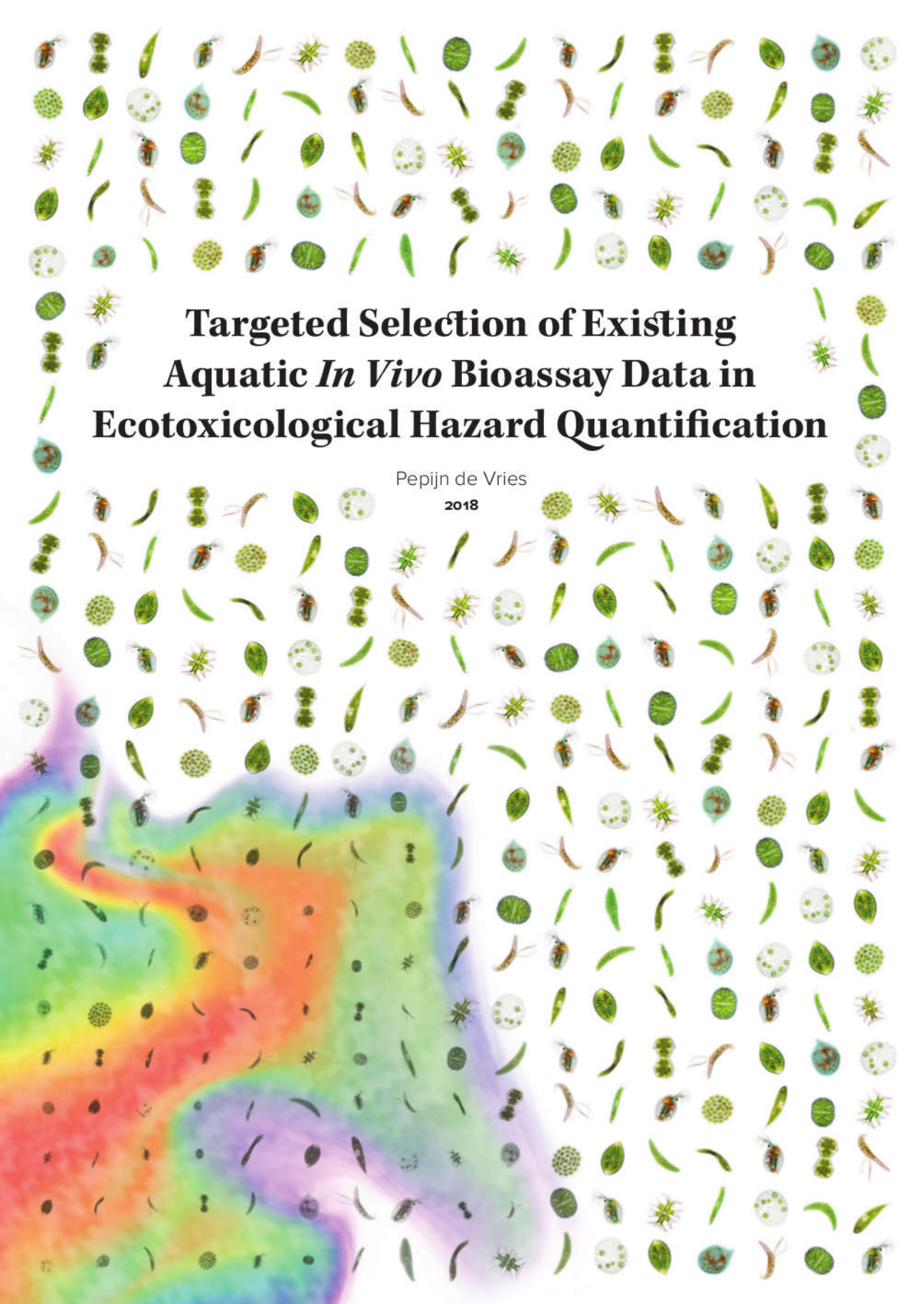
Targeted Selection of Existing Aquatic *In Vivo* Bioassay Data in Ecotoxicological Hazard Quantification

Pepijn de Vries

2018



Propositions

1. Guidelines inhibit critical thinking.
   (This thesis)
2. Although frowned upon by most (eco)toxicologists, No Observed Effect Concentrations (NOECs) should be used if they enable risk assessment based on existing bioassay data.
   (This thesis)
3. Sustainable harvesting only addresses the species of concern and should not be confused with sustainable fishing.
4. A theoretical model predicting physical properties (such as electrical and thermal conductance) is required to efficiently develop and apply new high entropy alloys.
5. The migration crisis is a blessing for the hosting society, under the condition that participation and assimilation is stimulated and secured.
6. Because of sheer bad luck Enkhuizen is currently not the capital of the Netherlands.

Propositions belonging to the thesis, entitled

"Targeted Selection of Existing Aquatic *In Vivo* Bioassay Data in Ecotoxicological Hazard Quantification".

Pepijn de Vries
Wageningen, 16 April 2018

Targeted Selection of Existing Aquatic *In Vivo* Bioassay Data in Ecotoxicological Hazard Quantification

Pepijn de Vries

Thesis committee

Promotor
Prof. Dr A.J. Murk
Professor of Marine Animal Ecology
Wageningen University & Research

Co-promotor
Dr E.M. Foekema, Researcher, Marine Animal Ecology
Wageningen University & Research

Other members
Prof. Dr P.J. van den Brink, Wageningen University & Research
Prof. Dr A.M.J. Ragas, Radboud University Nijmegen
Prof. Dr N.M. van Straalen, VU Amsterdam
Prof. Dr M.G. Vijver, Leiden University

This research was conducted under the auspices of the Research School for Socio-Economic and Natural Sciences of the Environment (SENSE).

Targeted Selection of Existing Aquatic *In Vivo* Bioassay Data in Ecotoxicological Hazard Quantification

Pepijn de Vries

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 16 April 2018
at 1.30 p.m. in the Aula.

Pepijn de Vries
Targeted Selection of Existing Aquatic *In Vivo* Bioassay Data in Ecotoxicological Hazard Quantification,
186 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in Dutch and English

ISBN 978-94-6343-730-1
DOI https://doi.org/10.18174/430498

Contents

1 General introduction 4
- 1.1 Ecotoxicological risk assessment . . . 4
- 1.2 Using existing bioassay data . . . 7
- 1.3 Research question . . . 9
- 1.4 Thesis outline . . . 9

2 Benford's Law applied to LC50 and NOEC data 15
- 2.1 Introduction . . . 16
- 2.2 Methods . . . 18
- 2.3 Results . . . 22
- 2.4 Discussion . . . 26
- 2.5 Conclusions . . . 30
- 2.A Appendix . . . 33

3 Towards risk assessment of elevated carbon dioxide levels 39
- 3.1 Introduction . . . 40
- 3.2 Methods . . . 41
- 3.3 Results . . . 45
- 3.4 Discussion . . . 48
- 3.5 Conclusions and recommendations . . . 55
- 3.A Appendix . . . 60

4 A species sensitivity distribution for temperature tolerance 71
- 4.1 Introduction . . . 72
- 4.2 Methods . . . 73
- 4.3 Results . . . 78
- 4.4 Discussion . . . 79
- 4.5 Conclusions . . . 85
- 4.A Appendix . . . 90

5 Flamingos at risk after oil fire incident 105
5.1 Introduction . . . 106
5.2 Method and materials . . . 108
5.3 Results . . . 114
5.4 Discussion . . . 123
5.5 Conclusions and recommendations . . . 127
5.A Appendix . . . 132

6 General discussion 140
6.1 Introduction . . . 140
6.2 Inherent quality of data . . . 144
6.3 Test species and conditions . . . 149
6.4 Relevance and context . . . 152
6.5 Conclusions . . . 153
6.6 Future perspectives . . . 154
6.A Appendix . . . 161

Glossary of acronyms and abbreviations 163

Summary 166

Samenvatting 170

Dankwoord 175

Curriculum vitae 177

List of publications 178
Peer reviewed articles . . . 178
Poster and platform presentations . . . 179
Selection of additional publications . . . 181

SENSE Certificate 184

Chapter 1

General introduction

1.1 Ecotoxicological risk assessment

Risk is defined as 'a situation involving exposure to danger' according to the Oxford online dictionary[1]. Ecotoxicological risk thus deals with the adverse effects ('danger') experienced by organisms in an ecosystem caused by exposure to one or more toxicants.

'Risk assessment' was initially addressed by the US National Research Council and focused on human health (NRC, 1983). After several iterations, it was eventually formally expressed by the EU as "*a process of evaluation, including the identification of the attendant uncertainties, of the likelihood and severity of an adverse effect(s)/event(s) occurring to man or the environment following exposure under defined conditions to a risk source(s)*" (EC, 2000b).

Risk assessment as such is widely applied to predict, monitor, evaluate and manage the (ecotoxicological) quality of environmental compartments. Although this thesis will mainly focus on the water column as such a compartment, many of the issues presented and discussed here are also applicable to other environmental compartments.

1.1.1 Applications in policy

Risk assessment is applied within many European policy frameworks. For instance, the Water Framework Directive aims to reach and ensure a good environmental status (including water quality) in river basins (EC, 2000a). Its counterpart, the Marine Strategy Framework Directive is meant to achieve the same goal in the marine environment (EC, 2008). As many regional seas extend beyond the EU, each European regional sea has its own convention, such as OSPAR

[1] `https://en.oxforddictionaries.com`, accessed on 2 September 2017

(`www.ospar.org`) for the North-East Atlantic. Risk assessment is applied in several OSPAR agreements. One example is the 'risk based approach' which aims to manage the environmental risk posed by produced water discharges from offshore oil and gas platforms (OSPAR, 2012). The EU directive concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) also calls for risk assessment in order to ensure environmental (and human) health, by focusing on substances that are produced or imported in bulk (EC, 2006). These are only a selection of European policy frameworks relying on risk assessment.

1.1.2 The risk management cycle

Despite the wide diversity of European policies and applications, the risk assessment process in each of them is always conducted following a similar set of rules (NRC, 1983; EC, 2000b; ECHA, 2003; Maltby, 2006; ECHA, 2008, 2011b). In the EU risk assessment was originally guided by the Technical Guidance Document which was developed for pesticides (ECHA, 2003). This was followed by the more recent guidelines by the European Chemicals Agency (ECHA) (e.g., ECHA (2008, 2011b)) which were developed for stakeholders of the REACH policy for a wide range of chemicals (EC, 2006). In general, ecotoxicological risk assessment follows the principles of the risk management cycle as depicted in Fig. 1.1. Risk assessment should always start with the problem formulation (Fig. 1.1), as this will dictate the methodology that should be followed and the type of data that is required.

Before risk of a substance posed to the environment can be assessed, the hazard of the substance needs to be determined. The hazard is regarded as the potency of a substance to cause adverse effects (EC, 2000b). This hazard is usually assessed in *in vivo* bioassays, where an effect on a test species is evaluated experimentally at different doses. The results of such experiments are often summarised in a single (no) effect concentration, to indicate the sensitivity of a species for the tested substance. For example, the 50% Effect Concentration (EC50) is frequently used. This is the interpolated concentration at which the effect is 50% between a baseline and maximum effect after a specific exposure duration. The No Observed Effect Concentration (NOEC) is the highest test concentration of a substance at which no significant (adverse) effects are observed within a specified exposure time. NOECs are criticised for multiple reasons, among which the fact that they depend on selected test concentrations and the statistical power resulting from the experimental setup (Jager, 2012). Despite these critiques the NOEC is still regularly used in environmental risk assessment.

These (no) effect concentrations can be determined for both acute (covering a short part of a species life-span) and chronic (covering a large(r) part of the

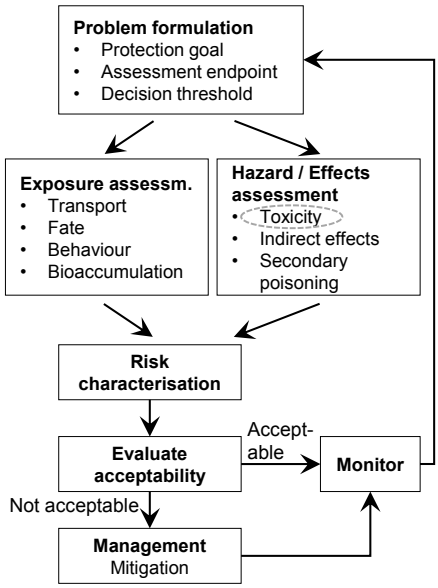


Fig. 1.1: The risk assessment process and management cycle, modified after Maltby (2006), reproduced by permission of The Royal Society of Chemistry. Aquatic *in vivo* bioassays (toxicity tests) are the focus in this theses and is therefore circled in this diagram.

species life-span). It is this type of aquatic *in vivo* bioassay data that this thesis will focus on. Following the problem formulation, this information is used in the step called hazard quantification, also often referred to as effect assessment (Fig. 1.1). Note that several other kinds of effect data are also available. For instance the No Effect Concentration, that interpolates the no effect level based on the effect of internal exposure levels on survival rate processes (Kooijman et al., 1996) as for instance implemented in the DEBtox model (Jager and Zimmer, 2012). However, this type of effect data is still underused in risk assessment, partly due to its limited availability.

The risk to the environment is of course not only determined by the potency of a substance to cause harm (the hazard). The actual or expected exposure concentration in the field is also important (Fig. 1.1). The risk can then be characterised using comparable information on the hazard and the exposure level (Fig. 1.1).

Once risk is characterised it needs to be evaluated and communicated whether the risk is acceptable or not (Fig. 1.1). Usually the rules for the acceptability of risk are set in the initial problem definition and goals. If risk is considered to be acceptable, it can be decided to keep monitoring the (the effects

occuring at the given) risk level. When the risk is assessed as being unacceptable, mitigating measures should be taken. The risk assessment should then be re-evaluated and new measures taken and developments monitored, in order to determine whether the adaptions were effective (Fig. 1.1).

Risk assessment often occurs in a tiered approach (Aagaard et al., 2013). Initially risk can be screened using the precautionary principle ('better safe than sorry') (Harremoës et al., 2001), for instance by applying large safety factors on available (no) effect concentrations (Craig, 2006). These factors should cover uncertainties when extrapolating from one species to another, from acute to chronic exposure and/or from laboratory to field conditions. By being very strict, this principle should ensure that situations classified as 'safe', truly are safe. However, if in such an initial evaluation the risk cannot be classified as 'safe', this does not automatically mean that it is *un*safe. In such cases higher tier risk assessment should be applied reducing the uncertainties thus the need for large extrapolation factors. Higher tier risk assessment moves away from the rough precautionary principles to a more realistic evaluation; from a generic to a more specific assessment. Higher tier risk assessment aims to achieve higher accuracy and certainty. Therefore in higher tier risk assessment the data requirements are higher (more and better quality data).

1.2 Using existing bioassay data

Usually, when a risk assessment process is started, existing bioassay data are initially applied for the hazard quantification, rather than performing experiments to generate new data (ECHA, 2011a). This is facilitated by extensive ecotoxicological databases such as US EPA ECOTOX (`www.epa.gov/ecotox`). There are several reasons for using existing data. Firstly, it is unethical to needlessly sacrifice additional experimental animals, which includes animals that are generally not covered by legislation (such as invertebrates). Secondly, it is practical, as generating new data costs time, means and expertise and probably often considered even more importantly: money. By reusing or recycling existing data, sustainable use of means and resources is promoted.

The premise of this thesis is therefore how existing bioassay data can be used effectively in risk assessment for slightly different situations. As illustrated before (Fig. 1.1) the objective for the risk assessment needs to be established first, which subsequently should dictate (hazard) data and risk assessment model requirements. Next available data needs to be identified, for which the databases (such as ECOTOX) can be used as a starting point. Once data have been identified, they need to be evaluated against specific quality criteria, for which guidelines can be used (ECHA, 2011b). Using the established objectives and quality of the data, a final selection of data can be made on which the risk assessment will be based. These steps will be further discussed below.

Data evaluation and selection procedures (i.e., the selection criteria) usually focus on the quality of the data and its adequacy for the intended purpose (ECHA, 2011b). ECHA (2011b) distinguishes between the relevance and reliability as data quality aspects. Statistical uncertainty in the data is not directly considered in the selection process by ECHA (2011b), but is instead addressed in the evaluation of the risk characterisation (ECHA, 2012), e.g. by applying extra safety factors (Craig, 2006; Craig et al., 2012) or a probabilistic approach (e.g., Aldenberg and Jaworska (2000)). Relevance of the data should reflect that representative substances, species, exposure routes and doses have been tested and all parameters affecting the toxicity end-points are considered (ECHA, 2011b). Traditionally, reliability of bioassay data is scored using the Klimisch et al. (1997) approach. This approach considers the level of standardisation that was applied in a test and how well test results are documented. Although uncertainty in test results is not directly considered, it is an important aspect of the data quality and should be considered (at the very least in a next iteration of the risk assessment cycle, Fig. 1.1).

It makes sense to only use bioassay data of sufficient quality by applying selection criteria. However, the guidelines (ECHA, 2011b) don't explicitly mention why the evaluation, and selection based thereon, focuses on the suggested criteria. It can only be assumed that this is to improve the accuracy, precision (Fig. 1.2) and/or reliability of the risk assessment.

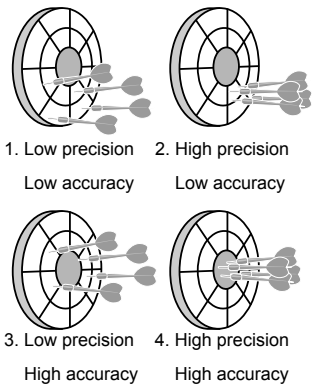


Fig. 1.2: Illustration of accuracy versus precision. Imagine the target environment to be the bullseye and each dart to represent a bioassay test result. When the tests poorly represent the target environment, this will result in low accuracy (top panels). When there is a lot of variation between the test results or uncertainty in each test results, this will result in low precision (lefthand panels). Note that the number of available test results (darts) can also affect the certainty and accuracy.

By selecting reliable data, the risk assessment is expected to become more credible, trustworthy and possibly more precise. This will improve the credi-

bility of risk assessment and will help with its communication to stakeholders. However, relying on less data can also increase the uncertainty (Aldenberg and Jaworska, 2000) and furthermore can introduce a bias (if it wasn't already there) in the risk assessment (Fox, 2015; Buonsante et al., 2014; Craig et al., 2012), making it less accurate.

Selecting relevant data, i.e., data that best represent the target ecosystem and the concerning exposure routes, will improve the accuracy of risk assessment (Fig. 1.2). Nonetheless laboratory tests need to be translated to field situations in risk assessment. By omitting less relevant data from risk assessment, the assessment may become more accurate (leaving a smaller gap between laboratory and field conditions) but it also leaves less data, making the assessment less precise (Aldenberg and Jaworska, 2000).

Risk assessment is thus clearly affected by the availability and data selection procedure, when relying on existing data only. How data selection affects the quality of risk assessment (accuracy, precision and credibility) is usually not made explicit and it is not evaluated whether the selection had the desired effect on the quality of the assessment. As explained above, a selection procedure can even worsen one or more of the quality aspects. Strictly following guidance rules without critical evaluation of its aim and consequences, is thus no guarantee for optimal risk assessment.

