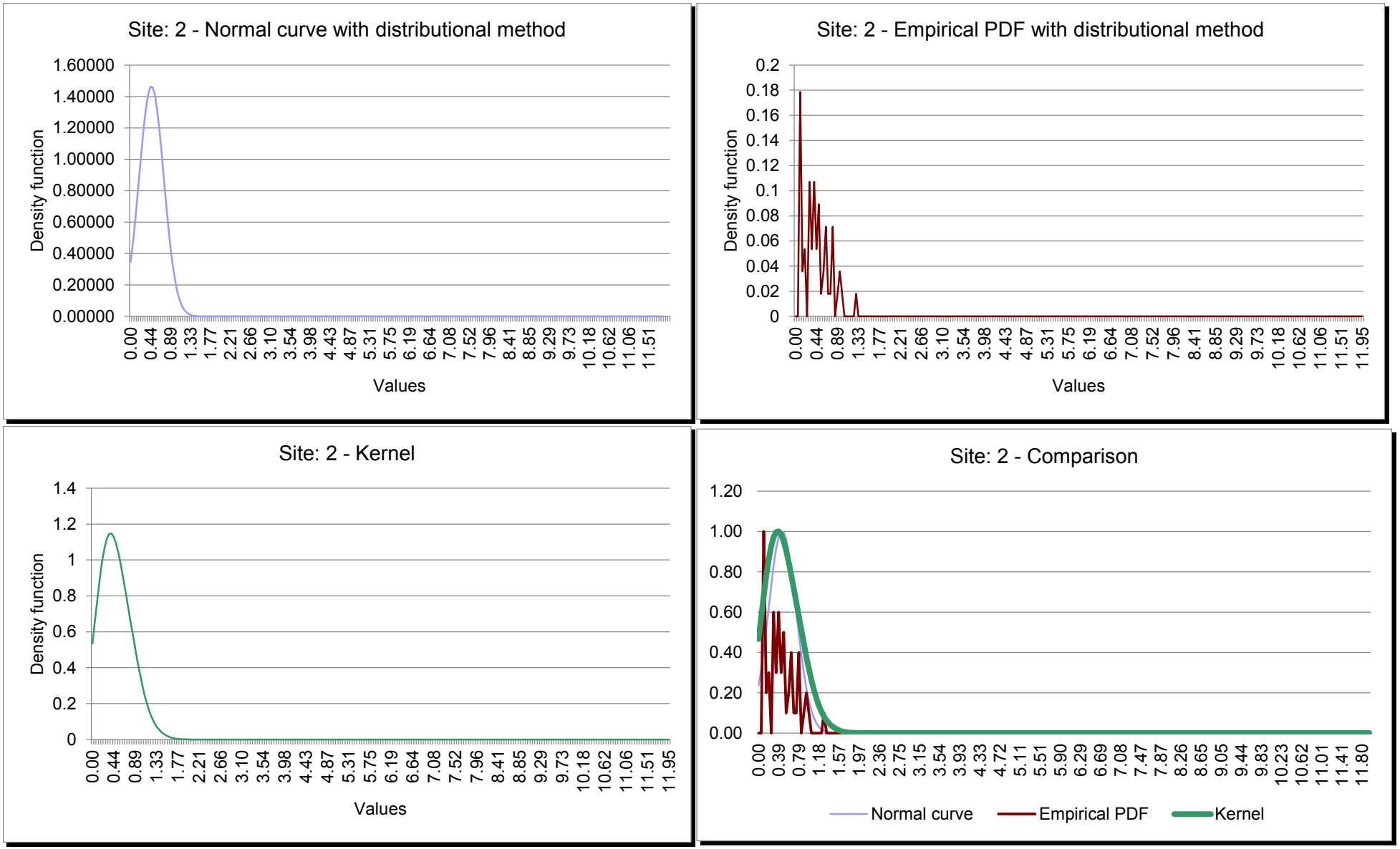


Figure 40: PEC graphs for station 2 (L'Airon)

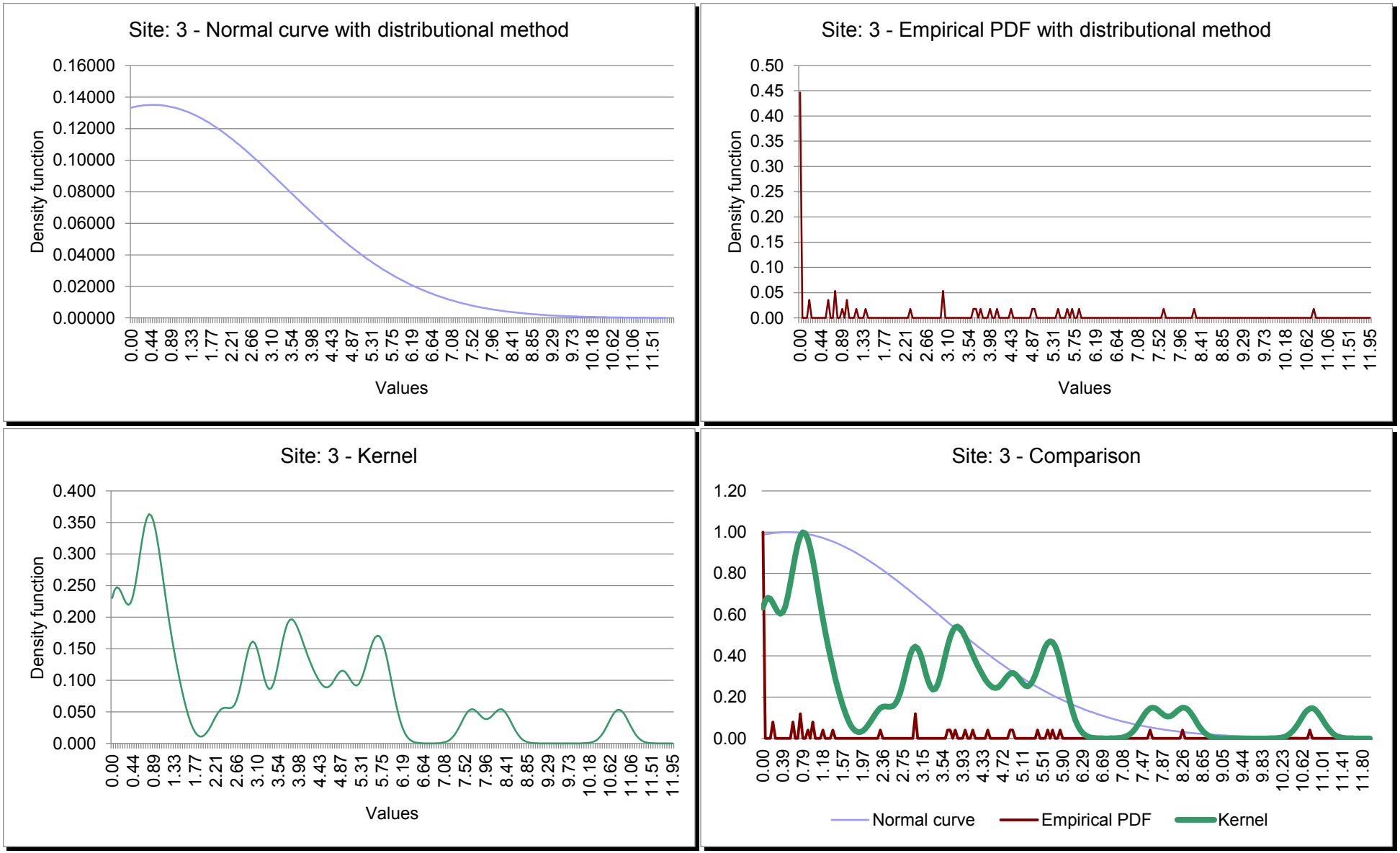


Figure 41: PEC graphs for station 3 (L'Yvrande)

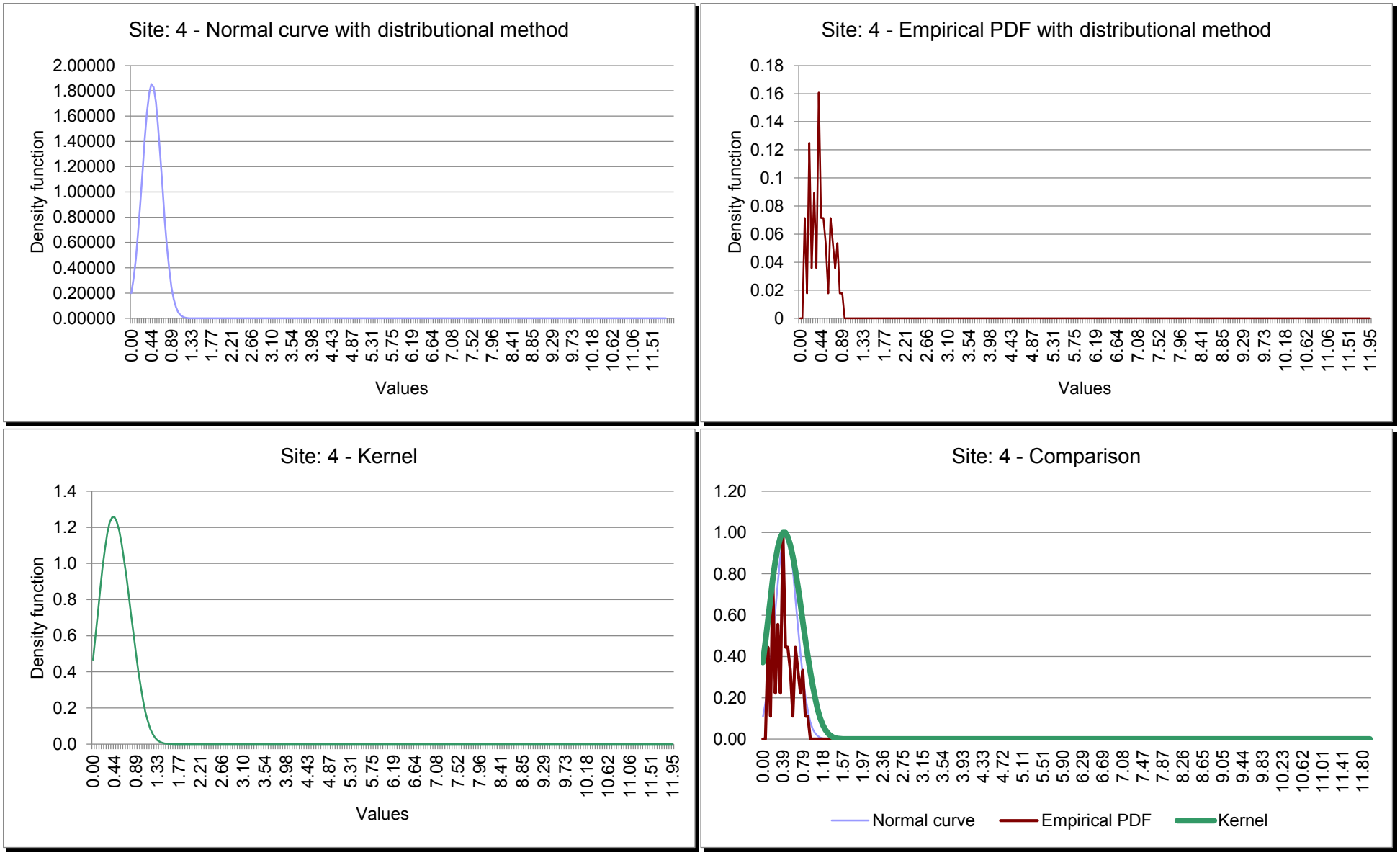


Figure 42: PEC graphs for station 4 (La Selune)

5.3.2. Potentially Affected Fraction (PAF)

As mentioned in paragraph 5.2, the forward use of the SSWD in Ecological Risk Assessment allows the quantification of environmental risk through the calculation of the Potentially Affected Fraction, which expresses the fraction of species which is expected to be affected above its no-effect level at a given environmental concentration.

In the case study, the analysis has been performed for the different taxonomic groups (as well as for all species) and trophic levels. The percentages of risk for the four analysed stations are reported in Table 10 below, for the different taxonomic groups and trophic levels. The first part reports the PAF percentages calculated with the use of the MCDA based SSWD methodology and the second with the conventional SSWD (where, all weights are equal).

5.3.2.1. PAF – Distributional method (winsorised) PEC

Risk - SSWD						
Station code	**All species**	**Invertebrates**	**Vertebrates**	**Primary Prod.**	**Primary Cons.**	**Secondary Cons.**
Station 1 - 03271415	16.4	16.7	11.7	16.6	8.6	27.3
Station 2 - 03271965	16.5	16.8	12.0	16.8	8.7	27.5
Station 3 - 03272235	36.4	35.3	39.3	36.4	32.3	41.6
Station 4 - 03272685	16.0	16.5	11.1	16.3	8.3	27.0
Risk - conventional SSWD						
Station code	**All species**	**Invertebrates**	**Vertebrates**	**Primary Prod.**	**Primary Cons.**	**Secondary Cons.**
Station 1 - 03271415	15.8	17.0	11.1	14.3	6.9	29.0
Station 2 - 03271965	15.9	17.1	11.3	14.4	7.0	29.1
Station 3 - 03272235	35.6	35.0	38.5	34.3	31.6	41.6
Station 4 - 03272685	15.5	16.7	10.5	14.0	6.7	28.7

Table 10: PAF percentages for different taxonomic groups and trophic levels, for the MCDA based SSWD and the conventional SSWD.

The assessment shows a significant risk from the cyanide concentrations for all the species at the four analysed stations and especially for the L'Yvrande station (station 3 - 03272235) at which the 36.4% of species is expected to be affected by the predicted environmental concetrations, while on the other hand, the risk is 16-16.5% for the other three stations, which definitely cannot be neglected.

The analysis of the PAF per trophic level shows, as expected, that the primary producers are the trophic level in higher risk, both for lower expected concentrations (27-27.5% in stations 1, 2, 4) and higher expected concentrations (41.6% in station 3) in comparison with the risk of primary consumers (8.3-8.7% and 32.3% respectively).

The analysis of the PAF per taxonomic groups shows, that Invertebrates are the taxonomic group in higher risk for lower expected concentrations since the PAF percentages (16.5-16.8%)

are higher than the percentages of the Vertebrates (11.1-12.0%) for stations 1, 2 and 4, where the predicted environmental concentrations are low. On the other hand, the reverse behaviour is observed for station 3, where the predicted environmental concentrations are higher than the other stations. Specifically, Vertebrates are found to be in higher risk for higher expected concentrations (39.3%) than Invertebrates (35.3%).

These results are in compliance with the outcomes of the SSWD graphs.

The comparison of the PAF percentages between the MCDA-based SSWD and the conventional SSWD shows that the MCDA-based SSWD shows a higher expected risk for all species, vertebrates, primary producers and primary consumers, whereas the percentages for invertebrates are almost the same for the two methods and the risk percentages for secondary consumers for stations 1, 2 and 4 are slightly lower for the MCDA based SSWD.

The following figures (Figure 43 to Figure 48) show the graphs of the Joint Probability Curve (JPC) and the related PAF percentage for each of the four stations based on the MCDA based SSWD for all species, inverterbates, vertebrates, primary producers, primary consumers and secondary consumers.