1.3 Research question

Under which conditions can *in vivo* bioassay data, that are discarded under current guidelines, still be used in hazard quantification and risk assessment? That is the main question this thesis will focus on. And vice versa: are there any data that are accepted under current guidelines, that should not be used in hazard quantification and risk assessment? Or in other words: how can the use of available *in vivo* bioassay data for hazard quantification in risk assessment be optimised.

1.4 Thesis outline

Chapter 2 starts with the core and addresses a key aspect of bioassay data quality: its reliability. Trustworthiness is an aspect of reliability that is generally ignored and is believed to be eliminated by focusing on other data quality aspects. Current approaches assume that selected data is free of (deliberate) errors. This chapter examines an approach for evaluating trustworthiness of toxicity data, by applying Benford's Law.

Chapters 3 and **4** revolve around bioassay data collection and data selection for stressors that are hard to quantify unambiguously. The approaches for such

situations are explored for bioassays of non-toxic stressors. These tests cannot rely on the level of standardisation that is available for testing of toxicants and no elaborate, standardised databases are available for them. Therefore, the data have to be collected from literature. **Chapter 3** concentrates on effects of elevated carbon dioxide levels (CO_2), a research field that is relatively young and surrounded with seemingly contradicting outcomes. The need for more studies about CO_2 effects was initiated in the early 2000s, triggered by increasing CO_2 levels in the atmosphere and the Kyoto protocol. This chapter considers how lack of standardisation affects the data selection process and the consequent hazard quantification and risk assessment. **Chapter 4** addresses temperature induced mortality effects. These effects have been studied since the 1950s, are therefore somewhat more standardised, and enjoys a larger data availability. Temperature increase, also has a relative and time-component to its quantification. The suggested approach allows for evaluation of the ecological risk of thermal discharges.

Chapter 5 is a more complex case study applying ecotoxicological risk assessment in practise. It studied how choices made in the selection and usage of bioassay data will depend on the context of data use. In this chapter existing water and sediment quality standards (based on existing bioassay test results) are used to evaluate the effects on food availability and of secondary poisoning of flamingo birds in a real-life field situation. This case study evaluates the relative importance of aspects, other than the utmost quality of hazard quantification, on the outcome of the risk assessment by addressing the full risk assessment process (Fig. 1.1).

Chapter 6 reflects on the results of the case studies presented in the previous chapters in combination with information from literature. Based on this information it will be evaluated how data selection criteria affect hazard quantification and thus risk assessment. It will be explored whether these selection criteria can be refined and guidelines can become more effective. Future perspectives on the application of existing bioassay data in ecological risk assessment will also be presented.

References

Aagaard, A., Brock, T., Capri, E., Duquesne, S., Filipic, M., Hernandez-Jerez, A. F., Hirsch-Ernst, K. I., Hougaard Bennekou, S., Klein, M., Kuhl, T., Laskowski, R., Liess, M., Mantovani, A., Ockleford, C., Ossendorp, B., Pickford, D., Smith, R., Sousa, P., Sundh, I., Tiktak, A., Van Der Linden, T., 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 11 (7), 1–268.

Aldenberg, T., Jaworska, J. S., 2000. Uncertainty of the hazardous concentration

and fraction affected for normal species sensitivity distributions. Ecotox Environ Safe 46 (1), 1–18.

Buonsante, V. A., Muilerman, H., Santos, T., Robinson, C., Tweedale, A. C., 2014. Risk assessment's insensitive toxicity testing may cause it to fail. Environ Res 135, 139–147.

Craig, P., 2006. Uncertainty factors in ectotoxicological risk: uncertainty factors in ectotoxicological risk management. 4th Edinburgh risk conference, Edinburgh.

Craig, P. S., Hickey, G. L., Luttik, R., Hart, A., 2012. Species non-exchangeability in probabilistic ecotoxicological risk assessment. J R Stat Soc A Stat 175 (1), 243–262.

EC, 2000a. 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. Off J Eur Comm L327, 1–72.

EC, 2000b. First report on the harmonisation of risk assessment procedures. part 1: The report of the scientific steering committee's working group on harmonisation of risk assessment procedures in the scientific committees advising the European Commission in the area of human and environmental health 26-27 October 2000.

EC, 2006. Corrigendum to 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, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union L136, 3–280.

EC, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (Text with EEA relevance). Off J Eur Comm L164, 19–40.

ECHA, 2003. Technical guidance document on risk assessment in support of Commission Directive 93/67/EEC on risk assessment for new notified substances, Commission regulation (EC) No 1488/94 on risk assessment for existing substances and of Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Tech. Rep. EUR 20418 EN/2, European Chemicals Agency, Ispra, Italy.

ECHA, 2008. Guidance on information requirements and chemical safety assessment chapter R.10: Characterisation of dose [concentration]-response for environment.

ECHA, 2011a. Guidance on information requirements and chemical safety assessment chapter R.3: Information gathering. version 1.1.

ECHA, 2011b. Guidance on information requirements and chemical safety assessment chapter R.4: Evaluation of available information. version 1.1.

ECHA, 2012. Guidance on information requirements and chemical safety assessment chapter R.19: Uncertainty analysis. version 2.1.

Fox, D. R., 2015. Selection bias correction for species sensitivity distribution modeling and hazardous concentration estimation. Environ Toxicol Chem 34 (11).

Harremoës, P., Gee, D., MacGarvin, M., Stirling, A., Keys, J., Wynne, B., Guedes Vaz, S., 2001. Late lessons from early warnings: the precautionary principle 1896-2000. Tech. Rep. 22, European Environment Agency, Copenhagen.

Jager, T., 2012. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. Environ Toxicol Chem 31 (2), 228–229.

Jager, T., Zimmer, E. I., 2012. Simplified dynamic energy budget model for analysing ecotoxicity data. Ecological Modelling 225 (Supplement C), 74–81.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharm 25 (1), 1–5.

Kooijman, S. A. L. M., Bedaux, J. J. M., Slob, W., 1996. No-effect concentration as a basis for ecological risk assessment. Risk Anal 16 (4), 445–447.

Maltby, L., 2006. Environmental risk assessment. In: Harrison, R. M., Hester, R. E. (Eds.), Chemicals in the Environment: Assessing and Managing Risk. Royal Society of Chemistry, Cambridge, UK, Ch. 5, pp. 84–101.

NRC, 1983. Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, DC.

OSPAR, 2012. OSPAR recommendation 2012/5 for a risk-based approach to the management of produced water discharges from offshore installations.

Photo: Pepijn de Vries

Chapter 2

Compliance of LC50 and NOEC Data with Benford's Law: an Indication of Reliability?

Reprinted from Ecotoxicology and Environmental Safety 98, Pepijn de Vries and Albertinka J. Murk (pages 171–178, copyright (2013)), with permission from Elsevier

Abstract

Reliability of research data is essential, especially when potentially far-reaching conclusions will be based on them. This is also, amongst others, the case for ecotoxicological data used in risk assessment. Currently, several approaches are available to classify the reliability of ecotoxicological data. The process of classification, such as using the Klimisch score, is time-consuming and focuses on the application of standardised protocols and the documentation of the study. The presence of irregularities and the integrity of the performed work, however, are not addressed. The present study shows that Benford's Law, based on the occurrence of first digits following a logarithmic scale, can be applied to ecotoxicity test data for identifying irregularities. This approach is already successfully applied in accounting. Benford's Law can be used as reliability indicator, in addition to existing reliability classifications. The law can be used to efficiently trace irregularities in large data sets of interpolated (no) effect concentrations such

as LC50s (possibly the result of data manipulation), without having to evaluate the source of each individual record. Application of the law to systems in which large amounts of toxicity data are registered (e.g., European Commission Regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals) can therefore be valuable.

2.1 Introduction

In ecotoxicological risk assessment of substances, the quality and hence the reliability of underpinning data are vital. Reliability of these data is usually assessed by applying a scoring system. Klimisch et al. (1997) proposed an approach that is used by risk assessors from regulatory agencies to classify the reliability of studies performed. Other approaches are also available, some have been evaluated by Ågerstrand et al. (2011). Such scoring methods usually assess whether laboratory experiments are well documented and conducted under standardised conditions. A problem with such classification methods is that they rely on the information provided and are time-consuming to perform. Such classifications cannot account for irregularities in the data, e.g., as result of (unintentional) errors made during the performance of the test, errors made while interpreting the test results, or even deliberate data manipulation.

The trustworthiness of data is also an issue in other fields using large data sets such as accounting. In that field, an approach has been developed based on the occurrence of first digits following a logarithmic scale, also called Benford's Law, Newcomb's Law or First Digit Law (Benford, 1938; Newcomb, 1881). It is successfully applied to identify suspicious book keeping (Rauch et al., 2011) or even fraud (Geyer and Williamson, 2004; Durtschi et al., 2004). In environmental science Benford's Law has been applied to identify irregularities in emission monitoring data (Dumas and Devine, 2000; Marchi and Hamilton, 2006) but to this date not to (eco)toxicicological data.

This study applies Benford's Law to ecotoxicological data (median lethal concentrations, LC50 and No Observed Effect Concentrations, NOEC) as a tool to quickly screen large amounts of data for anomalies, thereby dealing with an untouched aspect of quality.

2.1.1 Benford's Law

Benford's Law revolves around the first non-zero digit in numbers of a data set (e.g., digit '8' for the number 8.01, or '2' for the number 0.023). One might expect that each leading digit occurs with equal frequency (that is, the chance of finding the leading digit '1' is equal to that of finding digit '2', namely $\frac{1}{9} \approx 0.111$). However, Newcomb (1881) and later Benford (1938) (independently) observed that in many (but not all) data sets the leading non-zero digit '1' is more common

than '2', which in turn is more common than '3' and so on. Newcomb (1881) formulated this observation as follows:

$$Prob\,(D_1 = d_1) = \log_{10}\left(1 + \frac{1}{d_1}\right), \text{ for } d_1 = 1, \ldots, 9; \tag{2.1}$$

where the left-hand term indicates the probability that a first non-zero digit (D_1) equals a specific digit (d_1). So, according to Benford's Law, the fraction of leading digits equals '1' is $Prob\,(D_1 = 1) = \log_{10}(2) \approx 0.301$. The fraction that equals '2' is $Prob\,(D_1 = 2) = \log_{10}(1.5) \approx 0.176$, etc. There are numerous publications on the law. Fewster (2009) gives a simple introduction into the matter; more in-depth considerations are also available (e.g., Berger and Hill (2011)). For a nearly complete overview one could consult `http://www.benfordonline.net`.

Durtschi et al. (2004) and Fewster (2009) suggested a number of criteria for (accountant) data in order for them to comply with Benford's Law. Important criteria are: the data set is large; the data set spans several orders of magnitude; the mean of the set of numbers is greater than the median and the skewness is positive (which, for instance, is the case for log-normally distributed data). Furthermore, the data set is not comprised of assigned numbers; the numbers are not influenced by human choices; and the data set does not have a built in minimum and/or maximum. Log-normal and log-logistic distributions are commonly used to approximate the statistical distribution of ectoxicological data (e.g. Wheeler et al. (2002)). Note that data that are log-logistically distributed, have a mean that is greater than the median and have positive skewness. Analyses by De Zwart (2002) show that effect concentrations span orders of magnitude per substance (the variation will be larger when substances are combined). When proper test concentrations are selected, the interpolation of effect concentrations should not be affected by the concentrations selected for the experiment. Hence, it is expected that LC50 data, being the predominant endpoint, comply with Benford's Law. Other (no) effect concentrations interpolated from dose-response curves, such as the No Effect Concentration (Kooijman et al., 1996) and the Benchmark Dose (Crump, 1984), are also expected to conform to Benford's Law. No Observed Effect Concentration (NOEC) data on the other hand, are not expected to follow Benford's Law as they are based on the experimenter's choice of exposure concentrations.

In the present study the conformity to Benford's Law of both LC50 and NOEC data is tested. Subsets of LC50 data, mostly based on bibliometric meta-information associated with the data, are also tested.

2.2 Methods

2.2.1 Collecting toxicity data

The complete US EPA ECOTOX database (`ftp://ftp.epa.gov/pub/ecotox/ecotox_ascii_03_15_2012.exe`) is retrieved on 31st May 2012. All LC50 records are extracted from the database for analysis that report an effect concentration in μg/l, or decades of this unit; such as mg/l or g/l. Effect concentrations that are reported as 'greater than', 'less than' or 'approximately equal to' are omitted from analysis, as they indicate a minimum or maximum and thus do not comply with Benford's Law. In the same way, NOEC data are extracted from the database and subsequently further restricted to mortality endpoints, for comparability with the LC50 data. The routines for the restriction of the data and all those described below are implemented in R (version 2.12.2, The R Foundation for Statistical Computing, Vienna).

2.2.2 Classification of collected data

Benford's Law does not apply to individual records, rather it has to be applied to a complete data set or specific subsets thereof. Therefore, classification of the ecotoxicological data based on the meta-information is required. This classification is presented in Table 2.1 and is also used to study how characteristics of the studies that produced the toxicity data affect compliance of the data with Benford's Law. When meta-information is already categorical (for example whether the test substance is a pesticide or not), there is no need for classification. Numerical meta-information is divided into four classes with convenient intervals so that each class roughly contains the same amount of toxicity data (see Fig. 2.1 and Table 2.1 for the selected intervals).

As the same meta-information is not available or collected for all subsets of data, the relationship with some characteristics are only studied for specific sub-groups. The data are analysed in the following steps of increasing detail (see also Fig. 2.1 and Table 2.1): (A) the complete US EPA ECOTOX database; (B) LC50 records selected from the database; (C) LC50 records from the database for which its source is listed in Scopus (`http://www.scopus.com`).

In the first step (A), only the single matching characteristic is studied: the effect parameter (with the classes LC50 and NOEC). For the LC50 data in the second step (B) three characteristics are studied further: whether the tested substance is a pesticide; whether the source of the toxicity data is listed in Scopus; and the year of publication. Five characteristics are studied in the third step (C) with data whose source is listed in Scopus. These are all related to the source of the toxicity data: the number of citations to the source; the continent from which the first author published the data; the first author's h index; the number of co-authors associated with the first author; and the impact

Table 2.1: Characteristics used in analysis and the step to which it applies (see also Fig. 2.1). The steps are: (A) the complete US EPA ECOTOX database; (B) LC50 records selected from the database; (C) LC50 records from the database for which its source is listed in Scopus. Classification of ecotoxicological data before analysis is based on collected meta-information. This table gives the intervals or categories as selected in the present study for the classification of the data.

Characteristic	Step	Number of classes	Class intervals/categories			
			Class I	Class II	Class III	Class IV
Effect parameter	A	2	LC50	NOEC		
The substance is a pesticide	B	2	Yes	No		
The source of the data is listed in Scopus	B	2	Yes	No		
Publication year[2]	B	4	< 1970	$[1970, 1980)$	$[1980, 1990)$	≥ 1990
Number of citations to the source	C	4	< 5	$[5, 15)$	$[15, 20)$	≥ 20
Continent associated with the first author	C	8	See main text, section 2.2.3			
h Index of the first author	C	4	NA[3]	$[0, 5)$	$[5, 10)$	≥ 10
Size network first author[4]	C	4	< 10	$[10, 30)$	$[30, 80)$	≥ 80
Impact factor of the journal (2010)	C	4	< 1	$[1, 2.5)$	$[2.5, 3)$	≥ 3

[2] Via ECOTOX database.
[3] Not available; authors that have not published after 1995.
[4] Number of co-authors associated with the first author.

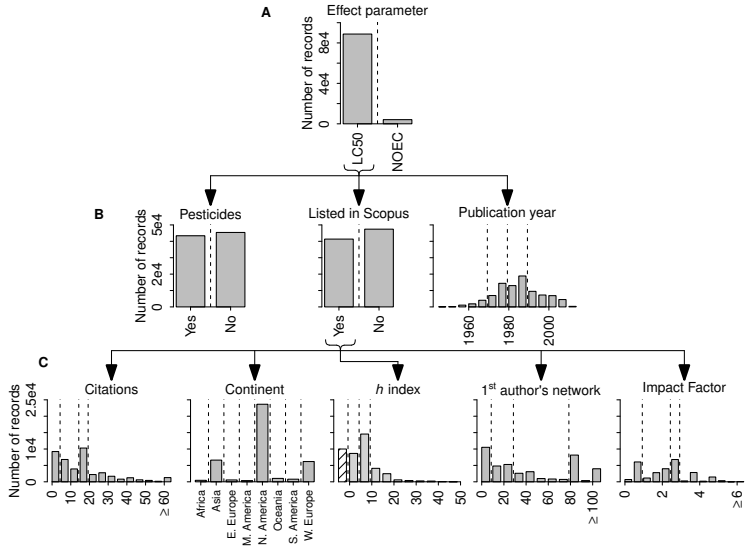


Fig. 2.1: Bar plots show the number of records in the ECOTOX database for which specific meta-information is available (see Table 2.1). Information is shown per level of detail (step): (A) the complete US EPA ECOTOX database; (B) LC50 records selected from the database; (C) LC50 records from the database for which its source is listed in Scopus. Within each step, the y-axis is scaled identically for all plots. Vertical dashed lines indicate class intervals as selected and specified in Table 2.1. The shaded bar in the 'h index' plot indicates the number of records for which no 'h index' was available as the authors in question have not published after 1995.

factor in 2010 of the journal in which the data were published. An overview of the characteristics studied in these steps is presented in Fig. 2.1 and Table 2.1. Most of the characteristics of the last two steps (B and C) required additional meta-information, which was not available from the ECOTOX database. The approach for the collection of this additional information is described below.

2.2.3 Collecting additional meta-information

In the present study a substance is considered to be a pesticide (Fig. 2.1 and Table 2.1) when it is listed as such, by CAS number, in Alan Wood's pesticide compendium (`http://www.alanwood.net/pesticides`, accessed on 9th May 2012). A substance is considered to be a non-pesticide when it is not listed there.

If available, information on the source for each LC50 record in the ECOTOX database is retrieved from Scopus (`http://www.scopus.com`, accessed on 3rd January 2013). This is done by searching Scopus using the combination of the first author's name, article title and publication year as listed in the ECOTOX

database. For each reference retrieved from Scopus the number of citations to the paper is recorded. Also the continent from which the publication was submitted (based on the affiliation address of the first author) is registered. For a better socioeconomic distinction, the continent of Eurasia is further divided into Eastern Europe (former Warsaw Pact countries), Western Europe and Asia. For the same reason we also distinguished Middle America from South and North America.

For each first author, additional information is collected from Scopus. Firstly the *h* index, which is the greatest number *h* such that *h* publications by the author have been cited at least *h* times. In Scopus the index is based on publications after 1995. Secondly, the total number of co-authors associated with the first author based on the used and other publications (even if the author is not first author for those) by the first author (also from Scopus).

For papers that are listed with a journal International Standard Serial Number in Scopus, the journal five-year impact factor is obtained from Thomson Reuters Journal Citation Report (2010 edition[5]) (characteristic 'impact factor' in Table 2.1). This characteristic is the average number of times papers, from the journal published in the past five years, have been cited in the reference year (2010, in this case).

2.2.4 Data analysis

Before the actual analysis, the association between the applied characteristics, after classification, is determined using Pearson's χ^2 test for categorical data. When characteristics are associated, it is impossible to tell which of those characteristics is responsible for possibly observed effects on compliance with Benford's Law.

Compliance with Benford's Law is also tested with χ^2 test statistics. For each digit i this test determines the difference between the observed fraction of leading non-zero digits (O_i) and the expected fraction (E_i), namely Benford's distribution ($E_i = Prob(D_1 = i)$, see Eqn. 2.1). The differences (expressed as the squared difference between the observed and expected fractions, divided by the expected fraction) for all nine digits are summed and multiplied with total number of observations:

$$\chi^2 = N \sum_{i=1}^{9} \frac{(E_i - O_i)^2}{E_i} \tag{2.2}$$

There is considerable variation in the amount of data in each of the classes (N_{class}) analysed in the present study, which is unfortunate as the χ^2 test is sensitive for sample size (Rauch et al., 2011). To overcome this problem, a random

[5] `http://admin-apps.webofknowledge.com/JCR`, accessed on 12th May 2012

subsample with a fixed size ($N_{sub} = 500$) is drawn from each previously specified class (Table 2.1). Sampling is performed without replacement as replacing samples will result in replicated samples, which may adversely affect compliance with Benford's Law. For each class a large number of such subsamples are simulated ($N_{sim} = 10,000$) and for each subsample χ^2 statistics are calculated (where $N = N_{sub} = 500$). Obviously, this is only possible if the size of the class (N_{class}) is larger than N_{sub}. Furthermore, this simulation becomes particularly meaningful when N_{class} is considerably larger than N_{sub} ($N_{class} > 10N_{sub}$). For each class we determine the fraction of the 10,000 simulated subsamples for which the calculated χ^2 value exceeds the critical value of 15.5073 ($\alpha = 0.05$). A class is considered noncompliant with Benford's Law when this fraction of rejected χ^2 tests is larger than the alpha level of 0.05. We also use this fraction as a relative indicator of how well a class complies with Benford's Law.

The χ^2 value is also calculated for each entire class, without subsampling simulation. As a reference, this χ^2 is again calculated but now assuming a uniform distribution of first digits ($E_i = \frac{1}{9}$). The χ^2 value calculated assuming a Benford distribution is compared with that assuming a uniform distribution.

2.3 Results

2.3.1 Collected data

The number of records found for each sub-group (class) is shown in Fig. 2.1 and Table 2.1. The complete ECOTOX database contains 607,679 ecotoxicological test results, of which 15% represents either of the selected effect parameters (LC50s and mortality NOECs) and is expressed in μg/L (or a decade of this unit). There are 22 times more LC50 values than mortality NOEC values in the database (Fig. 2.1A). Of the LC50 data, the fraction representing pesticides is similar to that of non-pesticides (Fig. 2.1B). The same is true for the fraction of records listed in Scopus and that not listed in Scopus (Fig. 2.1B). The records listed in Scopus are primarily peer reviewed papers.

Earliest publications on LC50 values in the database originate from the 1940s, followed by a growth in number of publications which peaked in the 1990s, after which the volume of LC50 publications declined (Fig. 2.1B).

There are relatively few records in the database with higher values for the bibliometric parameters (the number of citations to the paper, the first author's *h* index, the number of co-authors associated with the first author and the journal's five-year impact factor) (Fig. 2.1C). For the number of co-authors associated with the first author, an inexplicable peak is observed between 80 and 90 co-authors (Fig. 2.1C).

Most material (LC50 records that are listed in Scopus) is submitted from North America, followed by Asia and Western Europe respectively (Fig. 2.1C).

2.3.2 Data analysis

There is strong evidence that there is no association between any of the characteristics ($p < 0.001$, Pearson's χ^2 test), except for the characteristics 'listed in Scopus' and 'pesticides'. Apparently tests carried out with pesticides are slightly less listed in Scopus ($p = 0.55$, Pearson's χ^2 test).

The compliance with Benford's Law is expressed as the χ^2 value of the random subsamples (Fig. 2.2) and the fraction of those samples that exceed the critical value at $\alpha = 0.05$ (Fig. 2.3). The observed distribution of leading digits for both NOEC and LC50 data is shown in Fig. 2.4 together with the expected Benford distribution.

The χ^2 values of the random subsamples generally follow the same pattern as the χ^2 for the entire class (Fig. 2.3). For some classes (e.g., 'publication year') they deviate, probably due to the varying size of each class. The ratio between the χ^2 value assuming a uniform versus a Benford distribution of first digits ranges from 4.1 up to 901 with a median value of 207 (data not shown) for all classes. This shows as expected that the first digits are more likely to follow a Benford distribution rather than a uniform distribution.

In step B the publication year is particularly related to compliance with Benford's Law: more recently published LC50 data fit better to Benford's Law than older work (Fig. 2.2B and 2.3B). Slightly higher compliance with Benford's Law was found for non-pesticide data (0.06, the fraction of simulations exceeding the critical value) compared to pesticide data (0.07) and for data listed in Scopus (0.06) compared to data not listed (0.07) (Fig. 2.2B and 2.3B).

With increasing number of citations to the source of the data the compliance with Benford's Law decreases. However, the fit to Benford's distribution improves again for papers with twenty or more citations (Figs. 2.2C and 2.3C). The goodness of fit of LC50 data varies considerably among the continents associated with the laboratories of the first authors (Figs. 2.2C and 2.3C). Some of the 'continent classes' are relatively small ($N_{class} \leq 10N_{sub}$ or for Africa and Middle America even $N_{class} \leq N_{sub}$) so that the subsamples do not represent a random part of the total pool, which is required for a proper analysis. Eastern Europe (0.005, fraction of simulations exceeding the critical value) produced LC50 data that are most conform to Benford's Law, followed by North America and Asia (0.07 and 0.08, respectively). Data produced in Oceanea, South America and Western Europe (0.3, 0.7 and 0.2, respectively) deviate more from Benford's Law, where only the results of the latter continent is based on sufficient data ($N_{class} > 10N_{sub}$).

The compliance with Benford's Law increases with an increasing h index of the first author (Figs. 2.2C and 2.3C). The fit of LC50 data to Benford's Law decreases with increasing number of co-authors associated with the first author (Figs. 2.2C and 2.3C). There is no clear trend for the goodness of fit to Benford's Law as a function of the journal's impact factor (Figs. 2.2C and 2.3C).

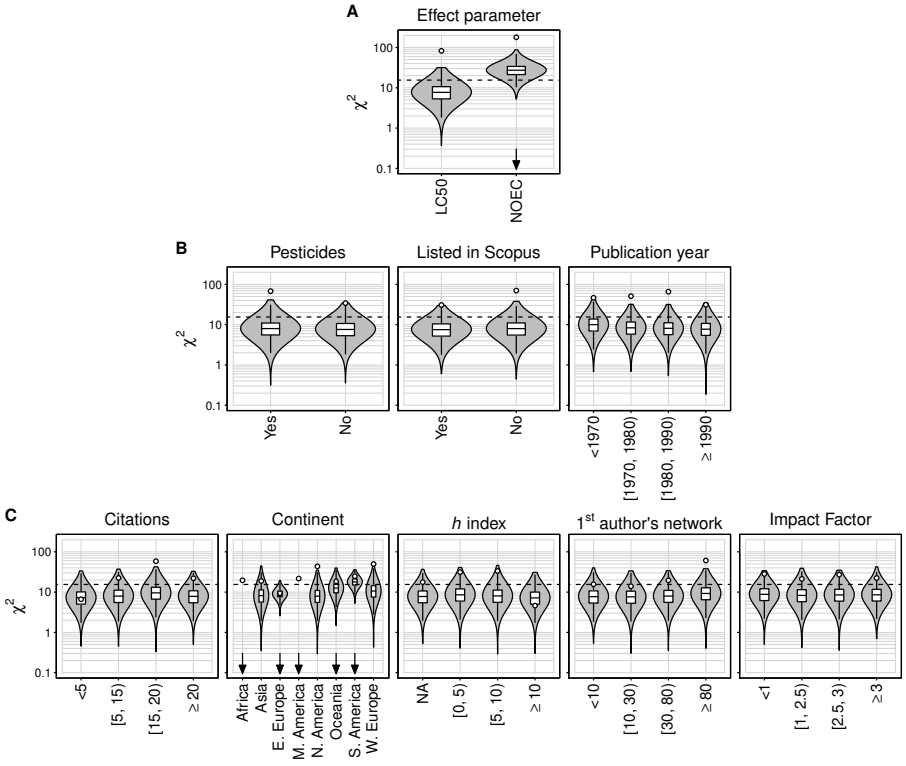


Fig. 2.2: The goodness of fit, expressed as χ^2, of observed leading non-zero digits against Benford's distribution of digits for the N_{sim} = 10,000 simulated subsamples shown for each class of each characteristic (Table 2.1 and Fig. 2.1). The dashed line shows the critical value of 15.5073 at α = 0.05, simulations above this value are in violation of Benford's Law. If sufficient data are available for subsampling ($N_{class} > N_{sub}$) the boxes indicate first (bottom), second (middle) and third (top) quartiles of the subsamples; whiskers indicate minimum and maximum values (excluding extremes outside 1.5 times the interquartile range). Grey violin shapes surrounding the boxes show the kernel density of the same data. The χ^2 value for the entire class without subsampling is indicated with a marker (∘). Classess for which simulations are less reliable ($N_{class} \leq 10N_{sub}$) are marked with arrows. Results are shown per step to which characteristics apply: (A) the complete US EPA ECOTOX database; (B) LC50 records selected from the database; (C) LC50 records from the database for which its source is listed in Scopus.

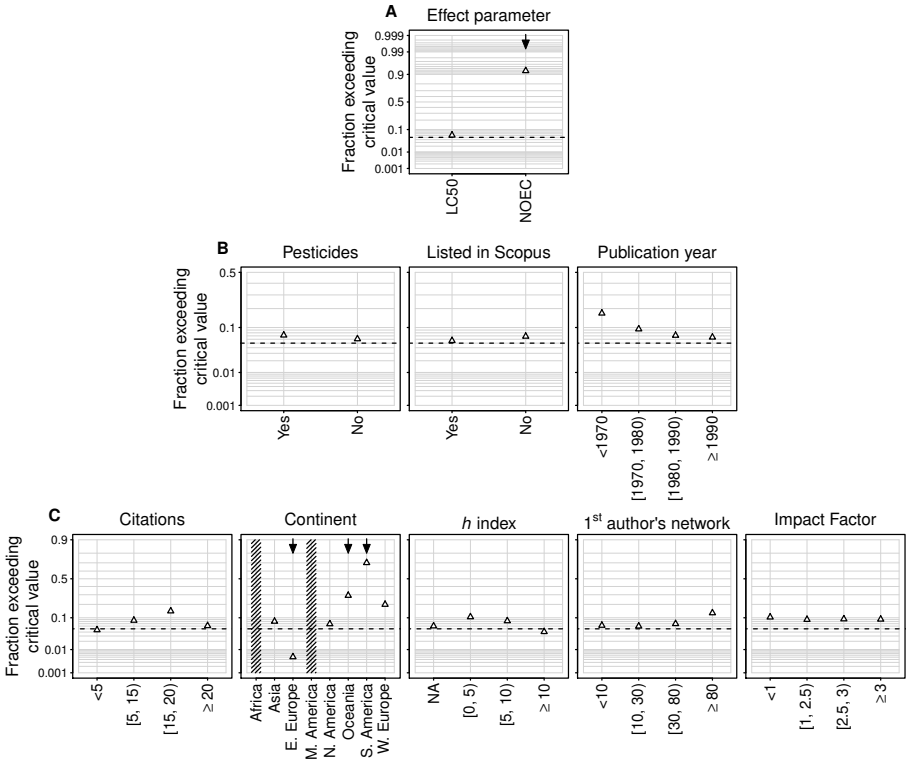


Fig. 2.3: The fraction of the N_{sim} = 10,000 simulated subsamples that is in violation of Benford's Law (tested with χ^2 statistics with α = 0.05), shown for each class of each characteristic (Table 2.1 and Fig. 2.1) (△). The dashed line shows the fraction equal to the α level at which was tested (0.05). Markers above the dashed line indicate that more simulated subsamples violate Benford's Law than expected from pure chance, for a class compliant with Benford's Law. Classess for which simulations are less reliable ($N_{class} \leq 10N_{sub}$) are marked with arrows, classes for which simulation was not possible ($N_{class} \leq N_{sub}$) are indicated with shaded bands. The y-axis is presented on a normal probability scale. Results are shown per step to which characteristics apply: (A) the complete US EPA ECOTOX database; (B) LC50 records selected from the database; (C) LC50 records from the database for which its source is listed in Scopus.

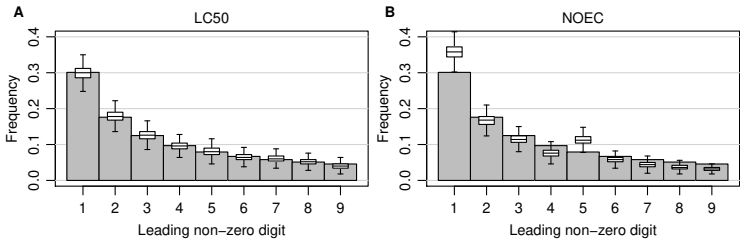


Fig. 2.4: Grey bars show the expected Benford distribution of leading non-zero digits ($E_i = Prob\,(D_1 = i)$) as calculated with Eqn. 2.1 (paragraph 2.1.1). White boxplots are observed frequencies (O_i) of leading digits for (A) LC50 data and (B) NOEC data. The boxes indicate first (bottom), second (middle) and third (top) quartiles of the N_{sim} = 10,000 subsamples, each containing N_{sub} = 500 data points (total amounts of data are shown in Fig. 2.1); whiskers indicate minimum and maximum values (excluding extremes outside 1.5 times the interquartile range).

2.4 Discussion

The present study investigates whether and, if so, how much publically available ecotoxicological data (LC50s and mortality NOECs) depart from Benford's Law. Such violations could indicate irregularities in the toxicity data. Also the possible relationship of several characteristics of the data sources with violations of Benford's Law is studied.

2.4.1 Discussion of results

Benford's Law can be used to find abnormal deviations from 'natural numbers'. The fact that they deviate from what is expected, does not explain why they deviate. It could be the result of scientific misconduct, but also because human choices influence the numbers. the latter is, i.e., the explanation for the difference in compliance with Benford's Law between LC50 and mortality NOEC (Figs. 2.2A, 2.3A and 2.4). NOECs are based on test concentrations and apparently scientists prefer such concentrations to start with the digits '1' and '5' (whole and half units) (Fig. 2.4B). This preference makes it impossible to apply Benford's Law to NOEC data. When NOECs are based on measured concentrations (rather than nominal concentrations), the fit to Benford's distribution improves but is still in violation of the law (data not shown). This provides an additional argument to the existing list (e.g., Jager (2012)) to avoid the use of NOECs in risk assessment.

As can be seen from Fig. 2.1, the number of published LC50 data is declining. The LC50 was selected in the present study to produce proof of principle of the applicability of Benford's Law for interpolated effect concentrations as opposed to chosen concentrations (such as NOECs). Current developments show a focus

on other interpolated effect concentrations, such as the EC50, the No Effect Concentration (Kooijman et al., 1996) or Benchmark Dose (Crump, 1984), endpoints for which Benford's Law is expected to apply as well.

Potentially, conflicts of interests could give rise to data manipulation and therefore lower compliance with Benford's Law. This may be the case for pesticides. However, only a small difference in compliance with Benford's Law was found between pesticides and non-pesticides (0.07 and 0.06 respectively, Figs. 2.2B and 2.3B). This is far less than, e.g., the difference in compliance found between the publication years (Figs. 2.2B and 2.3B). The same is true for the characteristic 'listed in Scopus' (Figs. 2.2B and 2.3B). Toxicity data whose source is listed in Scopus are generally published in peer-reviewed journals, whereas those that are not listed in Scopus are generally published as 'grey literature'. Hence, data from 'grey literature' is not necessarily less reliable, as peer reviews usually don't focus on the raw data produced, and tests performed under the OECD Principles of Good Laboratory Practice (GLP) (OECD, 1998) (where all raw data should be traceable) are often published in 'grey literature'. It would be interesting to compare compliance with Benford's Law of data that are either generated under GLP or not.

Violation of Benford's Law is reduced with increasing publication year (Figs. 2.2B and 2.3B). This is constistent with the idea that reliability of studies has improved over the last decades as a result of, for instance, standardisation of tests and the introduction of GLP (Purchase, 2004). Other factors have also changed over time which can be potential causes of the observed trend. For instance, the method for interpolating the LC50 value has changed over time. In early days it was not uncommon to derive effect concentrations from the dose-response curve using semi-graphical techniques (e.g., Litchfield and Wilcoxon (1949)). With computational power increasing over time, more sophisticated techniques such as maximum likelihood methods could also be applied (Newman, 2013). The formalisation and standardisation of LC50 testing over time can also have had an effect.

An explanation for the greater deviation from Benford's Law for first authors with a very large or very small network of co-authors (Figs. 2.2C and 2.3C) cannot be given. One could speculate that these reflect relatively unexperienced scientists or first authors that write reviews for large groups of collaborators who therefore have less control over the original data. Further analysis of the data is needed to explain these findings. It should be noted that the network size is that of the author at the date it was extracted from the Scopus database for this study and not at the moment of the publication of the data in question. This results in a discrepancy between older and newer publications. The same is true for most bibliometric characteristics (number of citations, h index, co-authors associated with first author and journal impact factor). However, we found no association between these bibliometric characteristics and characteristic 'publication year'

as pointed out earlier.

The *h* index of the first author seems to be a good indicator for compliance with Benford's Law. As the *h* index is an indicator of seniority and publication success, this is to be expected.

The number of times a publication is cited or the journal impact factor showed no evident relation with compliance with Benford's Law (Figs. 2.2C and 2.3C). This may be due to the fact that this study focused on LC50, and for external peers it is very hard to judge the reliability of the underpinning data.

Considerable variation exists between compliance with Benford's Law and the geographical origin of the data (characteristic 'continent', Figs. 2.2C and 2.3C). Indications of violations of Benford's Law for most continents are based on relatively small data sets ($N_{class} \leq 10 N_{sub}$). Conclusions for those continents should therefore be made with caution. An exception is perhaps Western Europe, where 20% of the simulations were in violation of Benford's Law (Fig. 2.3C), where only 5% was expected based on chance, and a considerable amount of data was available for this analysis. Why data from this continent show such deviations from Benford's Law is unclear and requires further study (see also Appendix 2.A).

2.4.2 Discussion of methods

In order to assess whether ecotoxicological data that are considered reliable according to Klimisch criteria (Klimisch et al., 1997) also comply to Benford's Law, thereby validating the applicability of Benford's Law, one ideally has a large dataset in which both reliable and unreliable data are scored. Such a dataset is presently not directly available. As indicated previously, the use of scoring approaches (e.g., Klimisch et al. (1997)) would be time-consuming for a large dataset. In addition, they do not account for aspects of reliability (e.g., integrity with which experiments were conducted). Moreover, extremes in this perspective, e.g., cases of scientific misconduct (data manipulation), are relatively rare and if they do occur they may already have been retracted from literature (Nigg and Radulescu, 1994).

The present study focuses on vast amounts of data in order to show the principles of Benfords Law. As a consequence, the characteristics studied here had to be derived from readily available information extracted from databases. Therefore not allowing the analyses of all relevant characteristics. One of the characteristics that would be interesting to include in future research would be the funding source (e.g., private or public) of the studies, as this may indicate conflicts of interests.