5.3.2.2. PAF – All data

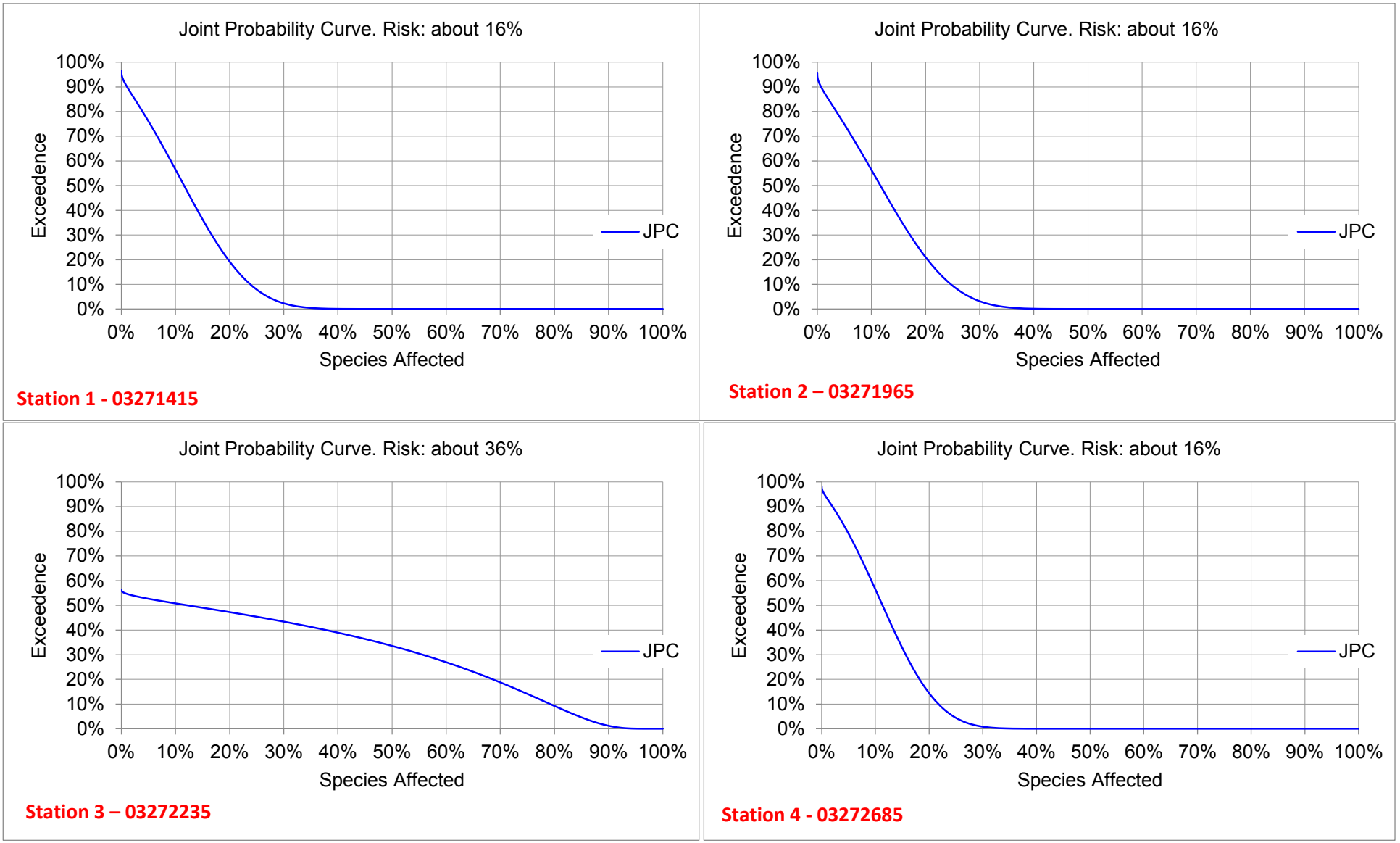


Figure 43: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for all species

5.3.2.3. PAF – Invertebrates

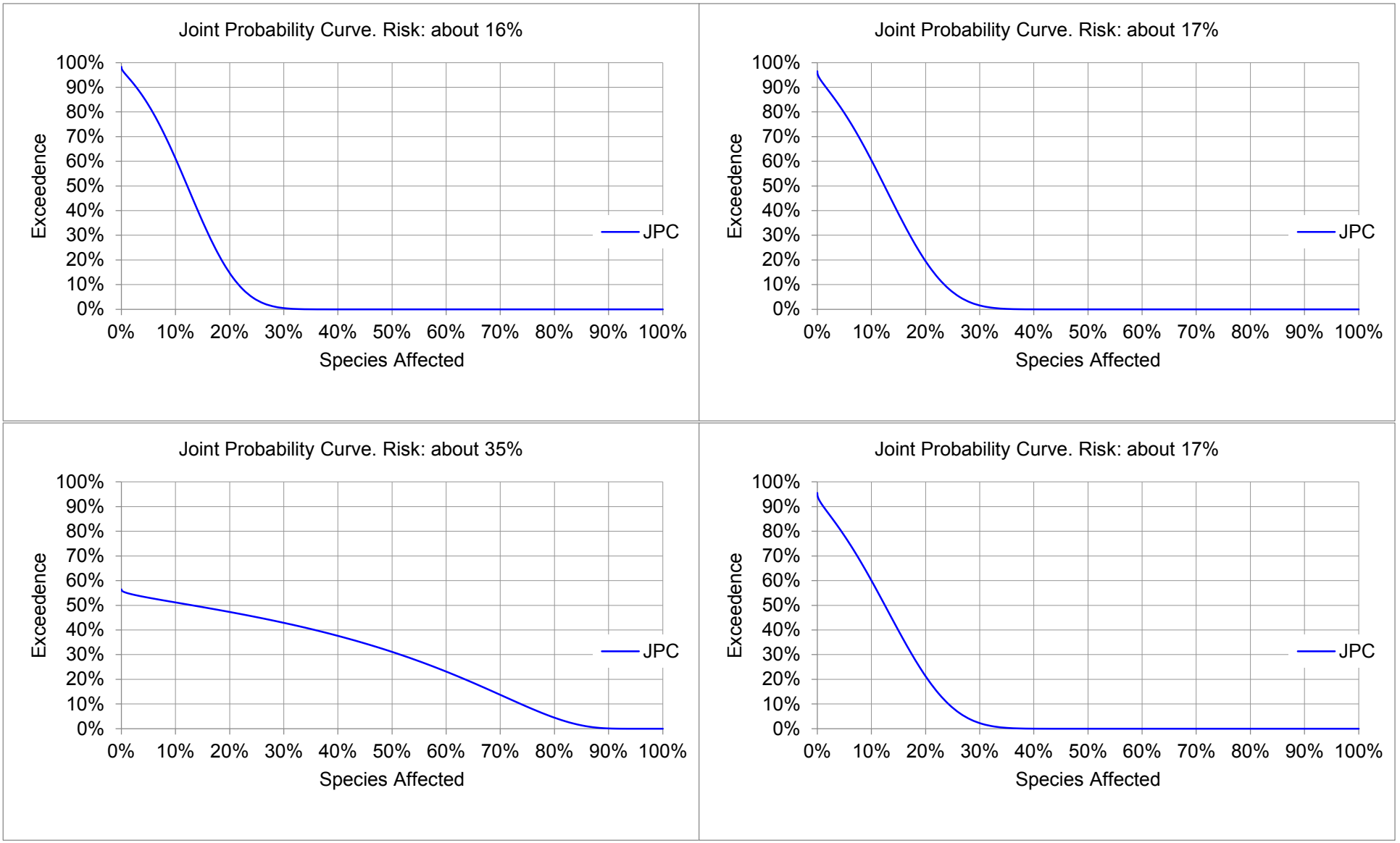


Figure 44: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for invertebrates

5.3.2.4. PAF – Vertebrates

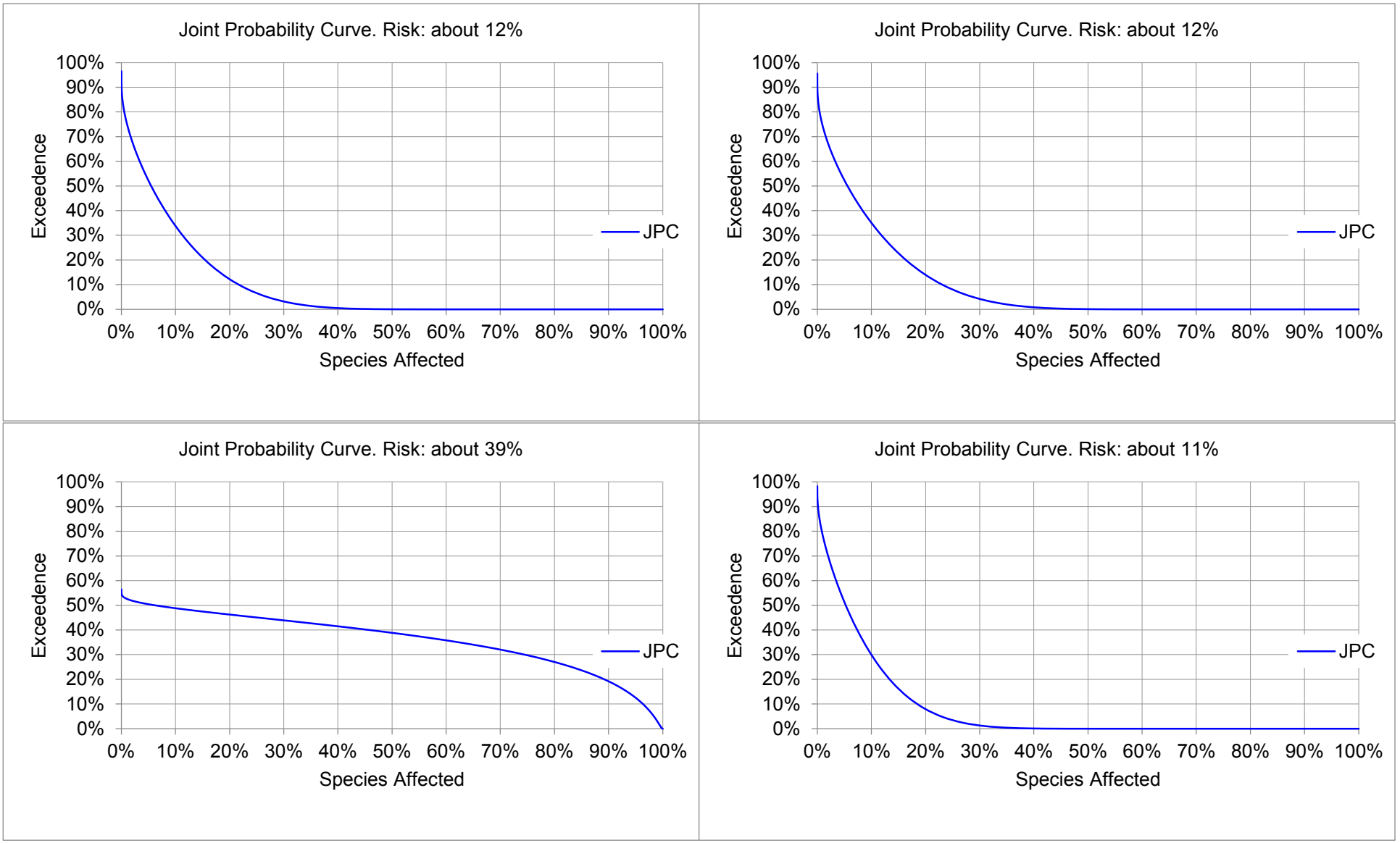


Figure 45: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for vertebrates

5.3.2.5. PAF – Primary producers

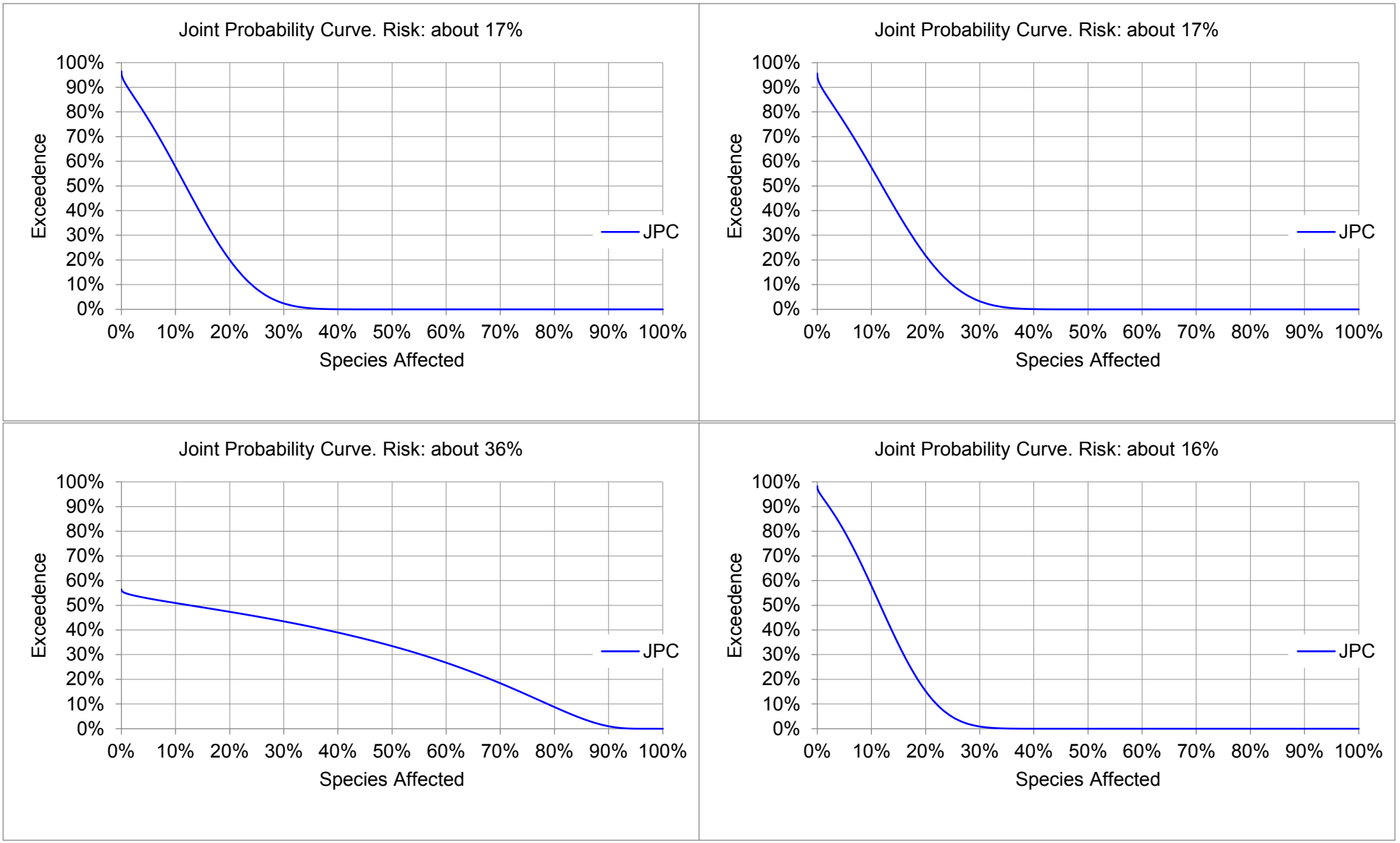


Figure 46: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for primary producers

5.3.2.6. PAF – Primary consumers

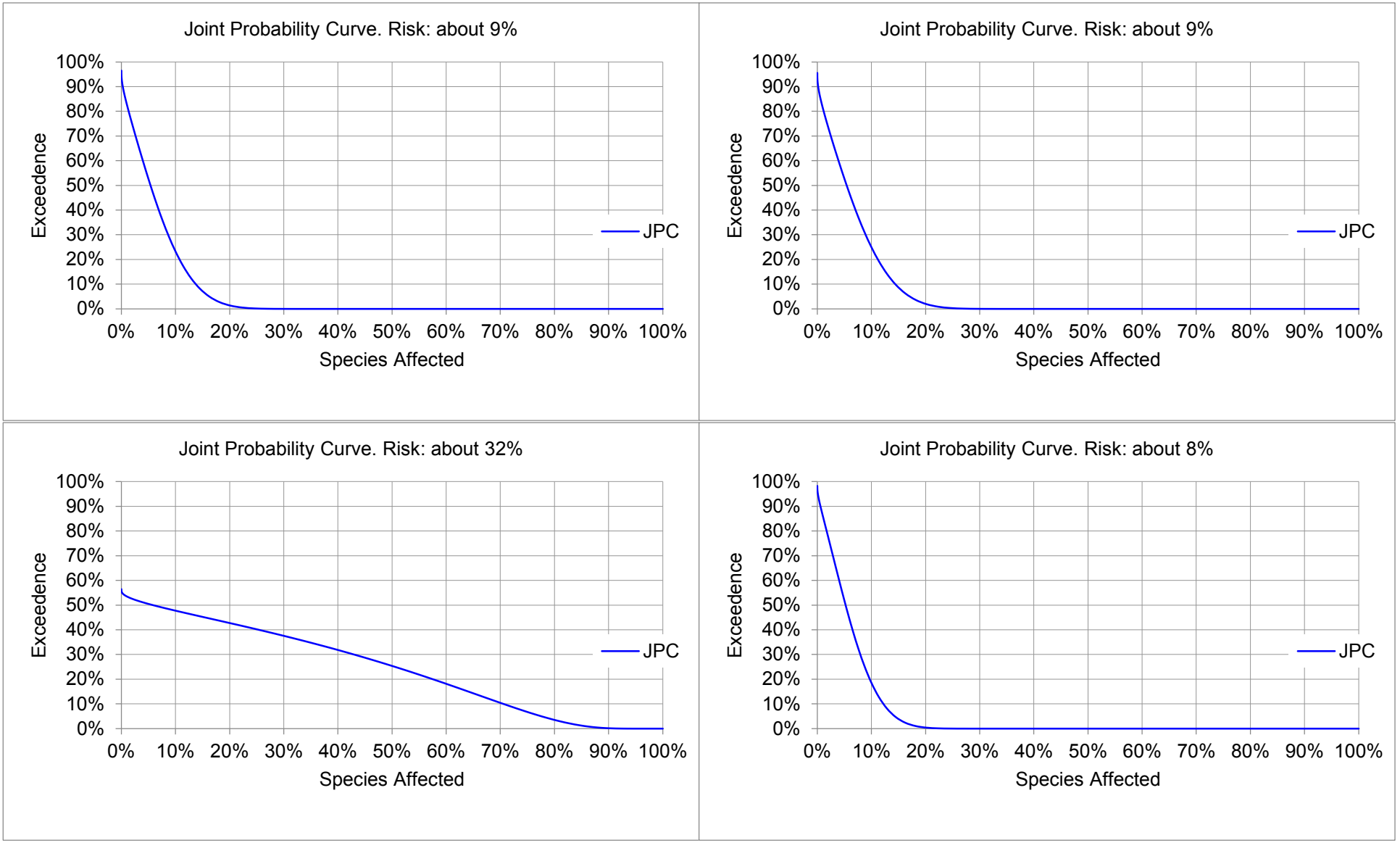


Figure 47: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for primary consumers

5.3.2.7. PAF – Secondary consumers

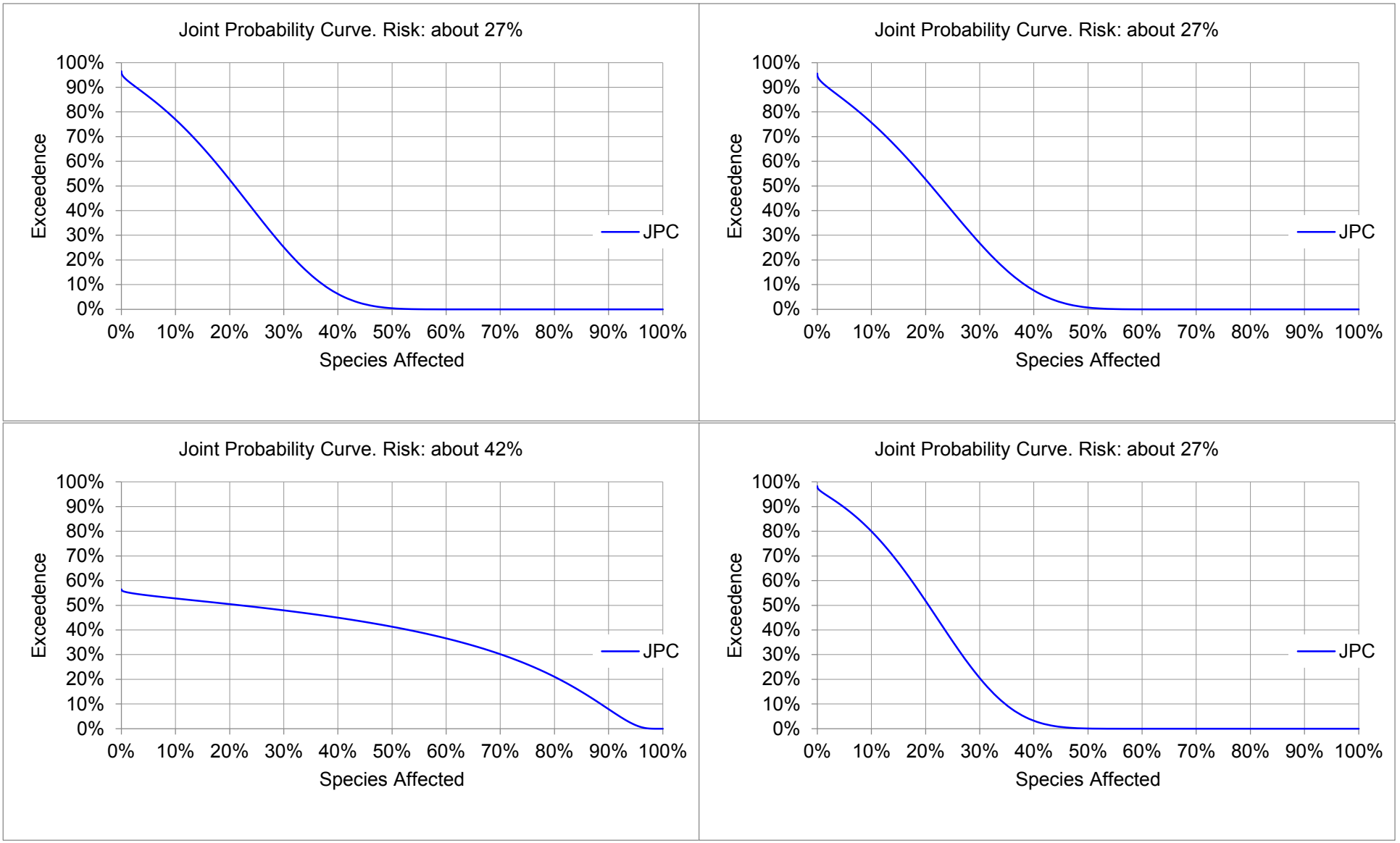


Figure 48: Joint Probability Curve (JPC) and PAF percentage for the four stations based on the SSWD for secondary consumers

6. Conclusions – Future considerations

The thesis presents the studies and research performed during the three years of my PhD programme. The activities were related with the AMORE research project and included three complex activities:

- the development of the MCDA-based WoE methodology and the underlying mathematical model,
- the supervision of the creation and development of the AMORE DSS as a standalone software application and
- the application of the DSS to the selected cyanide case study in France.