Classification of numerical characteristics was required as compliance with Benford's Law cannot be determined for individual data records. Consequently, choices were made as to the number of classes and the numerical intervals that defined the classes. In the present study four roughly equally sized classes

were used as this could be implemented consistently for all characteristics. Although more classes could be specified, less data remain to analyse per class. A disadvantage of working with categorical characteristics rather than numerical characteristics is that it is probably less sensitive for determining the association between characteristics.

The only (negative) association found in the dataset of the present study was between the characteristic 'listed in Scopus' and 'pesticide'. Many tests with pesticides are obligatory for registration or approval purposes and are therefore reported in 'grey literature', as they are of less interest to the scientific community. The association is likely caused by the fact that Scopus includes mostly peer-reviewed papers and little to no 'grey literature'. Because of this association, it is impossible to establish whether the compliance with Benford's Law relates to being listed in Scopus or to the test substance being a pesticide or not. Fortunately no association was found between the other characteristics.

There are many alternatives to determine the goodness of fit to Benford's Law (e.g., normalised Euclidean distance (d^*), distance measure (a^*) and Kuiper's test (V_N) as used by Tam Cho and Gaines (2007); Judge and Schechter (2009); Rauch et al. (2011)). The present study served as a first exploration of compliance of ecotoxicological data compliance with Benford's Law, where the focus is on relative comparisons rather than on strict null hypothesis testing. This is the reason only χ^2 testing with re-sampling is applied. Other tests should be included when null hypothesis testing becomes more critical.

2.4.3 Implication for data quality assessment

In the present study no direct comparison is made between compliance with Benford's Law and existing reliability indicators (e.g., Klimisch et al. (1997)). It is interesting to determine whether there is a correlation between existing reliability indicators and compliance with Benford's Law. If such correlations exist, applying existing indicators may suffice to cover all aspects of reliability. If not, it may be advisable to add testing compliance with Benford's Law.

Not only can Benford's Law serve as an indicator of reliability, it can also be used effectively to trace observed irregularities in a dataset to its source (Appendix 2.A). One way of doing this, is by analysing specific subsets of the data, as is done in the present study for specific characteristics (Fig. 2.1 and Table 2.1). Such analyses provide information on subgroups in which higher violations of Benford's Law are observed and narrows down the amount of data that needs to be evaluated in detail (see Appendix 2.A).

Another option is to extend Benford's Law to the second leading digit (or any subsequent digit for that matter; Hill (1995)). Deviation of a specific combination of the two leading digits from Benford's distribution can also strongly narrow down the amount of data and their sources to be evaluated (Appendix 2.A). Num-

bers that have been rounded to the first significant digit can interfere with this approach as this will lead to an overrepresentation of the second leading digit '0'. For example, if the number 0.83 is rounded to 0.8, or worse presented as 0.80, the second leading digit '0' is overrepresented (see Appendix 2.A for more details). It is therefore recommended that the number of significant digits for each number in a data set is reported for each record, in order be able to compensate for this phenomenon.

Screening techniques as described above (and illustrated in the Appendix 2.A) can be valuable in systems in which toxicity data are collected for large amounts of chemicals. Especially if potential conflicts in interests have been identified. For instance, in both the European Commission regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH; EC (2006)) and OSPAR's Harmonised Mandatory Control System for chemicals used in the offshore oil and gas industry (OSPAR, 2000), the industry is responsible for supplying toxicity data, where chemicals with low toxicity are favourable for the industry. In such cases evaluating reliability of the data is key. More specifically, the technique can be used after registration to narrow down a subset of registered data that requires a more detailed inspection of underpinning documentation.

The present study does not yet evaluate the applicability of Benford's Law for assessing scientific data more in general, e.g., EC50 data from a variety of validated or new studies. With such an evaluation the suggested increased rate of scientific fraud (Fang et al., 2012) can be investigated.

2.5 Conclusions

The methodology presented here can successfully identify deviations from Benford's Law for large data sets of interpolated (no) effect concentrations. This approach could be used as a quality indicator in addition to existing ones.

The application of Benford's Law can also be used to efficiently trace sources of irregularities in large data sets, without having to evaluate the source of each individual record.

Acknowledgment

Geke Lianne den Otter is acknowledged for proofreading and anonymous reviewers are for constructive feedback on the draft manuscript.

References

Ågerstrand, M., Breitholtz, M., Rudén, C., 2011. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessments of pharmaceutical substances. Environmental Sciences Europe 23 (1), 17.

Benford, F., 1938. The law of anomalous numbers. P Am Philos Soc 78 (4), 551–572.

Berger, A., Hill, T., 2011. Benford's Law strikes back: No simple explanation in sight for mathematical gem. Math Intell 33, 85–91.

Crump, K. S., 1984. A new method for determining allowable daily intakes. Toxicol Sci 4 (5), 854–871.

De Zwart, D., 2002. Observed regularities in species sensitivity distributions for aquatic species. In: Posthuma, L., Sutter II, G., Traas, T. (Eds.), Species sensitivity distributions in ecotoxicology. CRC Press, Boca Raton, FL, USA.

Dumas, C. F., Devine, J. H., August 2000. Detecting evidence of non compliance in self-reported pollution emissions data: an application of Benford's Law. In: American Agricultural Economics Association 2000 Annual meeting. Tampa, FL, USA.

Durtschi, C., Hillison, W., Pacini, C., 2004. The effective use of Benford's Law to assist in detecting fraud in accounting data. JFA 5, 17–34.

EC, 2006. Corrigendum to 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, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union L136, 3–280.

Fang, F. C., Steen, R. G., Casadevall, A., 2012. Misconduct accounts for the majority of retracted scientific publications. P Natl Acad Sci 109 (42), 17028–17033.

Fewster, R. M., 2009. A simple explanation of Benford's Law. Am Stat 63 (1), 26–32.

Geyer, C. L., Williamson, P. P., 2004. Detecting fraud in data sets using Benford's Law. Commun Stat-Simul C 33 (1), 229–246.

Hill, T. P., 1995. The significant-digit phenomenon. Am Math Mon 102 (4), 322–327.

Jager, T., 2012. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. Environ Toxicol Chem 31 (2), 228–229.

Judge, G., Schechter, L., 2009. Detecting problems in survey data using Benford's Law. J Hum Resour 44 (1), 1–24.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharm 25 (1), 1–5.

Kooijman, S. A. L. M., Bedaux, J. J. M., Slob, W., 1996. No-effect concentration as a basis for ecological risk assessment. Risk Anal 16 (4).

Litchfield, J. T., Wilcoxon, F., 1949. A simplified method of evaluating dose-effect experiments. J Pharmacol Exp Ther 96 (2), 99–113.

Marchi, S., Hamilton, J., 2006. Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. J Risk Uncertainty 32, 57–76.

Newcomb, S., 1881. Note on the frequency of use of the different digits in natural numbers. Am J Math 4 (1), 39–40.

Newman, M. C., 2013. Quantitative Ecotoxicology, 2nd Edition. CRC Press, Boca Raton, FL, USA.

Nigg, H. N., Radulescu, G., 1994. Scientific misconduct in environmental science and toxicology. JAMA-J Am Med Assoc 272 (2), 168–170.

OECD, 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. No. 1 OECD Principles of Good Laboratory Practice (as revised in 1997). Organisation for Economic Co-operation and Development, Paris.

OSPAR, 2000. OSPAR Decision 2000/2 on a Harmonised Mandatory Control System for the use and reduction of the discharge of offshore chemicals (as amended by OSPAR Decision 2005/1).

Purchase, I. F., 2004. Fraud, errors and gamesmanship in experimental toxicology. Toxicology 202 (1-2), 1–20.

Rauch, B., Göttsche, M., Brähler, G., Engel, S., 2011. Fact and fiction in EU-governmental economic data. Ger Econ Rev 12, 243–255.

Tam Cho, W. K., Gaines, B. J., 2007. Breaking the (Benford) Law. Am Stat 61 (3), 218–223.

Wheeler, J., Grist, E., Leung, K., Morritt, D., Crane, M., 2002. Species sensitivity distributions: data and model choice. Mar Pollut Bull 45 (1-12), 192–202.

2.A Appendix

2.A.1 Introduction

In this appendix two examples are provided on how Benford's Law can be used to effectively trace irregularities to a single source. The examples also show that caution is required when interpreting anomalies.

2.A.2 Example I

One strategy in tracing the source of irregularities in a data set is analysing compliance with Benford's Law of subsets drawn from the data. In this example the goodness of fit of leading digits to Benford's distribution is calculated for all individual publications, listed in Scopus, that contain more than 50 LC50 values. For this purpose χ^2 statistics is applied to each publication. This analysis showed that nearly half the publications exceed the χ^2 critical value of 15.5073 ($\alpha = 0.05$) for compliance (Fig. 2.5).

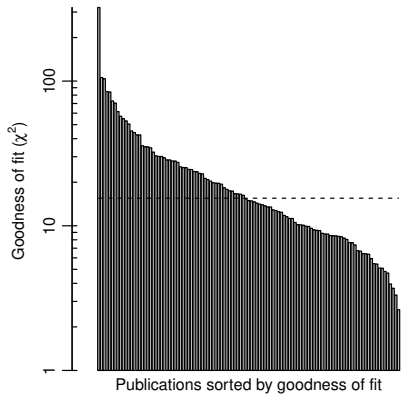


Fig. 2.5: Publications that contain more than 50 LC50 values and are listed in Scopus. Publications are sorted based on goodness of fit to Benford's distribution (leading digit), calculated with χ^2 statistics. Dashed line indicates critical value for the χ^2 statistics at $\alpha = 0.05$. For publications above the dashed line, it is unlikely that the first digits of the LC50 data are distributed conform the Benford distribution.

The publication with the worst fit to Benford's Law is selected for further inspection (left in Fig. 2.5). The leading digit '1' is clearly underrepresented whereas the digit '4' is overrepresented in the publication by Vedamanikam and Shazilli (2008) (Fig. 2.6). Overrepresentation of specific digits is not tested for significance in these examples, although this is possible with *z*-satistics as described by Durtschi et al. (2004).

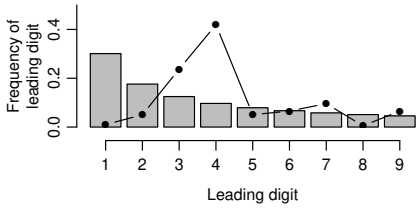


Fig. 2.6: Observed (•) frequency of leading digits in the publication (Vedamanikam and Shazilli, 2008) that deviates most from Benford's Law (left in Fig. 2.5; N = 199; χ^2 = 323). Expected Benford distribution is shown as grey bars.

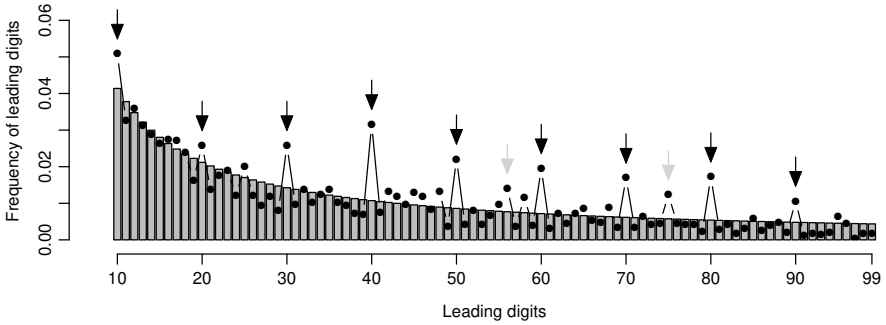


Fig. 2.7: Observed (•) frequency of the first two digits in a sub-selection of the data set. Only LC50 values listed in Scopus, published in the year 1990 or earlier and published from Western Europe are analysed. Grey bars indicates the expected Benford distribution $Prob((D_1, D_2) = (d_1, d_2))$ (Eqn. 2.3). Black arrows indicate overrepresentation of digit-combinations as a result of rounded numbers in the data set. Grey arrows show 'suspicious' overrepresentation of specific digit-combinations.

Upon closer inspection of the work by Vedamanikam and Shazilli (2008) there is actually a good reason for the deviation from Benford's Law. In the study by Vedamanikam and Shazilli (2008) the toxicity of a number of metals is tested for a selection of species and as a function of temperature, resulting in 199 LC50 values. Many of the LC50 values only shift within a small range as a result of the temperature dependence. For most combinations of test species and test substance there is less than a factor 20 difference between the minimum and maximum LC50 value in this publication. The data don't span several orders in magnitude and therefore don't comply to Benford's Law as this is one of the criteria listed in the main text. This could also explain why such a large part of the publications in general don't comply to Benford's Law (Fig. 2.5), if the LC50 values in those publications also don't span several orders in magnitude.

2.A.3 Example II

Example II uses an extension of Benford's Law to the combination of the first two leading digits (rather than only the first digit). This extension is expressed as (Hill, 1995):

$$Prob\left((D_1, D_2) = (d_1, d_2)\right) = \log_{10}\left(1 + 1/\left(10d_1 + d_2\right)\right), \quad \text{for } d_1 = 1, \ldots, 9 \text{ and } d_2 = 0, \ldots, 9; \tag{2.3}$$

where the left-hand term indicates the probability that the first (D_1) and second (D_2) digit in a 'natural' data set equals specific digits (d_1 and d_2 respectively). For example, the chance that the first digit is equal to '2' followed by the second digit '5' is $Prob\left((D_1, D_2) = (2, 5)\right) = \log_{10}\left(1 + \frac{1}{20+5}\right) \approx 0.017$. Note that the combination of first digits ranges from 10 up to 99.

First, only data are selected that are submitted from Western Europe and published in 1990 or earlier, as it is known from the analysis in the main text that the deviation from Benford's Law is relatively high for data with those characteristics. The first two digits of those data are analysed (Fig. 2.7).

The second digit '0' is clearly overrepresented in the analysed subset (blue arrows in Fig. 2.7). Most likely, this is the result of rounding data to the first significant digit. Overrepresentation of the digit combinations '56' and '75' (red arrows in Fig. 2.7) cannot be explained by the rounding of numbers. The data with first digit combination '75' is selected for closer inspection.

Note that in this procedure the amount of data that will be inspected is narrowed down considerably. The database contains 41,363 LC50 records that are listed in Scopus. After restricting data to publications from Western Europe and 1990 and earlier, 3,479 records remained. Next, records starting with the digits '75' are selected ending up with only 43 records, almost a thousand fold less than we started with.

Finally, similarities within the remaining 43 records are identified. It was found that 10 out of the 43 records were authored by Slooff and colleagues. These publications (Slooff et al., 1983; Slooff, 1983; Canton et al., 1985) are now analysed for compliance with Benford's Law (first significant digit only, Fig. 2.8). The leading digit '7' is overrepresented for these publications (Fig. 2.8) and may partially explain why the digit combination '75' is observed with such a high frequency in the selected data set (Fig. 2.7).

At first glance, the deviation of these publications from Benford's Law is less apparent than in the first example: a wide range of substances is tested on a wide variety of species, data is spanning several orders in magnitude. However, two of the publications (Slooff et al., 1983; Slooff, 1983) contain identical LC50 values for *Hydra oligactis* and *Lymnaea stagnalis*, probably the result of the same test. After removing these doubles, the leading digit '7' is slightly less overrepresented and the goodness of fit to Benford's distribution improves from

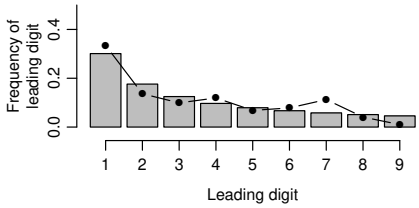


Fig. 2.8: Observed (•) frequency of leading digits in the publications authored by Slooff and others (Slooff et al., 1983; Slooff, 1983; Canton et al., 1985) as selected in example II (N = 293; χ^2 = 31.6). Expected Benford distribution is shown as grey bars.

χ^2 = 31.6 (N = 293) to χ^2 = 22.0 (N = 263), which is still greater than the critical value of 15.5073 (α = 0.05). The cause of the remaining deviation from Benford's distribution is still unknown.

References in Appendix 2.A

Canton, J. H., Slooff, W., Kool, H. J., Struys, J., Pouw, T. J. M., Wegman, R. C. C., Piet, G. J., 1985. Toxicity, biodegradability, and accumulation of a number of Cl/N-containing compounds for classification and establishing water quality criteria. Regul Toxicol Pharm 5 (2), 123–131.

Durtschi, C., Hillison, W., Pacini, C., 2004. The effective use of Benford's Law to assist in detecting fraud in accounting data. JFA 5, 17–34.

Hill, T. P., 1995. The significant-digit phenomenon. Am Math Mon 102 (4), 322–327.

Slooff, W., 1983. Benthic macroinvertebrates and water quality assessment: Some toxicological considerations. Aquat Toxicol 4 (1), 73–82.

Slooff, W., Canton, J. H., Hermens, J. L. M., 1983. Comparison of the susceptibility of 22 freshwater species to 15 chemical compounds. I. (sub)acute toxicity tests. Aquat Toxicol 4 (2), 113–128.

Vedamanikam, V. J., Shazilli, N. A. M., 2008. Comparative toxicity of nine metals to two Malaysian aquatic dipterian larvae with reference to temperature variation. B Environ Contam Tox 80, 516–520.