All the objectives have been fulfilled and a fully functional DSS has been developed, which utilises novel methodologies for performing probabilistic Ecological Risk Assessment.

The developed MCDA methodology, which successfully analyses the reliability and relevance of ecotoxicological data, is based on Multi-Attribute Value Theory (MAVT) and is combined with fuzzy logic and the use of causal relations, in the form of 'IF-THEN' rules, in a unique and innovative way. Thus, it allows the analysis of data based on an unambiguous set of hierarchically structured objective criteria and the quantitative assessment of data based on a Weight of Evidence framework. Based on the review of the existing frameworks for the analysis of ecotoxicological data (chapter 3), the proposed methodology is considered a step forward in the sector. The methodology addresses the identified flaws of the existing frameworks, through the use: (1) of unambiguous criteria, (2) of a quantitative assessment scoring system and (3) of clearly distinguished reliability and relevance concepts. In addition, the proposed methodology allows incorporating in the risk assessment the knowledge gathered from an expert panel and gives significant strength to the risk assessors for the performed assessments, through the use of previously not widely available information and expertise.

Even though the WoE framework has been designed specifically for the assessment of laboratory biotests for individual effects for the purposes of the AMORE project, the methodology is flexible and can be easily adapted to other types as well, e.g. laboratory biotests for population effects, multi species pseudo-field tests and modelling data.

This would require the creation of the respective hierarchically structured sets of criteria and the gathering of an expert panel for the evaluation of the importance of the new criteria hierarchies, with the use of the same type of questionnaire ('Questionnaire for expert consultation' – chapter 3.1.2). The latter is an assessment which can be performed in an easy way by any selected expert, through an online questionnaire. In this way, different probabilistic risk assessments could be performed with the use of the same aggregation methodology and the same software, in a very efficient way.

Regarding the DSS, it has been developed with the use of a rather simple programming language (VBA) and environment but it can run on any modern personal computer without the need for strong computing capabilities. Though, it is modular and scalable so that it can be improved, if necessary, in the future. This is possible not only by changing some of its elements (e.g. criteria hierarchy of the WoE methodology) but also by adding other modules/sub-modules.

The performed application of the AMORE DSS in the cyanide case study demonstrates a complete probabilistic Ecological Risk Assessment process with the use of Species Sensitivity Distributions and the utilisation of Multi-Criteria Decision Analysis.

Section C includes in detail the results of the application of the three modules of the software for the estimation of the Predicted Environmental Concentrations, the weighted Species Sensitivity Distributions and the Potentially Affected Fractions, with the aim of estimating the risk from the presence of cyanide in the Selune watershed in France.

The results, as analysed in Section C, provide to the Decision Maker/Risk Assessor a wealth of relevant information, which can be utilised for the different aspects of the Risk Assessment processes. The independent nature of the AMORE DSS modules, allows performing separately each process, therefore the user has the possibility of using the PEC or SDD modules stand-alone for a specific analysis or both the PEC and SSD modules for a complete ERA, based on the needs of his/her analysis. In our case study, both modules have been used since the available environmental exposure data and the ecotoxicological data are highly relevant to the analysed region and the desired risk assessment.

The statistical analysis of environmental exposure data has been used for predicting the concentrations of cyanide over time in the assessed area, based on the actual measurements at the stations. Furthermore, the analysis of the available ecotoxicological data for the toxicity of common European species, has allowed the production of relevant SSWD graphs that predict the sensitivity of species in the assessed area and the extraction of quality criteria. The SSWD results are available for all the data, as well as for the various taxonomic groups and trophic levels. The combination of the aforementioned results from modules 1 and 2, have allowed the estimation of the present risks for species in the watershed and the comparisons among taxonomic group and among trophic levels.

The case study results have shown that the MCDA based WoE methodology has a robust performance and allows the estimation of more reliable SSWDs, based on the innovative methodology for the analysis of ecotoxicological data. In addition, the conclusions drawn indicate that the methods proposed in this thesis have, in our opinion, potential for adoption within the risk assessment research fields. However, there are aspects that require further research in order to provide extra refined tools and methodologies and lead to more refined tiered risk assessments for pollutants in aquatic environments.

- Regularly during the research period, considerations and questions arose, regarding the tackled topics. Some of the most important ones are mentioned below, as they can be taken as future considerations: Group Decision processes and participatory processes require a high degree of common understanding, among all the involved parties. This is always a challenging task, both from the researcher's point of view, as well as from an expert and decision maker's point of view, since often different parties have different views and experiences on the assessed topic. For any similar research project, it is highly recommended, upon available funds restrictions, to plan as many live meetings, to ensure that all involved parties share the same visions and understanding and reach the maximum possible concensus.
- The selection of MCDA methods and the selection of the related aggregation operators is always a delicate topic. Even if the selected MCDA methodologies (MAVT, fuzzy logic) have been considered as appropriate for their incorporation in the

proposed MCDA-based scoring system for the assessment of ecotoxicological data, it is impossible to declare that some other possible methodology would be less or more efficient or appropriate for handling the topic.

- An open statistical issue in the SSD theory is the production of SSD graphs with small available datasets. Currently the DSS does not offer any specific functionality to handle this topic and therefore the exploration of how available Bayesian approaches could be combined with the proposed MCDA-based methodology for their possible addition as an option to the SSD module is considered highly interesting.

Future research could be performed both on the environmental side of the PhD thesis, as well on the mathematical one.

- The DSS is tailored for the assessment of ecotoxicological data from laboratory biotests with individual effects but it can be easily expanded to include other types of ecotoxicological data (Laboratory biotests - Population effects, Multi species pseudo-field tests, Modelling data). To this end, firstly the respective criteria hierarchies have to be developed and secondly experts should be gathered and asked to evaluate the criteria hierarchies through the 'Questionnaire for expert consultation'. The modularity of the DSS allows the fast and efficient incorporation of the new information. In addition, the DSS could be expanded to fit the needs of risk assessment for other environmental compartments (air, soil, sediments).
- Robust Ordinal Regression (Greco et al., 2008; Corrente et al., 2013b) is a MCDA method, which allows the induction of marginal value functions based on preference information collected from the Decision Maker(s). In our methodology, the criterion evaluation scores $\overline{\mathrm{CA}_1}$ are calculated through specific membership functions. It would be rather interesting to conduct research on the applicability of Robust Ordinal Regression for the induction of the proposed MCDA methodology membership functions.
- The proposed MCDA methodology utilises the concept of priority ordered sets of causal relations (IF-THEN rules), hierarchically structured based on experts' preferences. An alternative could be the 'Decision Rule approach' of the Dominance-based Rough Set Approach (DRSA) (Greco et al., 1998, 1999), which allows the extraction of decision rules based on the preferential information contained in the decision table of the problem under analysis, therefore a comparison of the two approaches could be interesting.

References

1. Ågerstrand M., Küster A., Bachmann J., Breitholtz M., Ebert I., Rechenberg B., Rudén C., 2011. Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. Environmental Pollution, Vol. 159, Issue 10, pp. 2487-2492.
2. Aldenberg T., 2007. Busy 1.0b Reference Manual. Institute for Risk Assessment Studies – IRAS, University of Utrecht, Netherlands.
3. American Public Health Association (APHA), 1915. Standard methods for the examination of water and wastewater. Vol. 2. American Public Health Association.
4. Angilella S., Corrente S., Greco S., Slowinski R., 2013. Multiple criteria hierarchy process for the Choquet integral. In 7th International Conference on Evolutionary Multi-Criterion Optimization, pages 475-489, Shaffield, UK, March 19-22.
5. ATSDR (Agency for Toxic Substances and Disease Registry), 2006. Toxicological profile for cyanide. Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA.
6. Aulenbach B.D., 1968. Water – Our second most important natural resource. 9 B.C.L. Rev 535.
7. Baccarelli A., Pfeiffer R., Consonni D., Pesatori A.C., Bonzoni M., Patterson D.G., Bertazzi P.A., Landi M.T, 2005. *Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study.* Chemosphere 60 pp. 898–906.
8. Barnthouse L.W., and Suter G.W. II. (eds.), 1986. User's Manual for Ecological Risk Assessment. ORNL-6251. Oak Ridge National Laboratory. Oak Ridge, Tennessee.
9. Baroudy I., Simonovic S., 2006. A Decision Support System for integrated risk management. Water Resources Research Report, no 51. University of Western Ontario.
10. Bellman R., Giertz M., 1973. *On the Analytic Formalism of the Theory of Fuzzy Sets*. Information Sciences, 5, 149-156.
11. Bose U., Davey A., Olson D., 1997. Multi-attribute Utility Methods in Group Decision Making: Past Applications and Potential for Inclusion in GDSS. Omega, Int. J. Mgmt Sci. Vol. 25, No. 6, pp. 691-706.
12. Bradbury S.P., Feijtel T.C.J., Van Leeuwen C.J., 2004. Meeting the Scientific Needs of Ecological Risk Assessment in a Regulatory Context. Environmental Science & Technology 2004 38 (23), 463A-470A.
13. Brandes LJ, Hollander H den, Meent D van de, 1996. SimpleBox 2.0: a nested multimedia fate model for evaluating the environmental fate of chemicals.
14. Brandt-Rauf P.W., Fallon L.F. Jr, Tarantini T., 1988. Health hazards of fire fighters: exposure assessment. Br J Ind Med 45(9):606–612.
15. Brauer V.F., Below H., Kramer A., 2006. The role of thiocyanate in the etiology of goiter in an industrial metropolitan area. Eur J Endocrinol 154(2):229–235.
16. Breitholtz M., Rudén C., Ove Hansson S., Bengtsson B.E., 2006. Ten challenges for improved ecotoxicological testing in environmental risk assessment. Ecotoxicology and Environmental Safety, Volume 63, Issue 2, February 2006, Pages 324-335.
17. Breton R. L., Gilron G., Thompson R., Rodney S., Teed S., 2009. A New Quality Assurance System for the Evaluation of Ecotoxicity Studies Submitted Under the New Substances Notification Regulations in Canada. Integrated Environmental Assessment and Management, Vol. 5, Number 1, 127–137