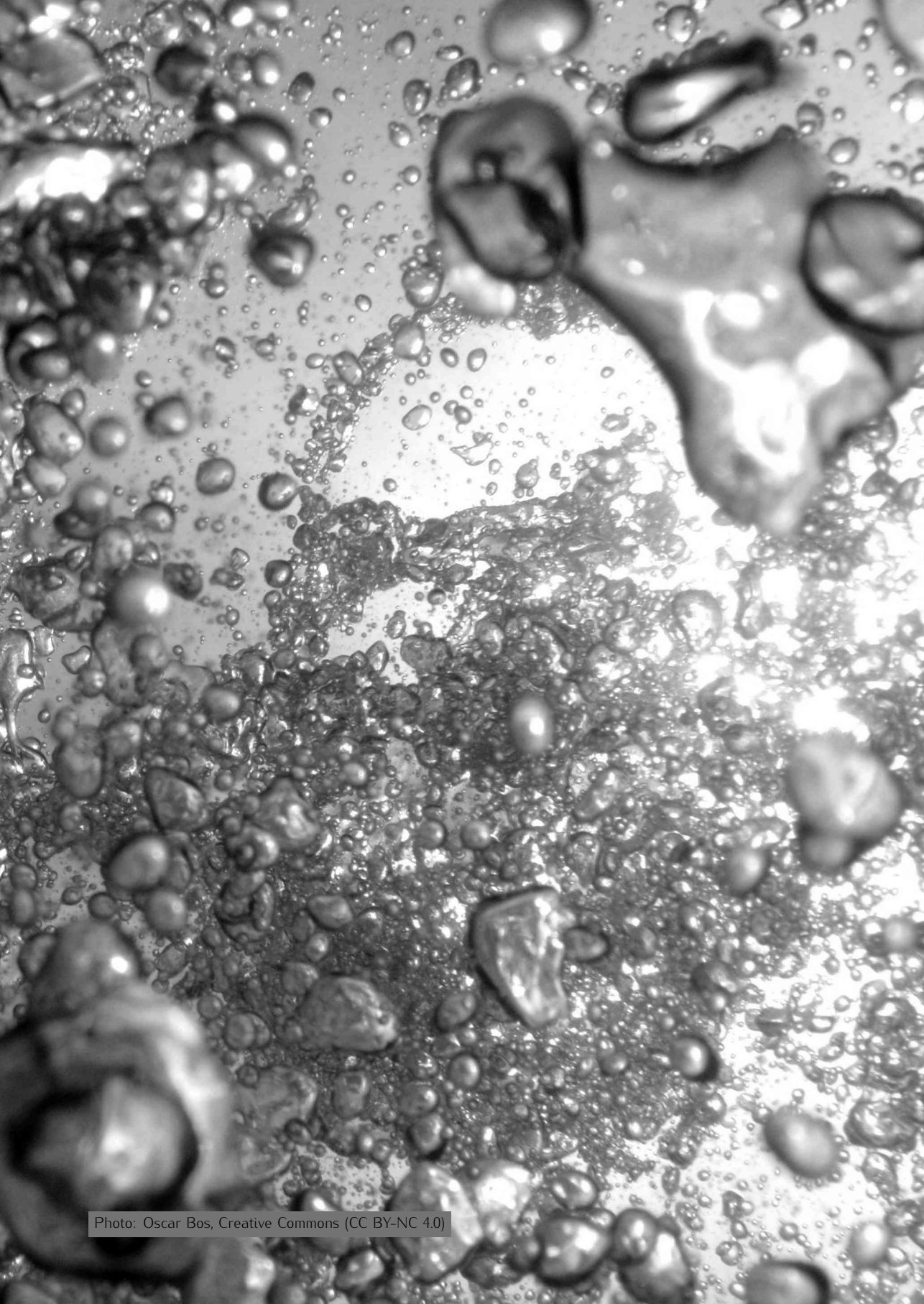


Photo: Oscar Bos, Creative Commons (CC BY-NC 4.0)

Chapter 3

Towards quantitative ecological risk assessment of elevated carbon dioxide levels in the marine environment

Reprinted from Marine Pollution Bulletin 73(2), Pepijn de Vries, Jacqueline E. Tamis, Edwin M. Foekema, Chris Klok and Albertinka J. Murk (pages 516–523, copyright (2013)), with permission from Elsevier

Abstract

The environmental impact of elevated carbon dioxide (CO_2) levels has become of more interest in recent years. This, in relation to globally rising CO_2 levels and related considerations of geological CO_2 storage as a mitigating measure. In the present study effect data from literature were collected in order to conduct a marine ecological risk assessment of elevated CO_2 levels, using a Species Sensitivity Distribution (SSD). It became evident that information currently available from the literature is mostly insufficient for such a quantitative approach. Most studies focus on effects of expected future CO_2 levels, testing only one or two elevated concentrations. A full dose-response relationship, a uniform measure of exposure, and standardised test protocols are essential for conducting a proper quantitative risk assessment of elevated CO_2 levels. Improvements are proposed to make future tests more valuable and usable for quantitative risk assessment.

3.1 Introduction

Carbon dioxide (CO_2) is a natural trace gas in the Earth's atmosphere, which is also formed by the combustion of fossil fuels. As a result of economic growth and industrialisation the atmosphere's concentration of CO_2 has grown over the last century (e.g., Wolff (2011)). As global warming is believed to be caused by rising CO_2 levels (e.g., Solomon et al. (2007)), authorities have set targets to reduce CO_2 emissions (e.g., United Nations (1998)). In order to achieve this goal, one of the solutions that is being considered (and in some cases already applied), is the capture and geological storage of CO_2, in for instance abandoned oil or gas reservoirs (Steeneveldt et al., 2006).

When stored sub-seabed, there is a risk, albeit small, that stored CO_2 is accidentally released into the aquatic environment. Some authors argue that when storage options other than depleted oil and gas fields are used, such as aquifers and coal seams, it may not be guaranteed that they retain integrity forever (Zwaan and Gerlagh, 2009; Zwaan and Smekens, 2009).

Leakage from artificial storage, whilst unlikely at well-planned and managed sites, could be in the form of sudden large releases. More likely it will involve seepage of small amounts of CO_2 over time (Zwaan and Smekens, 2009), which might result in locally elevated CO_2 levels. Quantitative risk assessment of elevated CO_2 levels on marine ecology, resulting from either increased air emission or accidental releases from storage, should be an important aspect in the license application process on geological storage as required by legislation (e.g., EC (2009)). However, such an assessment is currently unavailable.

Nonetheless, (physiological) effects of CO_2 on marine species are often studied, thus a great deal is known about potential effects of elevated CO_2 levels on these species. Shifts in pH as a result of elevated CO_2 levels are identified as an important factor resulting in physiological effects, particularly, for species that form calcareous tissues, such as corals (Hoegh-Guldberg, 2005). Kikkawa et al. (2004) indicate that the effects of water acidification by mineral acids such as hydrochloric and sulphuric acid are less than those caused by high CO_2 levels, when tested at the same water pH, as demonstrated in their study on eggs and larvae of red seabream (*Pagrus major*). Ishimatsu et al. (2005) indicate that this could very well be the case for other species as well. The latter was confirmed for Japanese flounder (*Paralichthys olivaceus*) (Hayashi et al., 2004), which supports the suggestion that exposure levels should be expressed as CO_2 levels, rather than a shift in pH units. CO_2 solubility in the water phase exceeds oxygen solubility which can reverse the normal outward diffusion of CO_2 from fish if CO_2 water concentrations are elevated (Ishimatsu et al., 2005).

A quantitative evaluation of median lethal CO_2 levels (LC50s) has rarely been conducted, but it appears that reported effect levels can vary largely, even within taxonomic groups like fish, as reviewed by Ishimatsu et al. (2005). Pörtner et al.

(2005) note in their review that, although acute and chronic as well as lethal and sub-lethal effects of CO_2 have been studied, the continuum between time- and concentration-dependent effects have not been studied. As a result critical thresholds limiting long-term survival cannot be determined.

A widely used technique for ecological risk assessment of toxicants is the Species Sensitivity Distribution (SSD) (Newman et al., 2000; Posthuma et al., 2002), which has recently been applied to non-toxic stressors as well (De Vries et al., 2008; Smit et al., 2008; Struijs et al., 2011). The technique has been extensively discussed and validated in ecotoxicology (e.g., Forbes and Forbes (1993); Forbes et al. (2001); Hose and Van den Brink (2004); Selck et al. (2002); Van Wijngaarden et al. (2005)). Basically, the SSD is the statistical distribution of species sensitivity, usually expressed as chronic no observed effect concentrations (NOECs) for a specific toxic compound for several representative species. An SSD can both be used to derive predicted no effect concentrations (PNECs) and to estimate the Potentially Affected Fraction (PAF) of species at risk posed by a specific exposure level.

For animals CO_2 can be considered as a toxicant, as it exerts adverse effects as a function of test species conditions, exposure duration and concentration. Specific issue for CO_2 is the complex carbonate chemistry which determines the exposure level and the fact that CO_2 is essential in respiratory pathways. Organisms have mechanisms to deal with CO_2, but this is also the case for toxic metals that are essential elements at low concentrations (e.g., (Goldhaber, 2003)). In addition, many toxicants also display complex chemistry affecting their availability and hence toxicity (e.g., Di Toro et al. (2001)).

In the present study, marine aquatic CO_2 effect data were collected in order to construct an SSD for quantitative risk assessment of elevated CO_2 levels in marine ecosystems. In addition to effect levels, information about experimental conditions and quality of reported data was collected as well, in order to perform a meta-analysis to assist the interpretation of the constructed SSD.

3.2 Methods

3.2.1 Carbonate chemistry

Carbon dioxide has a number of chemical species in the water phase (CO_2(aq), HCO_3^-, CO_3^{2-}, where the anions can be bound to numerous cations). A commonly used metric to denote CO_2 exposure is the partial pressure (pCO_2). However, not all collected studies have used the same carbon species or unit to express the exposure level. In the present study the so-called Seacarb model (Lavigne and Gattuso, 2011) was used to calculate missing carbon species for all experiments (where possible) and used it to express all exposures as pCO_2 in micro atmosphere (μatm).

The Seacarb model uses temperature and salinity as input data. For salinity a default value of 35 ppt was used when data were missing. If experimental temperature was not reported, it was assumed to be close to the test species optimum. For all dissociation and stability constants, the default values were used as provided by the model. In addition, a combination of any two CO_2–related parameters (pH, total alkalinity, concentration HCO_3^-, total dissolved inorganic carbon or pCO_2) is required as input. Preferably, the parameters were used as measured in the experiment. Otherwise, the parameters as calculated by the authors of the original paper were used. When partial pressure was reported as percentage or ppm, the total pressure was assumed to be standard (0.987 atm, McNaught and Wilkinson (1997)), in order to convert the partial pressure into μatm. When partial pressure was reported in kPa, the pressure was converted into μatm using a conversion factor of $9.87 \cdot 10^3$ μatm/kPa (Thompson and Taylor, 2008). When reported in Torr, a conversion factor of $1.32 \cdot 10^3$ μatm/Torr (Thompson and Taylor, 2008) was used. When the pCO_2 level in the control experiment was neither reported, nor calculable, the median level of the controls of all other experiments was used as a default. Default values were used in the construction of the SSD but were not included in statistical analyses.

3.2.2 Data collection

Using several search engines (including Scopus and Google Scholar) a search was performed for effects of elevated CO_2 conditions. Although non-exhaustive, available "grey" literature also was included in the dataset. In an SSD, each unique species is represented only once and several options exist to include multiple data for a single species (Wheeler et al., 2002). In the present study each unique species is recorded once in the dataset and when multiple studies on a single species were available peer reviewed literature was preferred over "grey" literature. Further, studies that tested a concentration range were preferred over studies testing only a single concentration and studies describing all test conditions were preferred over studies poorly describing them. If none of these criteria could be applied, the study with the lowest effect level was selected.

For each record (species), the following data were included in the dataset (if available): taxonomical information on the species; data required to calculate exposure levels, (see 'carbonate chemistry' section) for both control and treatment conditions; additional experimental conditions such as exposure duration, aeration/oxygen content and the number of concentrations tested next to the control.

Likewise, it was recorded whether the effect level was either a NOEC, Lowest Observed Effect Concentration (LOEC) or median effect concentration (EC50 or LC50). Most studies only indicated whether a significant effect (or not) was observed at specific exposure concentrations, when compared to the control ex-

periment. When no EC50 was available, the statistics from those reports were used to classify effect concentrations as either a NOEC or a LOEC. As a consequence, in case only a single concentration was tested, it was either a NOEC or a LOEC, depending on whether a significant effect was observed. A LOEC was only included in the dataset if neither a NOEC nor an EC50 was available. All effect types (e.g., mortality, reproductive success, calcification rate, etc.) and parameters (EC50, NOEC and LOEC) were used in the construction of the SSD.

3.2.3 Data subselection

For discussion purposes, a second SSD was constructed with a subselection of the data. This subselection was partly created using an indicative reliability score based upon the classification scheme proposed by Klimisch et al. (1997). Although the scheme applies to (eco)toxicological data, it can be translated to CO_2 effect data. Klimisch et al. (1997) differentiated between four classes. The first class, 'reliable without restrictions' (Klimisch et al., 1997), contains data that originate from well documented experiments that were performed according to (internationally) accepted guidelines. As such guidelines are not available for CO_2 exposures, CO_2 effect data couldn't be classified as such. In the second class, 'reliable with restrictions', data originate from experiments that were not performed under standard conditions, but are at least well documented and scientifically acceptable (Klimisch et al., 1997). The third class, 'not reliable', consisted of data from studies that were either not performed properly, or not sufficiently documented. In the present study, data were classified in this third class, when two or more experimental conditions (for instance, the pH level, information on aeration, oxygen levels or test medium type) were not reported. Otherwise, data were assigned to the second class.

The fourth class, 'not assignable', were studies for which insufficient information is available for classification, for instance, when it originates from a short abstract (Klimisch et al., 1997). In the present study the fourth class also is applied to indirect citations.

The dataset was first restricted to data from studies that were reliable (a score of two or better). In addition, the dataset was further restricted to those based upon at least three test concentrations. The applied selection criteria served as a proxy for data quality. In the remainder of this document, we refer to this subselection as the 'restricted dataset'.

3.2.4 Species sensitivity distribution and uncertainty

The SSD was constructed using a non-parametric bootstrapping technique (e.g., (Grist et al., 2002)). For this purpose, an empirical distribution function was generated from the data using Hazen plotting positions (Cunnane, 1978). Between plotting positions linear interpolation was applied, after log-transformation of the

CO_2 exposure level. In common bootstrapping, an observed or collected dataset is resampled, with replacement, a large number of times (in the present study 100,000 times). Each element in the original sample is assumed fixed in that case. However, there was considerable uncertainty in the underpinning data as well. Hence, in the present study, in each bootstrap sample, the data were assumed not to be fixed, but rather was randomly sampled from a statistical distribution describing its uncertainty. How this distribution for each record was obtained, is described below. Variation between the generated pseudo-samples (2.5%, 50% and 97.5% percentiles) was used to estimate the uncertainty in the constructed SSD.

For each data element a minimum and maximum value was derived between which a true conservative end-point (such as a NOEC or EC10) is expected to fall. When only a LOEC value was available, the conservative NOEC ($NOEC_C$) should be somewhere between this LOEC and the CO_2 concentration in the control experiment. When both a NOEC and a LOEC were available, the $NOEC_C$ was assumed to be between those two values. When only a NOEC was available, the $NOEC_C$ was assumed to fall between this NOEC and 1.4 times the same NOEC. The factor of 1.4 was based on the observed ratio between the collected LOECs and NOECs, which was 1.4 or higher. When only an EC50 was available, the $NOEC_C$ was assumed to fall between this EC50 and the EC50 divided by 1.4. These ranges reflect the majority of the uncertainty, estimated based on the available information in the dataset.

A cumulative distribution function, describing the likelihood of the value of the $NOEC_C$, for each species was now defined as follows: the median was assumed to lie at the geometric mean of the minimum and maximum value. We assumed it to be 99% certain that the $NOEC_C$ was between the derived minimum and maximum and log-normally distributed. These distributions were used in the bootstrapping procedure as described above.

3.2.5 Statistical analysis

Variance in the collected CO_2 effect level was analysed by applying a two-way ANOVA. Factors included in the analysis were the experimental temperature, phylum, the CO_2 level in the control experiments, exposure duration and salinity; where the latter three were log-transformed before analysis in order to normalise the data. An ANOVA analysis was also performed with phylum as the only factor, followed by Tukey's HSD test. Statistical analyses were not performed with the restricted dataset, due to the limited sample size.

3.3 Results

The composed dataset (Appendix 3.A) contained 67 records (implying data on 67 species). Ten records were from non-peer reviewed literature, five records were indirect cites, whereas the remaining 62 were direct. The set contained data on 11 different phyla. For four of the phyla only a single record was available (Foraminifera, Nemertea, Rhodophyta and Sipuncula), for the remaining seven phyla data on multiple species were available. Only for ubiquitous marine worm *Sipunculus nudus* the required experimental temperature was unavailable, therefore a default value of 20 °C was used. Most of the data (49 of the 67) were NOECs or LOECs based upon no more than three test concentrations. Only two NOECs in the set were based upon five test concentrations. The remaining 16 records were LC50 values, of which only six had a reliability score of 2. Hence, the severity of effects may differ among the data points, as different effect parameters (LC50, NOEC and LOEC) were included and they were mostly based upon limited number of test concentrations.

The restricted dataset consisted of nine species from four different phyla (Arthropoda, Chordata, Echinodermata and Nemertea). There were four LC50 values in the restricted set and five NOEC values.

Overall more than half (36 of the 67 records) had a reliability score of 2, indicating that they were well documented. A substantial fraction (21 out of the 67) of the data were based on experimental work that was not completely documented hence assigned unreliable (reliability score of 3). The reliability of the remaining ten records was scored 4 (not assignable).

Although Haptophyta and Heterokontophyta were affected at low CO_2 levels (Fig. 3.1), their sensitivity was not significantly different from other phyla. Cnidaria, Echinodermate and Mollusca were all significantly more sensitive than Chordata (Tukey's HSD test, $p < 0.05$). All other deviations were not significant. Considerable variation was observed for reported test conditions, specifically, CO_2 level, alkalinity, pH, temperature, salinity and exposure duration (Table 3.1). CO_2 levels in the controls ranged from minimum to maximum by a factor of five, whereas levels in the treatments ranged nearly five hundredfold (Table 3.1). There was a considerable overlap between the control levels in some studies and treatment levels in others (21% overlap of the probability density in the histogram shown in Fig. 3.2).

The factors included in the ANOVA analysis explained 77% of the variation in the effect level of CO_2 (Table 3.2). Most variation could be attributed to the phyla (50%). A significant part of the variation was explained by the CO_2 level in the control experiments (12%) and the exposure duration of the experiments (12%). Salinity had no noteworthy effect on the variation of the CO_2 effect levels.

Because of the large variation of the CO_2 level in the control experiments and the overlap with treatment levels (Fig. 3.2), the exposure level for the SSD

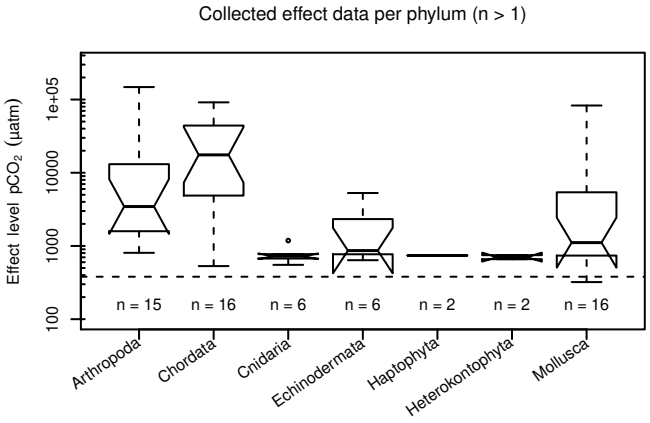


Fig. 3.1: Boxplot of CO_2 effect levels (LC50, NOEC and LOEC) per phylum. Boxes indicate first and third quartile, bold line indicates median. Whiskers indicate minimum and maximum (excluding outliers, which are shown as markers). Notches give a graphical indication of significant differences. Only phyla with data for more than one species are included in the plot (n = number of species). Cnideria, Echinodermata and Mollusca are significantly different from Chordata (Tukey's HSD test, $p < 0.05$), other differences are not significant. Dashed horizontal line indicates the median CO_2 level of all control experiments (381 μatm).

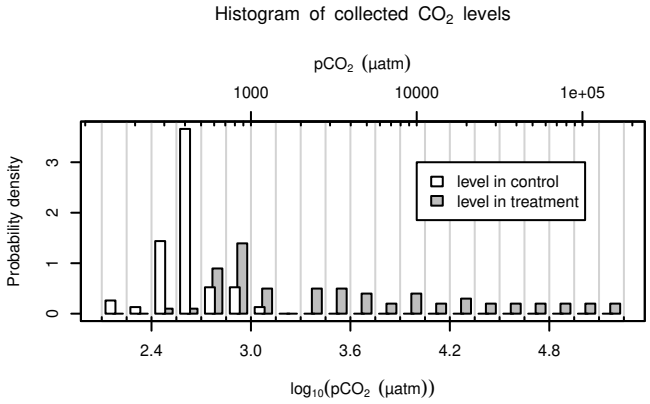


Fig. 3.2: Distribution of reported control and treatment (LC50, NOEC and LOEC) CO_2 levels. Vertical lines indicate histogram intervals.

Table 3.1: Ranges of test conditions in both the control and the treatment groups of the collected data. N is the number of records (out of the total of 67), of which N_{sc} records were estimated with the Seacarb model, 'min' is the minimum value, 'med' is the median value, 'max' the maximum value.

Test condition	Control					Treatment				
	N	N_{sc}	Min	Med	Max	N	N_{sc}	Min	Med	Max
Total alkalinity (mmol/kg)[6]	51	23	1.32	2.34	3.60	54	28	1.77	2.37	5.24
pH[6]	56	0	7.75	8.07	8.39	54	5	5.73	7.68	8.10
pCO2[6] (μatm)	51	35	166	381	1,086	67	30	321	2,359	148,039
Salinity (ppt)[7]	39	–	28.0	34.0	38.1	39	–	28.0	34.0	38.1
Temperature (°C)[7]	61	–	-0.5	17.5	30.0	61	–	-0.5	17.4	30.0
Exposure duration (days)[7]	–	–	–	–	–	65	–	0.07	8.0	365

[6] Contains values both as reported in the original publications and Seacarb model (Lavigne and Gattuso, 2011) estimates.
[7] Contains solely values as reported by the original publications.

Table 3.2: Two-way ANOVA of a linear model of pCO_2 (log-transformed) in the treatment versus experimental conditions. 28 records were omitted in the analysis due to missing values.

Variable	Degrees of freedom	Sum of squares (% of total)	F value	$Pr(>F)$
pCO_2 in control (μatm)[8,9]	1	1.67 (12%)	11.6	0.002
Temperature (°C)[9]	1	0.37 (3%)	2.53	0.13
Exposure duration (days)[8]	1	1.76 (12%)	12.2	0.002
Salinity (ppt)[8,9]	1	0.06 (0%)	0.44	0.52
Phylum	7	7.09 (50%)	7.02	$2 \cdot 10^{-4}$
Residuals	23	3.32 (23%)		

was expressed as a percentage relative to the control CO_2 levels that were set at 100% (Fig. 3.3).

The SSD based on the restricted dataset (Fig. 3.3B) is less conservative, when compared to the SSD based on the entire dataset (Fig. 3.3A). However, this is only true for the median of the bootstrap model. The SSD confidence intervals for the restricted dataset (Fig. 3.3B) are much wider than those based on the complete dataset (Fig. 3.3A).

3.4 Discussion

In the present study, CO_2 effect data of 67 aquatic species were collected, in order to construct an SSD for quantitative ecological risk assessment of elevated CO_2 levels. The exposure concentrations were all expressed in μatm and information about uncertainty in the effect data and experimental conditions were also included in the dataset, in order to assess the confidence of the constructed SSD and restriction that may apply.

3.4.1 Patterns among phyla

Enhanced CO_2 concentrations cause reduced pH and carbonate ion concentrations, and thus the level of calcium carbonate saturation. Calcifying organisms, such as corals and some mollusc and planktonic species, will have difficulty maintaining their external calcium carbonate skeletons under these conditions (Orr et al., 2005). This explains why the phylum of Cnidaria (in our case only corals) was most sensitive (Fig. 3.1).

[8]Log-transformed before analysis.

[9]Default values were not included in the analysis.

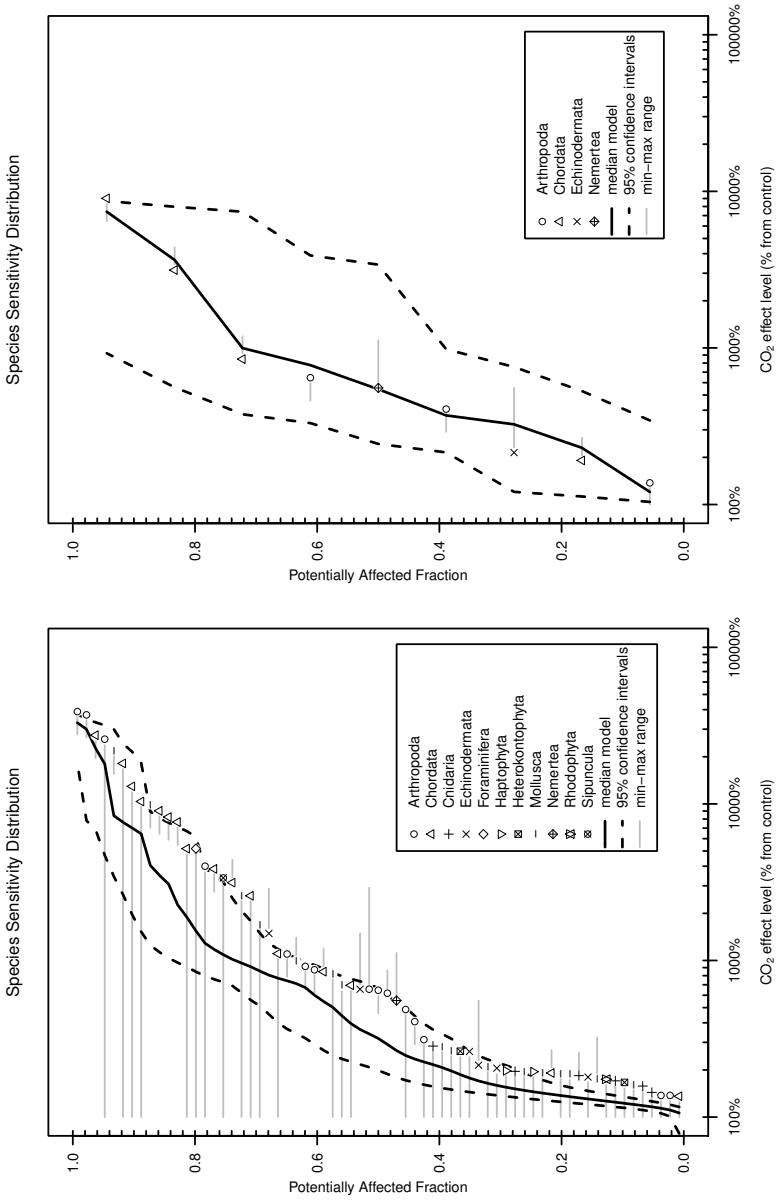


Fig. 3.3: Species Sensitivity Distributions of CO_2 effect levels expressed as a percentage of the CO_2 level in the control (set at 100%). Black lines show the median and 95% confidence interval of the bootstrapping model. Horizontal grey lines indicate the expected minimum and maximum range of 'true' conservative end-points (as explained in the method section of the main text). The SSDs are based upon the complete dataset (A) and the restricted dataset (B).

Echinodermata (e.g., sea urchins and sea stars) also appeared to be a sensitive phylum (Fig. 3.1). The species tested from this phylum were mostly sea urchins, which possess a calcareous skeleton and spikes which makes them vulnerable to effects on calcification.

Most Mollusca (e.g., mussels and snails) possess calcareous shells, and they were affected by a wide range of CO_2 levels. The mollusc median CO_2 effect levels were 16 times lower than those of Chordata, which was consistent with the observation by Melzner et al. (2009) who found that adult marine ectothermic Chordata (e.g. wolffish, salmon, Atlantic cod) are the most tolerant when exposed to chronic elevated CO_2 conditions, whereas invertebrates were generally less tolerant.

Positive effects on growth in algal species (Heterokontophyta) were included in the present analysis for the sake of completeness. The positive effect is probably due to increased availability of CO_2 for their respiration. Whether this effect should be considered positive or adverse for the environment is perhaps a more philosophical discussion that was outside the scope of the present study.

3.4.2 Quality of the data underpinning the SSD

Obviously, when data are to be used in a risk assessment, it is essential that they are of good quality. Guidelines exist for experimental work with CO_2. For instance Riebesell et al. (2010) provided excellent guidelines for dealing with practical and theoretical issues regarding marine CO_2 tests. The subject of the present study on quantitative risk assessment, however, was not addressed in these guidelines.

For probabilistic ecotoxicological risk assessment guidelines are available, including minimum requirements for quantity and quality of the underpinning data. The European Chemicals Agency (ECHA) has provided some generic criteria for the SSD approach in risk assessment of toxicants (ECHA, 2008a), such as the number of taxonomical groups and the total number of species to be included. These criteria can also be applied to non-toxic stressors. The same goes for the ECHA guidelines for the quality of laboratory data underpinning risk assessment (ECHA, 2008b). Below we discuss the quality of the CO_2 effect data collected in the present study following the two important aspects distinguished by ECHA: relevance and reliability.

Relevance

The relevance of performed experiments for the purpose of risk assessment was determined by ECHA (ECHA, 2008b) based on the following five criteria: (1) substance tested was representative; (2) appropriate species were tested; (3) appropriate route of exposure was tested; (4) appropriate doses were tested;

and (5) all critical parameters affecting end-points were considered. Each of these aspects is discussed below.

1. In the present study, only effect data of elevated CO_2 levels have been collected. Although this was the relevant substance for the present study, CO_2 affects the carbonate chemistry hence many physical characteristics of the aquatic compartment. The partial CO_2 pressure was used as proxy for the affected parameters. Whether this is the best proxy requires further study, but in the presently collected set of data it was the most frequently reported metric.

2. Many of the reviewed studies focused on species that were assumed to be sensitive for elevated CO_2 levels; almost a quarter of the dataset was composed of corals, coralline algae and sea urchins. This means that the species included in the SSD may not be fully representative for the marine ecosystem, although they were diverse.

   Furthermore, the end-points studied or modes of action for CO_2 are diverse amongst taxa (e.g., effects on calcification rate, enzymatic activity, survival, fertilisation success, etc.) and result in sensitive subgroups of species which may be selected for a more conservative risk assessment. Similar approaches have been proposed for toxicants with specific target modes of action (e.g., insecticides; Maltby et al. (2005)). In our dataset, however, no systematic difference was found between organisms with and without obvious calciferous structures (data not shown). It cannot be excluded that this becomes different when more standardised conditions are used with longer exposure durations.

   Life-stage of the test species can also affect their sensitivity towards CO_2 exposure (Kroeker et al., 2010, 2011). Information on species life-stage was not included in the present analysis. After including this factor in the ANOVA analysis, this did not reveal any relevant effect and contributed less to the variation in sensitivity than other factors (data not shown).

   Further evaluation of the species composition in the SSD is needed when it is to be applied for environmental risk assessment in a specific scenario.

3. In most of the retrieved studies, species were exposed to CO_2 by constantly bubbling with CO_2 enriched air, whereas the control was bubbled with ambient air. Whether this was the most appropriate exposure route also depends on the scenario for which the risk needs to be assessed. When pure CO_2 is released, this might lead to oxygen displacement, resulting in hypoxic effects. This was not reflected by the exposure route described before (bubbling with CO_2 enriched air). Most experiments also used a sudden transfer of test species from normal to elevated CO_2 levels, while in more realistic scenarios the transfer is probably more gradual. It

was shown, for instance, by Kikkawa et al. (2006) that juvenile Japanese sillago mortality at elevated CO_2 levels was much lower when levels were increased stepwise. The sudden transfer could be used in risk assessment to represent a worst case exposure scenario.

4. In toxicology, appropriate doses usually are determined first with a range finding test, after which a full dose-response curve is produced. This was not a common approach in CO_2 effect studies, as in more than half of the cases only one or two test concentrations were used in addition to the control level. These concentrations were usually based upon specific hypothetical future (or in some cases historical) CO_2 levels. The results indicated whether effects could be expected at those specific hypothetical levels, but did not provide information to quantify effects of other levels, which is required for quantitative risk assessment.

   The limited number of test concentrations even hampered the classification of test concentrations as either NOEC (the highest concentration that does not cause an effect significantly different from the control) or LOEC (the lowest concentration that does cause such an effect). Only a 'yes or no' occurrence of an effect could be determined, when only a single concentration was tested. For species for which both a LOEC and a NOEC was available, the ratio between the two ranged from 1.4 up to 40, with a median value of 2.0.

5. Critical parameters that affect the end-points should be considered, in order to assess the relevance of the data in risk assessment (ECHA, 2008b). In the present study we considered a few critical parameters: (a) CO_2 level in the control experiment; (b) temperature; (c) salinity; and (d) exposure duration. As noted before, the experimental conditions were highly variable (Table 3.1) and not always presented. Each of these parameters and their impact on the outcome of the risk assessment are discussed below.

   (a) Not only was the variation of the CO_2 level in the control experiments large (mainly due to variation in ambient levels), the levels also overlapped with those in the treatment (Fig. 3.2). Furthermore, the CO_2 level in the control experiments significantly contributed to the variation of the CO_2 level in the treatment (Table 3.2), which complicated the interpretation of the SSD (Fig. 3.3), as the level in the treatment was expressed relative to the level in the control. This suggests that acclimatisation may affect the observed effect, assuming that the level in the control was also the acclimatisation level. Another explanation could be that as most experiments only used one or two test concentrations, the selection of those levels may have been biased by the ambient CO_2 level used in the control. In any case, the level of

CO_2 under control conditions is an important aspect to consider in quantitative risk assessment.

(b) It was expected that temperature would be a relevant parameter, as it affects both chemical equilibria and physiological processes. However, it had a marginal and insignificant effect on the variation of the CO_2 effect levels (Table 3.2), even though the variation in test temperature is large, ranging from -0.5 up to 30 °C (Table 3.2). This indicates that temperature was not as important as expected. However, the effect of temperature may have been masked by other (unstudied) sources of variance.

(c) Salinity ranged from 28 up to 38 ppt in the reviewed experiments (Table 3.1). In the majority of the studies (56%) salinity varied between 33 and 35 ppt which are typical oceanic values (National Oceanographic Data Center, 2009). Salinity hardly contributed to the variation in the CO_2 effect levels (Table 3.2). It is therefore not considered an important factor in the ecological risk assessment of elevated CO_2 levels.

(d) The exposure duration ranged from a few hours up to a full year, and had a clear effect on the effect levels of CO_2 (Table 3.2). As with toxic compounds, it is to be expected that with longer exposure duration effect levels are lower. In the dataset used, both non-chronic as well as chronic effect data are included, but these should ideally be used in separate SSDs.

It can be concluded that the relevance of the presently collected data and the test conditions used are very diverse, mostly because no clear guidelines for CO_2 experiments in the context of risk assessment exist. For specific applicability of the SSD end-points should be selected that are relevant for the effects of CO_2 (e.g., Hendriks et al. (2010)). This requires further study and the currently available data did not allow selection of specific end-points.

Reliability

According to ECHA, the reliability of data should be assessed using the classification scheme as proposed by Klimisch et al. (1997). An indicative classification based on that study has been applied in the present case. No data could be classified as 'reliable without restrictions' as no (internationally) accepted guidelines are available for experiments for risk assessment of elevated CO_2 levels.

Not all data, presently collected, were reliable enough to fall in the classes with or without restrictions and therefore were not suitable for quantitative risk assessment. These data were tentatively included in the present study in order to get an overview that was as complete as possible. Unfortunately, Klimisch et al. (1997) did not specify which restrictions apply when data are classified as

reliable with restrictions. Those restrictions could be based on the relevance of the data as discussed above.

3.4.3 Uncertainty in SSD

The distribution of the effect data (Fig. 3.3) was clearly skewed (asymmetrical). The skewedness of the distribution could be partly explained by the fact that partial CO_2 pressure is scaled from 0 up to the total atmospheric pressure (0.986 atm under standard conditions), rather than up to infinity. In addition, many of the underpinning studies only tested a single elevated CO_2 concentration (based upon future predictions) instead of a full dose-response curve, which may also contribute to the asymmetry of the curve.

We determined the uncertainty in the SSD using a non-parametric bootstrapping technique. Even with the sample size used here (N = 67) bootstrapping techniques may not generate accurate confidence intervals (Van der Hoeven, 2001). However, in the present study the bootstrapping was not used to derive a valid PNEC, but to get an indication of uncertainty in the risk curve.

Uncertainty in the SSD, introduced as a consequence of the use of the Seacarb model to calculate effect levels, were not evaluated in this paper. The uncertainty introduced by the Seacarb model was expected to be outweighed by the uncertainty resulting from the lack of test concentrations, which ranged from a factor 1.4 up to 259. As some sources of uncertainty (e.g., lack of reliability) were difficult to quantify and therefore not included, the actual uncertainty in the SSD probably was larger than derived here (Fig. 3.3).

In order to assess ecological risks, the SSD curves need to be evaluated against exposure levels, such as for example described for several scenarios by Blackford et al. (2008). However, given the limitations as described above, such an evaluation is not performed in the present study and we restrict ourselves to providing an indication of 5% hazardous concentrations (HC5), which is often applied as a 'safe' threshold level in ecotoxicology. This was 116% (108% – 125%, 95% confidence interval) CO_2 compared to the control set at 100% based on the complete dataset (Fig. 3.3A). After applying selection criteria, insufficient data remained to be able to derive an HC5 using the bootstrapping technique (Fig. 3.3B). The HC50 level determined for the complete dataset was 318% (198%–663%, 95% confidence intervals, Fig. 3.3A), and for the restricted dataset 545% (243%–3,414%, 95% confidence intervals, Fig. 3.3B). Although the latter option used better (but still not ideal) quality data, the confidence intervals at the HC50 level were much wider as a result of the smaller sample size after restriction of the data.

3.5 Conclusions and recommendations

From the collected CO_2 effect data explicable effects could be extracted for sensitivity of phyla and test conditions. Unfortunately, the number of test concentrations often was too limited to properly quantify a (no) effect level. In addition, the experimental conditions were highly variable and not always chronic. The resulting uncertainty in the SSD derived in the present study for exposure of marine ecosystems to CO_2, makes the application for ecological risk assessment indicative at best.

Quantitative ecological risk assessment of elevated CO_2 exposure would benefit from internationally accepted standardised guidelines which improve the relevance and reliability of the experiments.

Such guidelines could make use of already developed guidelines (e.g., OECD Guidelines for testing toxic compounds, European Chemicals Agency, 2008a, b; Riebesell et al. (2010)), and include a definition of a proper test concentration range, a narrow range of CO_2 levels in the control experiment and (realistic) acclimatisation levels.

Acknowledgments

Research presented in this document was funded by CATO-2 (the Dutch national research and development program for CO_2 capture, transport and storage) and RISCS (funded by the EC 7th Framework Programme and by industry partners ENEL I&I, Statoil, Vattenfall AB, E.ON and RWE), which is gratefully acknowledged. We are also pleased to acknowledge Geke L. den Otter for assisting with proof-reading of the manuscript. Suggestions by the editor and anonymous reviewers are much appreciated.

References

Blackford, J. C., Jones, N., Proctor, R., Holt, J., 2008. Regional scale impacts of distinct CO_2 additions in the North Sea. Mar Pollut Bull 56 (8), 1461–1468.

Cunnane, C., 1978. Unbiased plotting positions – a review. J Hydrol 37 (3-4), 205–222.

De Vries, P., Tamis, J. E., Murk, A. J., Smit, M. G. D., 2008. Development and application of a species sensitivity distribution for temperature-induced mortality in the aquatic environment. Environ Toxicol Chem 27 (12), 2591–2598.

Di Toro, D. M., Allen, H. E., Bergman, H. L., Meyer, J. S., Paquin, P. R., Santore, R. C., 2001. Biotic ligand model of the acute toxicity of metals. 1. technical basis. Environ Toxicol Chem 20 (10), 2383–2396.