18. Burstein F., Holsapple C.W. (Eds.), *Handbook on Decision Support Systems 1: Basic Themes*, Springer, Heidelberg, 2008, pp. 163–189.
19. Corrente S., Greco S., Slowinski R., 2012. Multiple criteria hierarchy process in robust ordinal regression. Decision Support Systems, 53:660–674.
20. Corrente S., Greco S., Slowinski R., 2013a. Multiple criteria hierarchy process with ELECTRE and PROMETHEE. Omega, 41:820–846.
21. Corrente S., Greco S., Kadzinski M., Słowinski R., 2013b. Robust Ordinal Regression in preference learning and ranking. Machine Learning, vol.2-3, no. 93, pp. 381–422.
22. Craciun M.V., Neagu D., Minzu V., Bumbaru S., 2006. Software tool for toxicity prediction of pesticides, candidate pesticides and their derivatives.
23. Critto A., Suter II G.W., 2009. Environmental Risk Assessment. Chapter on A. Marcomini et al. (eds.), Decision Support Systems for Risk-Based Management of Contaminated Sites.
24. Davey A., Olson D., 1998. *Multiple Criteria Decision Making Models in Group Decision Support.* Group Decision and Negotiation, 7, 55–75.
25. De Sanctis G., Gallupe B., 1985. *Group decision support systems: a new frontier*. Database, pp. 3-10, Winter.
26. Duboudin C., Ciffroy P., Magaud H., 2004. Effects of data manipulation and statistical methods on species sensitivity distribution. Environmental Toxicology and Chemistry Vol. 23, No. 2, pp. 489–499.
27. ECHA, 2008a. Guidance on information requirements and chemical safety assessment. Chapter R.7b: Endpoint specific guidance. Helsinki, Finland
28. ECHA, 2008b. Guidance on Information Requirements and Chemical Safety Assessment Chapter R. 10: Characterisation of Dose [Concentration] - Response for Environment. Helsinki, Finland
29. ECHA, 2010. Practical guide 2: How to report weight of evidence. Helsinki, Finland
30. ECOFRAM, 1999. ECOFRAM Aquatic and Terrestrial Final Draft Reports, USEPA.
31. EEA, 1998. Environmental Risk Assessment: Approaches, Experiences and Information Sources – Chapter 6: Ecological Risk Assessment. Environmental Issues Series No. 4.
32. EPA, 2005. Environmental quality criteria reference document for cockburn sound (2003 - 2004). Environmental Protection Authority Report 20, Australia.
33. European Commission, 1996. Technical guidance document in support of commission directive 93/67/EEC on risk assessment for new notified substance and commission regulation 5EC no 1488/94 on risk assessment for existing substance. Part II. Office for Official Publications of the European Communities, Luxembourg.
34. European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. OJ L 327, 22/12/2000, p. 1–73
35. European Commission, 2003. Technical guidance document in support of commission directive 93/67/EEC on risk assessment for new notified substances and commission regulation (EC) No 1488/94 on risk assessment for exiting substances. Luxembourg.
36. European Commission, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, Official Journal of the European Union, L396, pp. 1–849.
37. European Commission, 2008a. Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 on environmental quality standards in the field of water policy, amending and subsequently repealing Council Directives 82/176/EEC, 83/513/EEC,

84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council. OJ L 348, 24/12/2008, p. 84–97
38. European Commission, 2008b. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. OJ L 353.
39. European Commission, 2011. Common Implementation Strategy for the Water Framework Directive (2000/60/EC), Guidance Document No. 27: Technical Guidance for Deriving Environmental Quality Standards. Tech. Rep. 2011-055.
40. Ferson S., 2002. RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers. Lewis Publishers, Boca Raton, Florida
41. Figueira J., Greco S., Ehrgott M., (eds) 2005. Multiple Criteria Decision Analysis: state of the art surveys. Springer, Boston.
42. Finlay P., Marples C., 1992. *Strategic Group Decision Support Systems: A Guide for the Unwary*. Long Range Planning, Vol. 25, No. 3, pp. 98 to 107, 1992
43. FOCUS, 2001. FOCUS Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC. Report of the FOCUS Working Group on Surface Water Scenarios, EC Document Reference SANCO/4802/2001-rev.2. 245 pp.
44. Forbes VE., Callow P., 2002. Species sensitivity distributions revisited: A critical appraisal. Hum Ecol Risk Assess 8(3):473-492.
45. Giove S., Brancia A., Satterstrom F. K., Linkov I., 2009. Decision Support Systems and Environment: Role of MCDA. Chapter on A. Marcomini et al. (eds.), Decision Support Systems for Risk-Based Management of Contaminated Sites.
46. Greco S., Matarazzo B., Slowinski R., 1998. A new rough set approach to evaluation of bankruptcy risk". In Operational Tools in the Management of Financial Risk, edited by C. Zopounidis, pp. 121- 136. Dordrecht, Boston: Kluwer Academic Publishers.
47. Greco S., Matarazzo B., Slowinski R., 1999. The use of rough sets and fuzzy sets in MCDM. In Advances in Multiple Criteria Decision Making, edited by T. Gal, T. Hanne and T. Stewart, chapter 14, pp. 14.1-14.59. Dordrecht, Boston: Kluwer Academic Publishers.
48. Greco S., Mousseau V., Słowinski R., 2008. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European Journal of Operational Research, vol. 191, no. 2, pp. 416–436.
49. Hellawell J.M., 1986. Biological indicators of freshwater pollution and environmental management.
50. Hendley P., Giddings J., 1999. *ECOFRAM Aquatic Report*, ECOFRAM Aquatic Workgroup Co-chairs.
51. Hickey G.L., 2010. Ecotoxicological Risk Assessment: Developments in PNEC Estimation. Durham theses, Durham University.
52. Hobbs D. A., Warne M., Markich S.J., 2005. Evaluation of Criteria Used to Assess the Quality of Aquatic Toxicity Data. Integrated Environmental Assessment and Management, Vol. 1, N. 3, 174–180
53. Holsapple C.W., 2008. *DSS architecture and types*, Chapter in: F. Burstein, C.W. Holsapple (Eds.), *Handbook on Decision Support Systems 1: Basic Themes*, Springer, Heidelberg, 2008, pp. 163–189.
54. Hommen U, Poethke H J, Dulmer U, Ratte HT. 1993. Simulation models to predict ecological risks of toxins in freshwater systems. ICES Journal of marine Science 50, 337-347

55. Jelassi T., Kersten G., Zionts S., 1990. An introduction to Group Decision and Negotiation Support.
56. Keen P.G.W., Scott-Morton M.S., 1978. *Decision Support Systems: an Organizational Perspective*, Addison-Wesley, Reading.
57. Keeney R., Raiffa H., 1976. Decisions with multiple objectives: preferences and value trade−offs. Wiley, New York.
58. Kilgour M., Eden C., 2010. *Handbook of Group Decision and Negotiation*. Springer
59. Klement E.P., Slany W., 1994. *Fuzzy logic in artificial intelligence*. Christian Doppler Laboratory Technical Reports 67.
60. Klimisch H.J., Andreae M., Tillmann U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25, 1-5.
61. Koksalan M., Wallenius J., Zionts S., 2011. *Multiple Criteria Decision Making: From Early History to the 21st Century*. World Scientific.
62. Levy J.K., Hartmann J., Li K.W., An Y.B., Asgary A., 2007. *Multi-criteria decision support systems for flood hazard mitigation and emergency response in urban watersheds,* American Water Resources Assosiation, Vol. 43, Issue 2, P. 346-358.
63. Linkov I., Loney D., Cornier S., Satterstrom K.F., Bridges T., 2009. Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Science of the Total Environment, 407, 5199-5205.
64. Linkov I., Welle P., Loney D., Tkachuk A., Canis L., Kim J.B., Bridges T., 2011. Use of multicriteria decision analysis to support weight of evidence evaluation. Risk analysis, Vol. 31, N. 8, 1211-1225
65. Mamdani E.H., 1977. Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE transactions on computers, V. C-26, Issue 12, 1182-1191.
66. Marakas G.M., 1999. Decision support systems in the twenty-first century. Upper Saddle River, N. J., Prentice Hall. 1999
67. Marcomini A., Suter II G.W., Critto A., 2009. *Decision Support Systems for Risk-Based Management of Contaminated Sites,* Springer.
68. Meinzen-Dick R., Appasamy P.P., 2002. Urbanization and Intersectoral Competition for Water. In Finding the Source: the Linkages between Population and Water. Woodrow Wilson International Center for Scholars: Washington, DC; 27–51.
69. OECD, 1992. Guideline for testing of chemicals - Fish, acute toxicity test (Method 203).
70. OECD, 2002. Guideline for testing of chemicals - Draft revised Guideline 201: Freshwater alga and cyanobacteria, growth and inhibition test.
71. OECD, 2004. Guideline for testing of chemicals. Daphnia sp., acute immobilization test and reproduction Test Method 202.
72. Olness A., 1995. Water quality: prevention, identification and management of diffuse pollution. Journal of Environmental Quality 24.2: 383-383.
73. Olson J.R., Rueter H., 1987. Extracting expertise from experts: Methods for knowledge acquisition. Expert Systems, Vol. 4, No 3.
74. Oracle Corporation, 2008. Crystal Ball manual: Oracle Crystal Ball, Fusion addition Release 11.1.1.3 Statistical Guide.
75. Posthuma L., Traas T.P ., Suter II G.W., 2002. *Species Sensitivity Distribution in ecotoxicology*, CRC Press LLC, chapters 1, 4, 5, 15.
76. Power D.J., 2002. *Decision support systems: concepts and resources for managers*, Quorum Books.

77. Schneider K., Schwarz M., Burkholder I., Kopp-Schneider A., Edler L., Kinsner-Ovaskainen A., Hartung T., Hoffmann S., 2009. “ToxRTool”, a new tool to assess the reliability of toxicological data. Toxicology Letters 189, 138–144
78. Solomon K., Giesy J., Jones P., 2000. *Probabilistic risk assessment of agrochemicals in the environment*, Crop Protection, No. 19, pp. 649-655.
79. Solomon K., Sibley P., 2002. New concepts in ecological risk assessment: where do we go from here?. Marine Pollution Bulletin, Volume 44, Issue 4, April 2002, Pages 279-285, ISSN 0025-326X.
80. Steinmaus C, Miller M.D., Howd R., 2007. Impact of smoking and thiocyanate on perchlorate and thyroid hormone associations in the 2001-2002 National Health and Nutrition Examination Survey. Environ Health Perspect 115(9):1333–1338.
81. Suter II G.W., 2006. Ecological risk assessment. CRC press.
82. Traas T.P., Meent D., Posthuma L., Hamers T., Kater B. J., Zwart D. D., Aldenberg T., 2002. *The potentially affected fraction as a measure of ecological risk*. Chapter in: Posthuma L., Traas T.P., Suter II G.W., 2002. *Species Sensitivity Distribution in ecotoxicology*, CRC Press LLC.
83. Tsuge K., Kataoka M., Seto Y., 2000. Cyanide and thiocyanate levels in blood and saliva of healthy adult volunteers. J Health Sci 46(5):343–350.
84. US EPA, 1983. Water Quality Standards Regulation (40 CFR 131).
85. US EPA, 1984. Ambient water quality criteria for cyanide. Washington, DC, United States Environmental Protection Agency.
86. US EPA, 1991. Technical support document for water quality-based toxics control. US Environmental Protection Agency. 505/2-90- 001. Washington, DC.
87. US EPA, 1994. Water Quality Standards Handbook: Second Edition.
88. US EPA, 1998. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. Risk Assessment Forum, US Environmental Protection Agency, Washington, DC.
89. US EPA, 2002. AQUATOX - Modeling environmental fate and ecological effects in aquatic ecosystems, Washington, DC.
90. US EPA, 2010. Toxicological review of hydrogen cyanide and cyanide salts. In Support of Summary Information on the Integrated Risk Information System (IRIS). Washington, DC.
91. van Straalen N.M., Denneman C.A.J., 1989. Ecotoxicological Evaluation of Soil Quality Criteria. Ecotoxicology and environmental safety 18, 241-251.
92. van Vlaardingen P.L.A., Traas T.P., Wintersen A.M., Aldenberg T., 2004. ETX 2.0. A Program to Calculate Hazardous Concentrations and Fraction Affected, Based on Normally Distributed Toxicity Data. RIVM report 601501028/2004.
93. Verdonck F.A.M., Jaworska J., Janssen C.R., Vanrolleghem P.A., 2002. Probabilistic ecological risk assessment framework for chemical substances. In Integrated Assessment and Decision Support, Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society (IEMSS) (Vol. 1, pp. 144-149).
94. Vincke P., 1992. Multi-criteria decision aid, John Wiley and Sons, Chichester.
95. WFD-UKTAG (Water Framework Directive - United Kingdom Technical Advisory Group), 2012. Proposed EQS for Water Framework Directive Annex VIII substances: cyanide (free). Edinburgh, Scotland.
96. Warne M.St.J., Westbury A.M., Sunderam R.I.M., 1998. A compilation of data on the toxicity of chemicals to species in Australasia Part 1: Pesticides. Australasian Journal of Ecotoxicology 4:93–144

97. Weed, D.L., 2005. Weight of Evidence: A Review of Concept and Methods. Risk Analysis, Vol. 25, No. 6, pp. 1545-1557
98. Wheeler J.R., Grist E.P.M., Leung K.M.Y., Morritt D., Crane M., 2002. Species sensitivity distributions: data and model choice. Marine Pollution Bulletin 45 (2002) 192–202.
99. WHO (World Health Organization), 2001. Report on Integrated Risk Assessment. WHO/IPCS/IRA/01/12. World Health Organization, Geneva, Switzerland.
100. Zadeh L.A., 1965. Fuzzy sets. Information and Control, 8, 338-353.
101. Zambelli P., Lora C., Spinelli R., Tattoni C., Vitti A., Zatelli P., Ciolli M., 2012. *A GIS decision support system for regional forest management to assess biomass availability for renewable energy production,* Environmental modelling and software 38, 203-213.
102. Zhao J., Juliang J., Qizhong G., Yaqian Ch., Mengxiong L., Tinoco L., 2014. Forewarning model for water pollution risk based on Bayes theory. Environmental Science and Pollution Research 21.4: 3073-3081.

Annex A – Hierarchical criteria structure

Line of Evidence	Ecotox category	Criteria group	#	Criterion and associated question
1. Experimental Reliability	1.1. Quality assurance	1.1.1 Guideline	1	*1.1.1.1. Guideline criterion*
				Did the test strictly follow an international guideline and/or national guideline?
		1.1.2. GLP	2	*1.1.2.1. GPL criterion*
				Did the test strictly follow Good Laboratory Practices?
	1.2 Substance identification and monitoring	1.2.1. Substance identification	3	*1.2.1.1. Substance identity criterion*
				Was the test substance identified by unambiguous information?
			4	*1.2.1.2. Substance purity criterion*
				Was the purity of the substance given?
		1.2.2. Substance loss and monitoring	5	*1.2.2.1. Loss ability criterion*
				Was the potential evolution in the concentration of the test substance negligible during the test period (by e.g. degradation, volatilization, formation of by-products, sorption onto container surface)?
			6	*1.2.2.2. Concentration monitoring criterion*
				Was the concentration of the tested substance measured during the test period ?
			7	*1.2.2.3. Loss acceptability criterion*
				Was the concentration loss of the tested substance during the test period acceptable?
			8	*1.2.2.4. Speciation criterion*
				Does the chemical monitoring allow to discriminate the different chemical species when relevant (dissolved vs particulate forms, ionic vs neutral forms etc)?
			9	*1.2.2.5. By-products criterion*
				Was the additional ecotoxicity of by-products formed during the test period evaluated?
	1.3 Organisms	1.3.1. Organisms identification and physiology	10	*1.3.1.1. Species identity criterion*
				Was the tested species variety and/or strain and/or isolate given?
			11	*1.3.1.2. Organisms physiology criterion*
				Was the age/length/weight/gender of the tested organisms given and suitable for running the test?
		1.3.2 Culture design	12	*1.3.2.1. Acclimatation criterion*
				Is the acclimatation duration (i.e. contact of the organisms with the investigated media during a given time period before introduction of the toxicant) adapted?
			13	*1.3.2.2. Organism feeding criterion*
				Do you consider that feeding of organisms over the test period is well stated and adapted?
	1.4 Test design	1.4.1. Controls	14	*1.4.1.1. No toxicant control criterion*
				Were appropriate 'No toxicant controls' (i.e. test under the same conditions without the investigated substance) used?
			15	*1.4.1.2. No toxicant control results criterion*
				Do the results obtained for no toxicant controls meet the validity criteria of the test?
			16	*1.4.1.3. Reference substance criterion*
				For the tested organism and effect, does a dose-response relationship exist for a reference substance?
			17	*1.4.1.4. Reference substance results criterion*
				Do the results obtained for the reference substance meet the expected toxicity range?
		1.4.2. Physico-chemical conditions	18	*1.4.2.1. Temperature criterion*
				Was the test temperature suitable for tested organisms?
			19	*1.4.2.2. pH criterion*
				Was the test pH suitable for tested organisms?
			20	*1.4.2.3. Light criterion*
				Was the light intensity and photoperiod stated and suitable for tested organisms?
			21	*1.4.2.4. Dissolved oxygen criterion*
				Was Dissolved oxygen concentration suitable for tested organisms?
			22	*1.4.2.5. Other physic-chemical parameters criterion*
				Were the other physico-chemical parameters (e.g. organic matter, hardness) known as potentially influent vs toxicity suitable for tested organisms?
		1.4.3. Exposure conditions	23	*1.4.3.1. Exposure system criterion*
				Do you consider that the bias generated by the exposure system (e.g. static; semistatic; flow through) is negligible and can be ignored?
			24	*1.4.3.2. Exposure route criterion*
				Is the route of exposure reliable for the species?