EC, 2009. Directive 2009/31/EC of the European Parliament and of the Council of 23 April 2009 on the geological storage of carbon dioxide and amending Council Directive 85/337/EEC, European Parliament and Council Directives 2000/60/EC, 2001/80/EC, 2004/35/EC, 2006/12/EC, 2008/1/EC and Regulation (EC) No 1013/2006. Off J Eur Union L140, 114–135.

ECHA, 2008a. Guidance on information requirements and chemical safety assessment chapter R.10: Characterisation of dose [concentration]-response for environment.

ECHA, 2008b. Guidance on information requirements and chemical safety assessment chapter R.4: Evaluation of available information.

Forbes, T. L., Forbes, V. E., 1993. A critique of the use of distribution-based extrapolation models in ecotoxicology. Funct Ecol 7, 249–254.

Forbes, V. E., Calow, P., Sibly, R. M., 2001. Are current species extrapolation models a good basis for ecological risk assessment? Environ Toxicol Chem 20 (2), 442–447.

Goldhaber, S. B., 2003. Trace element risk assessment: essentiality vs. toxicity. Regul Toxicol Pharm 38 (2), 232–242.

Grist, E. P. M., Leung, K. M. Y., Wheeler, J. R., Crane, M., 2002. Better bootstrap estimation of hazardous concentration thresholds for aquatic assemblages. Environ Toxicol Chem 21 (7), 1515–1524.

Hayashi, M., Kita, J., Ishimatsu, A., 2004. Comparison of the acid-base responses to CO_2 and acidification in Japanese flounder (*Paralichthys olivaceus*). Mar Pollut Bull 49 (11-12), 1062–1065.

Hendriks, I. E., Duarte, C. M., Álvarez, M., 2010. Vulnerability of marine biodiversity to ocean acidification: A meta-analysis. Estuar Coast Shelf S 86 (2), 157–164.

Hoegh-Guldberg, O., 2005. Low coral cover in a high-CO_2 world. J Geophys Res 110, C09S06.

Hose, G. C., Van den Brink, P. J., 2004. Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. Arch Environ Con Tox 47 (4), 511–520.

Ishimatsu, A., Hayashi, M., Lee, K.-S., Kikkawa, T., Kita, J., 2005. Physiological effects on fishes in a high-CO_2 world. J Geophys Res 110, C09S06.

Kikkawa, T., Kita, J., Ishimatsu, A., 2004. Comparison of the lethal effect of CO_2 and acidification on red sea bream (*Pagrus major*) during the early developmental stages. Mar Pollut Bull 48 (1-2), 108–110.

Kikkawa, T., Sato, T., Kita, J., Ishimatsu, A., 2006. Acute toxicity of temporally vary-

ing seawater CO_2 conditions on juveniles of Japanese sillago (*Sillago japonica*). Mar Pollut Bull 52 (6), 621–625.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharm 25 (1), 1–5.

Kroeker, K. J., Kordas, R. L., Crim, R. N., Singh, G. G., 2010. Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. Ecol Lett 13 (11), 1419–1434.

Kroeker, K. J., Kordas, R. L., Crim, R. N., Singh, G. G., 2011. Response to technical comment on 'meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms'. Ecol Lett 14 (9), E1–E2.

Lavigne, H., Gattuso, J.-P., 2011. Seawater carbonate chemistry with R. calculates parameters of the seawater carbonate system. code and/or corrections were contributed by Andreas Andersson, Jean-Marie Epitalon, Bernard Gentili, Andreas Hofmann, Jim Orr, Aurelien Proye and Karline Soetaert. version 2.4.2.

Maltby, L., Blake, N., Brock, T. C. M., Van den Brink, P. J., 2005. Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems. Environ Toxicol Chem 24 (2), 379–388.

McNaught, A. D., Wilkinson, A., 1997. Compendium of chemical terminology: IUPAC recommendations, 2nd Edition. Blackwell Science, Oxford, International Union of Pure Applied Chemistry.

Melzner, F., Gutowska, M. A., Langenbuch, M., Dupont, S., Lucassen, M., Thorndyke, M. C., Bleich, M., Pörtner, H.-O., 2009. Physiological basis for high CO_2 tolerance in marine ectothermic animals: pre-adaptation through lifestyle and ontogeny? Biogeosciences 6 (10), 2313–2331.

National Oceanographic Data Center, 2009. World ocean atlas.

Newman, M. C., Ownby, D. R., Mézin, L. C. A., Powell, D. C., Christensen, T. R. L., Lerberg, S. B., Anderson, B.-A., 2000. Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. Environ Toxicol Chem 19 (2), 508–515.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., Yool, A., 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437, 681, 686.

Pörtner, H. O., Langenbuch, M., Michaelidis, B., 2005. Synergistic effects of temperature extremes, hypoxia, and increases in CO_2 on marine animals: From earth history to global change. J Geophys Res 110, C09S10.

Posthuma, L., Suter II, G. W., Traas, T. P., 2002. Species Sensitivity Distribution in Ecotoxicology. CRC Press LLC, Boca Raton, Fl., USA.

Riebesell, U., Fabry, V., Hansson, L., Gattuso, J.-P., 2010. Guide to best practices for ocean acidification research and data reporting.

Selck, H., Riemann, B., Christoffersen, K., Forbes, V. E., Gustavson, K., Hansen, B. W., Jacobsen, J. A., Kusk, O. K., Petersen, S., 2002. Comparing sensitivity of ecotoxicological effect endpoints between laboratory and field. Ecotox Environ Safe 52 (2), 97–112.

Smit, M. G. D., Holthaus, K. I. E., Trannum, H. C., Neff, J. M., Kjeilen-Eilertsen, G., Jak, R. G., Singsaas, I., Huijbregts, M. A. J., Hendriks, A. J., 2008. Species sensitivity distributions for suspended clays, sediment burial, and grain size change in the marine environment. Environ Toxicol Chem 27 (4), 1006–1012.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., T., Miller, H. (Eds.), 2007. Climate Change 2007: The Physical Science Basis – Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Steeneveldt, R., Berger, B., Torp, T. A., 2006. CO_2 capture and storage: Closing the knowing–doing gap. Chem Eng Res Des 84 (9), 739–763.

Struijs, J., De Zwart, D., Posthuma, L., Leuven, R. S. E. W., Huijbregts, M. A. J., 2011. Field sensitivity distribution of macroinvertebrates for phosphorus in inland waters. Integrated Environ Assess Manag 7 (2), 280–286.

Thompson, A., Taylor, B. N., 2008. Guide for the Use of the International System of Units (SI)-National Institute of Standards and Technology Special Publication 811, 2008 Edition. US Government Printing Office, Washington, D.C., USA.

United Nations, 1998. Kyoto protocol to the United Nations framework convention on climate change.

Van der Hoeven, N., 2001. Estimating the 5-percentile of the species sensitivity distributions without any assumptions about the distribution. Ecotoxicology 10 (1), 25–34.

Van Wijngaarden, R. P. A., Brock, T. C. M., Van den Brink, P. J., 2005. Threshold levels for effects of insecticides in freshwater ecosystems: A review. Ecotoxicology 14 (3), 355–380.

Wheeler, J. R., Grist, E. P. M., Leung, K. M. Y., Morritt, D., Crane, M., 2002. Species

sensitivity distributions: data and model choice. Mar Pollut Bull 45 (1-12), 192–202.

Wolff, E. W., 2011. Greenhouse gases in the earth system: a palaeoclimate perspective. Phil T Roy Soc A 369 (1943), 2133–2147.

Zwaan, B., Gerlagh, R., 2009. Economics of geological CO_2 storage and leakage. Climatic Change 93 (3-4), 285–309.

Zwaan, B., Smekens, K., 2009. CO_2 capture and storage with leakage in an energy-climate model. Environ Model Assess 14 (2), 135–148.

3.A Appendix

3.A.1 Summary of data included in analysis

Table 3.3: The data analysed, including all meta-information is to extensive to present completely in this appendix. Only a summary of the CO_2 effect data analysed is presented, here. A complete set of the data is available as supplemental information to the original publication (De Vries et al., 2013). pCO_2 is in μatm.

Species name	Effect parameter	pCO_2 control	pCO_2 treatment	Reference
Arthropoda				
Acartia steueri	NOEC	355	2,329	Kurihara and Shirayama (2004)
Acartia tsuensis	NOEC	380	2,349	Kurihara and Ishimatsu (2008)
Calanus pacificus	LC50	381[10]	3,323	Sato et al. (2005)
Euchaeta marina	LC50	1,086	7,007	Watanabe et al. (2006)
Gaidius variabilis	LC50	790	1,086	Watanabe et al. (2006)
Heterostylites major	LC50	790	8,685	Watanabe et al. (2006)
Homarus gammarus	LOEC	166	806	Arnold et al. (2009)
Marsupenaeus japonicas	LC50	381[10]	141,130	Kikkawa et al. (2008)
Metamphiascopsis hirsutus	LOEC	493	19,738	Sato et al. (2005)
Metridia pacifica	LC50	381[10]	3,493	Sato et al. (2005)
Neocalanus cristatus	LC50	849	3,454.2	Watanabe et al. (2006)
Panulirus Cygnus	LOEC	381[10]	98,692	Ishimatsu et al. (2005b)
Paraeuchaeta birostrata	LC50	790	1,085.6	Watanabe et al. (2006)
Penaeus japonicas	LC50	381[10]	148,038	Ishimatsu et al. (2005a)
Semibalanus balanoides	LOEC	273	851	Findlay et al. (2010)

Continued on next page

[10] pCO_2 not reported nor calculable for this species; default value used.

Table 3.3 continued

Species name	Effect parameter	pCO_2 control	pCO_2 treatment	Reference
Chordata				
Acanthochromis polyacanthus	NOEC	292	559	Munday et al. (2011)
Amphiprion percula	LOEC	391	531	Munday et al. (2009)
Anarhichas minor	NOEC	497	15,599	Foss et al. (2003)
Anguilla anguilla	LOEC	381[10]	39,474	Cruz-Neto and Steffensen (1997)
Careproctus trachysoma	LOEC	381[10]	19,738	Ishimatsu et al. (2005b)
Dicentrarchus labrax	LC50	687	61,974	Grøttum and Sigholt (1996)
Euthynnus affinis	LC50	333	91,586	Kikkawa et al. (2003)
Gadus morhua	LOEC	406	4,487	Melzner et al. (2009)
Mustelus manazo	LOEC	381[10]	69,085	Ishimatsu et al. (2005b)
Oncorhynchus mykiss	LOEC	381[10]	9,869	MacKenzie and Perry (1997)
Pagrus major	LC50	336	12,929	Kikkawa et al. (2003)
Paralichthys olivaceus	LC50	338	27,831	Kikkawa et al. (2003)
Salmo salar	NOEC	625	5,303	Fivelstad et al. (1998)
Seriola quinqueradiata	LOEC	381[10]	49,346	Ishimatsu et al. (2005b)
Sillago japonica	LC50	331	25,364	Kikkawa et al. (2003)
Sparus aurata	LOEC	501	3,482	Michaelidis et al. (2007)
Cnideria				
Acropora digitifera	LOEC	417	1,185	Suwa et al. (2010)
Acropora palmata	LOEC	468	673	Albright et al. (2010)
Astrangia poculata	LOEC	394	772	Holcomb et al. (2010)
Cladocora caespitosa	NOEC	376	692	Rodolfo-Metalpa et al. (2010)
Lophelia pertusa	LOEC	350	552	Maier et al. (2009)
Stylophora pistillata	LOEC	444	756	Reynaud et al. (2003)

Continued on next page

Table 3.3 continued

Species name	Effect parameter	pCO_2 control	pCO_2 treatment	Reference
Echinodermata				
Echinometra mathaei	NOEC	355	2,329	Kurihara and Shirayama (2004)
Hemicentrotus pulcherrimus	NOEC	355	5,290	Kurihara and Shirayama (2004)
Pisaster ochraceus	LOEC	375	770	Gooding et al. (2009)
Sterechinus neumayeri	NOEC	425	912	Ericson et al. (2010)
Strongylocentrotus franciscanus	LOEC	355	639	Reuter et al. (2011)
Tripneustes gratilla	LOEC	312	819	Sheppard Brennand et al. (2010)
Foraminifera				
foraminifera	LOEC	381[10]	19,738	Caldeira and Akai (2005)
Haptophyta				
Emiliania huxleyi	LOEC	380	741	De Bodt et al. (2010)
Phaeocystis globosa	LOEC	375	740	Wang et al. (2010)
Heterokontophyta				
Ecklonia radiate	LOEC	396	660	Connell and Russell (2010)
Phaeodactylum tricornutum	LOEC	285	749	Wu et al. (2010)
Mollusca				
Argopecten irradians	LOEC	393	738	Talmage and Gobler (2010)
Cavolinia inflexa	LOEC	381	794	Comeau et al. (2010)
Crassostrea gigas	LOEC	433	1,143	Lannig et al. (2010)
Crassostrea virginica	LOEC	300	2,457	Beniash et al. (2010)
Haliotis kamtschatkana	LOEC	169	321	Crim et al. (2011)
Illex illecebrosus	LOEC	381[10]	6,415	Caldeira and Akai (2005)
Limacina helicina	NOEC	213	375	Lischka et al. (2011)
Littorina obtusata	LOEC	381[10]	1079	Ellis et al. (2009)

Continued on next page

Table 3.3 continued

Species name	Effect parameter	pCO_2 control	pCO_2 treatment	Reference
Mercenaria mercenaria	LOEC	378	740	Talmage and Gobler (2010)
Mytilus edulis	LOEC	451	867	Gazeau et al. (2010)
Mytilus galloprovincialis	LOEC	501	3,482	Michaelidis et al. (2005)
Octopus vulgaris	LOEC	381[10]	9,869	Ishimatsu et al. (2005b)
Saccostrea glomerata	LOEC	370	592	Parker et al. (2009)
Sepia lycidas	LC50	381[10]	82,902	Kikkawa et al. (2008)
Sepia officinalis	NOEC	455	4,559	Gutowska et al. (2008)
Sepioteuthis lessoniana	LC50	381[10]	37,503	Kikkawa et al. (2008)
Nemertea				
Parborlasis corrugatus	NOEC	425	2,359	Ericson et al. (2010)
Rhodophyta				
Lithophyllum cabiochae	LOEC	393	686	Martin and Gattuso (2009)
Sipuncula				
Sipunculus nudus	LOEC	296	9,968	Langenbuch and Pörtner (2004)

End of Table 3.3

References in Appendix 3.A

Albright, R., Mason, B., Miller, M., Langdon, C., 2010. Ocean acidification compromises recruitment success of the threatened Caribbean coral *Acropora palmata*. P Natl Acad Sci 107 (47), 20400–20404.

Arnold, K. E., Findlay, H. S., Spicer, J. I., Daniels, C. L., Boothroyd, D., 2009. Effect of CO_2-related acidification on aspects of the larval development of the European lobster, *Homarus gammarus* (L.). Biogeosciences 6 (8), 1747–1754.

Beniash, E., Ivanina, A., Lieb, N. S., Kurochkin, I., Sokolova, I. M., 2010. Elevated level of carbon dioxide affects metabolism and shell formation in oysters *Crassostrea virginica*. Mar Ecol Prog Ser 419, 95–108.

Caldeira, K., Akai, M., 2005. In: Metz, B., Davidson, O., De Coninck, H., Loos, M., Meyer, L. (Eds.), Carbon Dioxide Capture and Storage. Cambridge University Press, Cambridge, UK, Ch. 6 Ocean Storage.

Comeau, S., Gorsky, G., Alliouane, S., Gattuso, J.-P., 2010. Larvae of the pteropod *Cavolinia inflexa* exposed to aragonite undersaturation are viable but shell-less. Mar Biol 157 (10), 2341–2345.

Connell, S. D., Russell, B. D., 2010. The direct effects of increasing CO_2 and temperature on non-calcifying organisms: increasing the potential for phase shifts in kelp forests. Philos T Roy Soc B 277 (1686), 1409–1415.

Crim, R. N., Sunday, J. M., Harley, C. D. G., 2011. Elevated seawater CO_2 concentrations impair larval development and reduce larval survival in endangered northern abalone (*Haliotis kamtschatkana*). J Exp Mar Biol Ecol 400 (1-2), 272–277.

Cruz-Neto, A. P., Steffensen, J. F., 1997. The effects of acute hypoxia and hypercapnia on oxygen consumption of the freshwater European eel. J Fish Biol 50 (4), 759–769.

De Bodt, C., Van Oostende, N., Harlay, J., Sabbe, K., Chou, L., 2010. Individual and interacting effects of pCO_2 and temperature on *Emiliania huxleyi* calcification: study of the calcite production, the coccolith morphology and the coccosphere size. Biogeosciences 7 (5), 1401–1412.

De Vries, P., Tamis, J. E., Foekema, E. M., Klok, C., Murk, A. J., 2013. Towards quantitative ecological risk assessment of elevated carbon dioxide levels in the marine environment. Mar Pollut Bull 73 (2), 516–523.

Ellis, R. P., Bersey, J., Rundle, S., M., H.-S. J., Spicer, J. I., 2009. Subtle but significant effects of CO_2 acidified seawater on embryos of the intertidal snail, *Littorina obtusata*. Aquat Biol 5 (1), 41–48.

Ericson, J. A., Lamare, M. D., Morley, S. A., Barker, M. F., 2010. The response

of two ecologically important antarctic invertebrates (*Sterechinus neumayeri* and *Parborlasia corrugatus*) to reduced seawater pH: effects on fertilisation and embryonic development. Mar Biol 157 (12), 2689–2702.

Findlay, H. S., Kendall, M. A., Spicer, J. I., Widdicombe, S., 2010. Relative influences of ocean acidification and temperature on intertidal barnacle post-larvae at the northern edge of their geographic distribution. Estuar Coast Shelf S 86 (4), 675–682.

Fivelstad, S., Haavik, H., Løvik, G., Olsen, A. B., 1998. Sublethal effects and safe levels of carbon dioxide in seawater for Atlantic salmon postsmolts (*Salmo salar* l.): ion regulation and growth. Aquaculture 160 (3-4), 305–316.

Foss, A., Røsnes, B. A., Øiestad, V., 2003. Graded environmental hypercapnia in juvenile spotted wolffish (*Anarhichas minor* Olafsen): effects on growth, food conversion efficiency and nephrocalcinosis. Aquaculture 220 (1-4), 607–617.

Gazeau, F., Gattuso, J.-P., Dawber, C., Pronker, A. E., Peene, F., Peene, J., Heip, C. H. R., Middelburg, J. J., 2010. Effect of ocean acidification on the early life stages of the blue mussel *Mytilus edulis*. Biogeosciences 7 (7), 2051–2060.

Gooding, R. A., Harley, C. D. G., Tang, E., 2009. Elevated water temperature and carbon dioxide concentration increase the growth of a keystone echinoderm. P Natl Acad Sci 106 (23), 9316–9321.

Grøttum, J. A., Sigholt, T., 1996. Acute toxicity of carbon dioxide on European seabass (*Dicentrarchus labrax*): Mortality and effects on plasma ions. Comp Biochem Phys A 115 (4), 323–327.

Gutowska, M. A., Pörtner, H. O., Melzner, F., 2008. Growth and calcification in the cephalopod *Sepia officinalis* under elevated seawater pCO_2. Mar Ecol Prog Ser 373, 303–309.

Holcomb, M., McCorkle, D. C., Cohen, A. L., 2010. Long-term effects of nutrient and CO_2 enrichment on the temperate coral *Astrangia poculata* (Ellis and Solander, 1786). J Exp Mar Biol Ecol 386 (1-2), 27–33.