Table 11: Hierarchical criteria structure of the assessment methodology based on LoEs, Categories, Criteria groups and Criteria-Questions

2. Statistical Reliability	2.1. Test design	2.1.1. Concentration design	25	*2.1.1.1. Concentration number criterion* Was the number of tested concentrations acceptable?
			26	*2.1.1.2. Concentration spacing criterion* Was the concentration spacing acceptable?
		2.1.2. Replicates	27	*2.1.2.1. Replicates criterion* Was the number of replicates per concentration in agreement with the requirement of the statistical method applied?
			28	*2.1.2.2. Individuals number criterion* Was the number of individuals per replicate in agreement with the requirement of the statistical method applied?
	2.2. Assumptions	2.2.1. Hypothesis-testing assumptions	29	*2.2.1.1. Precision criterion* In case of Hypothesis-testing, were Precision (e.g. MSD or PMSD) criteria explicitly given and acceptable?
			30	*2.2.1.2. Power criterion* In case of Hypothesis-testing, were Power (type II error rate) and type I error rate explicitly given and acceptable?
			31	*2.2.1.3. Variance homogeneity criterion* In case of Hypothesis-testing, was the homogeneity of variance between treatments explicitly verified and acceptable?
		2.2.2. Regression model assumption	32	*2.2.2.1. Regression model selection criterion* Was there some mechanistic support to the choice of regression model?
			33	*2.2.2.2. Regression model comparison criterion* Were several regression models tested and was the selected model the best-fitting?
			34	*2.2.2.3. Prior assumption criterion* In case of Bayesian NEC, was the prior justified and acceptable?
	2.3. Estimation quality	2.3.1. Summary statistics quality	35	*2.3.1.1. Confidence interval criterion* Were the confidence interval for summary statistics indicated and was the confidence interval acceptable?
			36	*2.3.1.2. Goodness of fit criterion* Was the fit acceptable, with a justification of goodness of fit?
			37	*2.3.1.3. Comparative robustness criterion* Was the robustness of the summary statistics tested by comparison with alternative assumptions or approaches?
3. Biological Relevance	3.1. Duration	3.1.1. Acute test relevance	38	*3.1.1.1. Acute duration criterion* In case of acute test, was the test duration acceptable?
			39	*3.1.1.2. Deterministic ACR criterion* In case of acute test, is there a robust deterministic Acute-to-Chronic Ratio able to convert Acute summary statistics to Chronic summary statistics?
			40	*3.1.1.3. Probabilistic ACR criterion* In case of acute test, is there a robust probabilistic Acute-to-Chronic Ratio able to convert Acute summary statistics to Chronic summary statistics?
		3.1.2. Chronic test relevance	41	*3.1.2.1. Life-cycle criterion* In case of chronic test, did the test cover an acceptable range of the species life-cycle?
	3.2 Test design	3.2.1. Physico-chemical conditions	42	*3.2.1.1. Temperature relevance criterion* Was temperature used for the lab test suitable for the natural system targeted by the risk assessment?
			43	*3.2.1.2. pH relevance criterion* Was pH used for the lab test suitable for the natural system targeted by the risk assessment?
			44	*3.2.1.3. Salinity relevance criterion* Was salinity used for the lab test suitable for the natural system targeted by the risk assessment?
			45	*3.2.1.4. Dissolved oxygen relevance criterion* Was dissolved oxygen used for the lab test suitable for the natural system targeted by the risk assessment?
			46	*3.2.1.5. Other physico-chemical parameters relevance criterion* Were all the other physico-chemical parameters used for the lab test suitable for the natural system targeted by the risk assessment?
		3.2.2. Doses relevance	47	*3.2.2.1. Background criterion* Were the tested concentrations at a 'realistic' environmental level compared to the measured or predicted natural geochemical background concentration?
			48	*3.2.2.2. Ambient concentration criterion* Were the tested concentrations at a 'realistic' environmental level compared to the measured or predicted ambient concentration upstream of the investigated site?
	3.3. Biological endpoint	3.3.1. Summary statistics	49	*3.3.1.1. Hypothesis-testing relevance* If the summary statistics is derived from hypothesis-testing, it is considered that the calculated summary statistics are at least as robust as those derived from an alternative approach?
			50	*3.3.1.2. Effect bound relevance* Was the summary statistics a lower bound of effect on population?
		3.3.2. Sensitivity	51	*3.3.2.1. Most sensitive endpoint criterion* If different endpoints were measured for the investigated biotest, is the reported endpoint the most sensitive one?
			52	*3.3.2.2. Mode of Action criterion* If information related to Mode of Action is available, is there any justification that the reported endpoint is among the most sensitive ones?
		3.3.3. Population relevance	53	*3.3.3.1. Population dynamics criterion* Can the measured effect be considered to be an adverse effect directly affecting population dynamics?
			54	*3.3.3.2. Behavioral effect criterion* Can the measured effect be considered to be a behavioral effect directly affecting population dynamics?
			55	*3.3.3.3. Molecular/Cellular effect criterion* Did the biotest measure effect at molecular/cellular scale directly affecting population dynamics?
	3.4. Organism relevance	3.4.1. Ecological traits	56	*3.4.1.1. Habitat criterion* For the tested lab species, are the ecological traits related to habitat suitable for the natural system targeted by the risk assessment?
		3.4.2. Biological traits	57	*3.4.2.1. Food criterion* For the tested lab species, are the biological traits related to food preferences and feeding habits suitable for the natural system targeted by the risk assessment?

Table 11(continued): Hierarchical criteria structure of the assessment methodology based on LoEs, Categories, Criteria groups and Criteria-Questions

Annex B – Abstract

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Panagiotis Isigonis _____ matricola: 955934 _____

Dottorato: Scienze Ambientali _____

Ciclo: 27th _____

Titolo della tesi[1] : A Decision Support System for Probabilistic Ecological Risk Assessment (PERA) of pollutants on aquatic ecosystems _____

Abstract:

The PhD thesis is related with the development of a fully functional, modular Decision Support System (DSS) for performing probabilistic Ecological Risk Assessment (ERA) of pollutants in aquatic environments. The Decision Support System is a 3-module software, which integrates the use of Multi-Criteria Decision Analysis (MCDA) methods for the quantitative assessment of the reliability and relevance of ecotoxicological data. Ecotoxicological data are vital components of the Ecological Risk Assessment processes and specifically the Effect assessment part. An innovative MCDA based methodology has been fully developed (i.e. from the definition of the conceptual framework to the software implementation in collaboration with an experienced programmer) for the assessment of ecotoxicological data, which are used for the creation of weighted Species Sensitivity Distribution (SSWD) graphs. The MCDA based methodology allows the assessment of ecotoxicological data based on a set of 57 distinctive criteria and gives the possibility to researchers to classify and rank ecotoxicological data, based on their various characteristics. The knowledge base of the MCDA based methodology, which is required for the DSS, has been implemented with the support of scientists and experts on ecotoxicology.

A case study application has been performed for the analysis of the ecological risk from the presence of cyanide in the Sélune watershed, at the Manche region of the Lower Normandy in the north-west part of France. Environmental exposure data of cyanide (CN) have been collected from the Water Agency of 'Seine-Normandie' and used in the Exposure Assessment module, while ecotoxicological data for cyanide gathered from peer-reviewed publications have been analysed with the use of the proposed MCDA based methodology, in the Effect Assessment module. The ecological risk assessment process was concluded with the calculation of the risk indices in the last module of the DSS.

Firma dello studente

Panagiotis Isigonis

[1] Il titolo deve essere quello definitivo, uguale a quello che risulta stampato sulla copertina dell'elaborato consegnato.

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Panagiotis Isigonis _____ matricola: 955934 _____

Dottorato: Scienze Ambientali _____

Ciclo: 27th _____

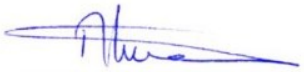
Titolo della tesi[2]: A Decision Support System for Probabilistic Ecological Risk Assessment (PERA) of pollutants on aquatic ecosystems _____

Abstract:

La tesi di dottorato proposta presenta lo sviluppo di un sistema di supporto alle decisioni modulare, completamente funzionante, volto alla valutazione probabilistica del rischio ecologico derivante da agenti inquinanti nei sistemi acquatici. Il sistema di supporto alle decisioni (DSS) è un software composto di tre moduli, che integra l'utilizzo di diverse tipologie di analisi multicriteriale (MCDA) per la valutazione quantitativa dell'attendibilità e rilevanza dei dati ecotossicologici. Tali dati ecotossicologici rappresentano componenti vitali della valutazione del rischio ecologico soprattutto relativamente alla fase di valutazione degli effetti. Una metodologia di analisi multicriteriale innovativa per la valutazione dei dati ecotossicologici utilizzati per la costruzione di weighted species sensitivity distributions (SSWD), è stata sviluppata in tutte le sue parti (i.e. dalla definizione del framework concettuale fino all'implementazione del software). Il processo MCDA di valutazione dei dati ecotossicologici proposto si basa sulla valutazione di 57 criteri distintivi e consente di ottenere la classificazione e l'ordinamento dei dati valutati sulla base delle loro caratteristiche. La knowledge base necessaria al corretto funzionamento dell'applicazione è stata collezionata grazie alla collaborazione di ricercatori ed esperti tossicologi.

L'analisi del rischio ecologico proposta è stata applicata ad un caso di studio relativo alla presenza di Cianuro nel fiume Sélune nella regione della bassa Normandia nella parte nord-ovest della Francia. I dati di esposizione del Cianuro utilizzati nel modulo di esposizione sono stati presi dal sito dell'agenzia per l'acqua 'Seine-Normandie', mentre i dati ecotossicologici, ricavati da pubblicazioni scientifiche, sono stati valutati attraverso la metodologia MCDA proposta nel modulo di valutazione dell'effetto. La valutazione del rischio ecologico è stata portata a termine attraverso il calcolo degli indici di rischio con il modulo di rischio del DSS.

Firma dello studente



Panagiotis Isigonis

[2] Il titolo deve essere quello definitivo, uguale a quello che risulta stampato sulla copertina dell'elaborato consegnato.