Ishimatsu, A., Hayashi, M., Lee, K.-S., Kikkawa, T., Kita, J., 2005a. Physiological effects on fishes in a high-CO_2 world. J Geophys Res 110.

Ishimatsu, A., Kikkawa, T., Hayashi, M., Lee, K.-S., Murata, K., Kumagai, E., Kita, J., 2005b. Acute physiological impacts of CO_2 ocean sequestration on marine animals. In: Proceedings of the 7th Conference on Greenhouse Gas Control Technologies. Vancouver, Canada.

Kikkawa, T., Ishimatsu, A., Kita, J., 2003. Acute CO_2 tolerance during the early developmental stages of four marine teleosts. Environ Toxicol 18 (6), 375–382.

Kikkawa, T., Watanabe, Y., Katayama, Y., Kita, J., Ishimatsu, A., 2008. Acute CO_2

tolerance limits of juveniles of three marine invertebrates, *Sepia lycidas*, *Sepioteuthis lessoniana*, and *Marsupenaeus japonicus*. Plank Benth Res 3 (3), 184–187.

Kurihara, H., Ishimatsu, A., 2008. Effects of high CO_2 seawater on the copepod (*Acartia tsuensis*) through all life stages and subsequent generations. Mar Pollut Bull 56 (6), 1086–1090.

Kurihara, H., Shirayama, Y., 2004. Effects of increased atmospheric CO_2 on sea urchin early development. Mar Ecol Prog Ser 274, 161–169.

Langenbuch, M., Pörtner, H. O., 2004. High sensitivity to chronically elevated CO_2 levels in a eurybathic marine sipunculid. Aquat Toxicol 70 (1), 55–61.

Lannig, G., Eilers, S., Pörtner, H. O., Sokolova, I. M., Bock, C., 2010. Impact of ocean acidification on energy metabolism of oyster, *Crassostrea gigas* – changes in metabolic pathways and thermal response. Mar Drugs 8, 2318–2339.

Lischka, S., Büdenbender, J., Boxhammer, T., Riebesell, U., 2011. Impact of ocean acidification and elevated temperatures on early juveniles of the polar shelled pteropod *Limacina helicina*: mortality, shell degradation, and shell growth. Biogeosciences 8 (4), 919–932.

MacKenzie, W. M., Perry, S. F., 1997. The effects of hypercapnia on branchial and renal calcium fluxes in the rainbow trout (*Oncorhynchus mykiss*). J Comp Physiol B 167 (1), 52–60.

Maier, C., Hegeman, J., Weinbauer, M. G., Gattuso, J.-P., 2009. Calcification of the cold-water coral *Lophelia pertusa*, under ambient and reduced pH. Biogeosciences 6 (8), 1671–1680.

Martin, S., Gattuso, J.-P., 2009. Response of mediterranean coralline algae to ocean acidification and elevated temperature. Glob Change Biol 15 (8), 2089–2100.

Melzner, F., Göbel, S., Langenbuch, M., Gutowska, M. A., Pörtner, H.-O., Lucassen, M., 2009. Swimming performance in Atlantic cod (*Gadus morhua*) following long-term (4–12 months) acclimation to elevated seawater. Aquat Toxicol 92 (1), 30–37.

Michaelidis, B., Ouzounis, C., Paleras, A., Pörtner, H. O., 2005. Effects of long-term moderate hypercapnia on acid–base balance and growth rate in marine mussels *Mytilus galloprovincialis*. Mar Ecol Prog Ser 293, 109–118.

Michaelidis, B., Spring, A., Pörtner, H. O., 2007. Effects of long-term acclimation to environmental hypercapnia on extracellular acid-base status and metabolic capacity in Mediterranean fish *Sparus aurata*. Mar Biol 150 (6), 1417–1429.

Munday, P. L., Donelson, J. M., Dixson, D. L., Endo, G. G. K., 2009. Effects of ocean

acidification on the early life history of a tropical marine fish. Philos T Roy Soc B 276 (1671), 3275–3283.

Munday, P. L., Gagliano, M., Donelson, J. M., Dixson, D. L., Thorrold, S. R., 2011. Ocean acidification does not affect the early life history development of a tropical marine fish. Mar Ecol Prog Ser 423, 211–221.

Parker, L. M., Ross, P. M., O'Conner, W. A., 2009. The effect of ocean acidification and temperature on the fertilization and embryonic development of the sydney rock oyster *Saccostrea glomerata* (Gould 1850). Glob Change Biol 15 (9), 2123–2136.

Reuter, K. E., Lotterhos, K. E., Crim, R. N., Thompson, C. A., Harley, C. D. G., 2011. Elevated pCO_2 increases sperm limitation and risk of polyspermy in the red sea urchin *Strongylocentrotus franciscanus*. Glob Change Biol 17 (1), 163–171.

Reynaud, S., Leclercq, N., Romaine-Lioud, S., Ferrier-Pagés, C., Jaubert, J., Gattuso, J.-P., 2003. Interacting effects of CO_2 partial pressure and temperature on photosynthesis and calcification in a scleractinian coral. Glob Change Biol 9 (11), 1660–1668.

Rodolfo-Metalpa, R., Martin, S., Ferrier-Pagès, C., Gattuso, J.-P., 2010. Response of the temperate coral cladocora caespitosa to mid- and long-term exposure to pCO_2 and temperature levels projected for the year 2100 AD. Biogeosciences 7 (1), 289–300.

Sato, T., Watanabe, Y., Toyota, K., Ishizaka, J., 2005. Numerical simulation of mortality of zooplankton caused by direct injection of carbon dioxide in the ocean. In: Proceedings of the Fourth Annual Conference on Carbon Capture and Sequestration DOE/NETL.

Sheppard Brennand, H., Soars, N., Dworjanyn, S. A., Davis, A. R., Byrne, M., 2010. Impact of ocean warming and ocean acidification on larval development and calcification in the sea urchin *Tripneustes gratilla*. PLoS ONE 5 (6).

Suwa, R., Nakamura, M., Morita, M., Shimada, K., Iguchi, A., Sakai, K., Suzuki, A., 2010. Effects of acidified seawater on early life stages of scleractinian corals (Genus *Acropora*). Fisheries Sci 76 (1), 93–99.

Talmage, S. C., Gobler, C. J., 2010. Effects of past, present, and future ocean carbon dioxide concentrations on the growth and survival of larval shellfish. P Natl Acad Sci 107 (40), 17246–17251.

Wang, Y., Smith, W. O., Wang, X., Li, S., 2010. Subtle biological responses to increased CO_2 concentrations by *Phaeocystis globosa* Scherffel, a harmful algal bloom species. Geophys Res Lett 37 (9).

Watanabe, Y., Yamaguchi, A., Ishida, H., Harimoto, T., Suzuki, S., Sekido, Y., Ikeda, T., Shirayama, Y., Mac Takahashi, M., Ohsumi, T., Ishizaka, J., 2006. Lethality of

increasing CO_2 levels on deep-sea copepods in the western North Pacific. J Oceanogr 62 (2), 185–196.

Wu, Y., Gao, K., Riebesell, U., 2010. CO_2-induced seawater acidification affects physiological performance of the marine diatom *Phaeodactylum tricornutum*. Biogeosciences 7 (9), 2915–2923.

Photo: Pepijn de Vries

Chapter 4

Development and application of a species sensitivity distribution for temperature-induced mortality in the aquatic environment

Reprinted from Environmental Toxicology and Chemistry 27(12), Pepijn de Vries, Jacqueline E. Tamis, Albertinka J. Murk and Mathijs G. D. Smit (pages 2591–2598, copyright (2008)), with permission from Wiley Materials

Abstract

Current European legislation has static water quality objectives for temperature effects, based on the most sensitive species. In the present study a species sensitivity distribution (SSD) for elevated temperatures is developed based on temperature sensitivity data (mortality) of 50 aquatic species. The SSD applies to risk assessment of heat discharges that are localised in space or time. As collected median lethal temperatures (LT50 values) for different species depend on the acclimation temperature, the SSD is also a function of the acclimation temperature. Data from a thermal discharge in The Netherlands are used to show the applicability of the developed SSD in environmental risk assessment. Although restrictions exist in the application of the developed SSD, it is con-

cluded that the SSD approach can be applied to assess the effects of elevated temperature. Application of the concept of SSD to temperature changes allows harmonisation of environmental risk assessment for stressors in the aquatic environment. When a synchronisation of the assessment methods is achieved, the steps to integration of risks from toxic and non-toxic stressors can be made.

4.1 Introduction

As a result of risk mitigating measures, the chemical state of many waters has improved substantially over the last 20 years (Sheahan et al., 2001). Other nonchemical stressors, however, can also affect the ecological status of water bodies. Thermal water discharges for instance, can cause environmental effects by elevating the temperature of the receiving water. These discharges can take place at power plants or other industrial plants where the surface water is used for cooling purposes. Possible effects of elevated temperatures can, for instance, be mortality of aquatic species or algal blooms, depending on the receiving water bodies.

Carter and colleagues provided a rationale for the evaluation of thermal induced biological effects caused by thermal discharges (Carter et al., 1979). In this approach the estimated risk is based on the most sensitive species only. A similar approach was used for current European legislations (EC, 2006). The current standard for waters, capable of supporting cyprinids, is that the weekly monitored temperature downstream of the emission should not be increased with more than 3 °C relative to the unaffected temperature (EC, 2006). This limit may, however, be exceeded 2% of the time (EC, 2006).

Most of the time heat discharges coincide with additional stressors such as hypoxia and toxic biocides or anti-fouling agents. Therefore, in those situations a multi-stress approach, integrating risk for those stressors in one overall indicator instead of evaluating all risks separately, would be most obvious as has been suggested before for toxic stress (De Zwart and Posthuma, 2005). As effects induced by temperature increase depend on the initial temperature and the natural tolerance of the species as well (McErlean et al., 1969), this type of risk assessment par excellence asks for an area-specific approach including all these aspects.

In the year 2013 the current legislation will be replaced by legislation based on the Water Framework Directive (EC, 2000). This new legislation will be area-specific and aims at the integration of principles for protection and sustainable use of water. To allow a more advanced risk limit (being location specific and allowing integration of these principles) in the new legislation than the current European 3 °C limit, a new risk assessment approach must become available (EC, 2006; Kerkum et al., 2004).

We suggest the use of a species sensitivity distribution (SSD) for temperature

effects, potentially in combination with toxic and non-toxic stressors to assess the effects of thermal discharges that are localised in space or time. An SSD describes the mean sensitivity and range of sensitivity among biota for a specific stressor (Aldenberg et al., 2002). This method could be used for generic risk assessment, but also allows for location-specific assessment by only incorporating local species. In the present study we introduce an SSD for temperature-induced mortality in the aquatic environment as a potential tool for risk assessment of thermal discharges. As this approach is based on acute data, it primarily reflects acute exposures of organisms that swim or drift into the heated water. Of course it implies the loss of resident species that could not sustain in the warmed water. In a simple case study the applicability of the SSD is demonstrated.

4.2 Methods

4.2.1 Determination of effect levels

In contrast to the determination of toxicity endpoints such as no-observed-effect concentration or 50% effect concentration, no standardised test protocol exists to determine exposure metrics of nontoxic stressors (Smit et al., 2008). The present study focuses on risk assessment of temperature increases, for which mortality effect levels were collected from literature. Two parameters are frequently used to express mortality effects of increased temperature, namely: the temperature causing 50% mortality after a specific exposure period (Urban, 1994), and the time it takes at a certain temperature to reach 50% mortality (McMahon et al., 1995). Both indicators are in the literature referred to as the LT50 but in the present study we use only the first category, since our interest is in the effect of temperature and not of time.

A similar but more refined endpoint, the upper incipient lethal temperature, represents the temperature at which 50% of the population theoretically could survive indefinitely (Beitinger et al., 2000; Jobling, 1981). In the present study, LT50 and incipient lethal temperature are considered equal and both are referred to as LT50.

As thermally induced mortality (LT50) depends on the acclimation temperature (T_a) (McErlean et al., 1969), the effect of a sudden temperature increase in test-systems with different water temperatures will differ as well. The temperature tolerance interval (TTI) is developed by Urban (1994) to describe the interval by which the temperature can increase above the T_a without killing more than 50% of the population. The relation between TTI and LT50 is described by Eqn. 4.1 (Urban, 1994).

$$\mathrm{TTI} = \mathrm{LT50} - T_a \tag{4.1}$$

It should be mentioned that the linearity of the relationship between TTI and T_a is somewhat artificial, since the TTI is by definition a function of T_a (Eqn. 4.1). The slope of the function deviates from -1, due to the apparent relationship between the T_a and the LT50. The relationship between T_a and TTI has been shown to be linear for a number of aquatic species (McErlean et al., 1969) (Appendix 4.A.1, Fig. 4.7). In the present study we used this relationship to quantify the TTI at a specific T_a. We consider the TTI as the parameter describing the sensitivity of a species towards a temperature change at specific T_a and is therefore used to construct the species sensitivity distribution. For each species (i), the TTI_i can be described by Eqn. 4.2. As test conditions, more specifically the T_a, may differ between species in literature, we use linear regression to derive a_i and b_i, where a_i is the slope and b_i is the intercept parameter for each species (i). This way we are no longer bound to discrete T_a values and can include all species at each T_a.

$$TTI_i = a_i T_a + b_i \tag{4.2}$$

4.2.2 Derivation of the species sensitivity distribution

When constructing an SSD, the choice of the distribution function to fit of the data and describe the distribution is arbitrary and usually based on best-fit results (Wheeler et al., 2002; Smit et al., 2001). Log-logistic and log-normal distributions often are used for toxic stressors, because effect concentrations can differ between species by several orders of magnitude (Gaddum, 1945). In the present study, with temperature intervals as effect parameters, the differences are less than one order of magnitude and, therefore, a normal (Gaussian) distribution was used which also described the effect data best (Appendix 4.A.2, Fig. 4.8). The general equation for a cumulative normal distribution is:

$$F(x) = \frac{1}{2}\left(1 + ERF\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right); \tag{4.3}$$

where ERF is the error function, x is an exposure metric, μ is the average exposure metric and σ is the standard deviation of all observed exposure metrics. For normal distribution calculations (Eqn. 4.3), the function NORMDIST (x; mean; standard deviation; cumulative) in Excel® 2003 is used (Microsoft, Redmond, WA, USA). The average and the standard deviation define the SSD as a function of the T_a. Eqns. 4.3 to 4.7 describe the derivation of the T_a dependent SSD, where N is the number of species included. For each species (i) the parameters (a_i and b_i) that describe TTI_i as a function of T_a (Eqn. 4.2) are derived. The average TTI

(μ_{TTI}) is calculated from the following equation:

$$\mu_{TTI} = \frac{1}{N}\sum_{i=1}^{N}(a_i T_a + b_i) = \frac{T_a}{N}\sum_{i=1}^{N} a_i + \frac{1}{N}\sum_{i=1}^{N} b_i \tag{4.4}$$

This can be rewritten as:

$$\mu_{TTI} = \mu_a T_a + \mu_b; \tag{4.5}$$

showing that the average TTI for a given T_a also is a function of the average regression coefficients a_i and b_i from Eqn. 4.2 (μ_a and μ_b respectively). The standard deviation of the TTI of all included species (σ_{TTI}) is written as a function of T_a as well based on the derivation given in Appendix 4.A.3:

$$\sigma_{TTI} = \sqrt{T_a^2\sigma_a^2 + \sigma_b^2 + c\,T_a}; \tag{4.6}$$

where c is defined as:

$$c = \frac{2}{N-1}\sum_{i=1}^{N}(a_i - \mu_a)(b_i - \mu_b) \tag{4.7}$$

These equations show that with the σ_{TTI} and μ_{TTI}, the SSD can be described as a function of T_a once μ_a, μ_b, σ_a, σ_b, and c are quantified, based on the regression coefficients a and b determined for a set of aquatic species. The regression coefficients will be determined from experiments that comply with the selection criteria described below. Although several guidelines exist for the data composition (e.g., number of taxonomic groups and species), the number of data required for a successful assessment is not fixed (Suter II et al., 2002). The reliability of the SSD will increase with a higher number of data, however, in the present study the availability of data is limiting.

4.2.3 Selection criteria

When multiple data are available for a single species, there are several options for including them in an SSD (Wheeler et al., 2002). The lowest effect value could be used but another option is to use the geometric mean of all available data. However, the first option would mean reducing the dataset on which the SSD is based, and the latter would not allow the linear regression method used in the present study. We choose to include all data, if exposure periods were similar, for the determination of the regression coefficients a and b. If the squared Pearson correlation between TTI and T_a for the combined data sets is less then

0.9, additional selection criteria were used, based on life stage. The youngest life stage is used when the combined data show low correlation.

The exposure period varied between the studies reported in the literature. Although for the incipient lethal temperature the endpoint is determined theoretically after an indefinite exposure, the tests usually limit exposure periods to 96 h or at least 24 h, but sometimes even as short as 0.5 h. The acclimation period was generally more than 96 h. In one study, the acclimation periods differed between 6 and 204 h. Some publications did not report the duration of the acclimation period (Appendix 4.A.4, Table 4.1). The acclimation and exposure period only were used to select data, when multiple data for a species were available. In that case the longest exposure and acclimation period were used.

The origin of the species is not included in the selection criteria. Although, every species has an optimum temperature range which depends on the geographical location of the species. Species with a narrow range are called stenothermal and species with a large range are eurythermal. There are arctic, as well as tropical stenothermal species. In the marine ecosystem, most species are stenothermal. The marine eurythermal species mainly occur in the coastal areas. Most fresh- and brackish water species are eurythermal (Hartholt and Jager, 2004). However, there are no indications that the relationship between the acclimation temperature and temperature tolerance should be different between regions as the latter is expressed relative to the acclimation temperature.

At higher temperatures, the solubility of oxygen in water is reduced. If in an experimental setup the water was not aerated or at least monitored, the observed effect might be (partially) caused by hypoxia as well. Studies that were explicitly conducted under hypoxic conditions were not included in the construction of the SSD.

A linear relationship is assumed between TTI and T_a. For most species, the relationship was quite strong. Species for which this was not the case (the squared Pearson correlation, $r^2 < 0.9$) were excluded from further analysis. Four species (with r^2 between 0.86 and 0.90) were dismissed for this reason.

4.2.4 Example of application of the SSD

After the SSD for temperature is developed, it will be applied to calculate the potentially affected fraction (PAF) of aquatic species in a real life situation, where a power plant in Velsen-Noord, in the North Sea Canal in The Netherlands (Fig. 4.1) uses canal water to cool its processes. The water temperature has been monitored over time, both upstream and downstream of the power plant, and can be found in the waterbase database (`http://www.waterbase.nl`). This public database contains validated measurements of the Directorate-General for Water Affairs of the Dutch Ministry of Transport, Public Works and Water Management. The effects on the water temperature are used to estimate, based

on the SSD, what fraction of species (PAF) will potentially be affected. The PAF is determined from the normal distribution (Eqn. 4.3) fitted to the sensitivity data. The standard deviation and average values, required for normal distribution calculations are determined with Eqns. 4.5 and 4.6. These equations require a background or acclimation temperature, where the upstream temperature is used. The difference between the upstream and downstream temperatures is used as the exposure metric x, in the normal distribution (Eqn. 4.3).

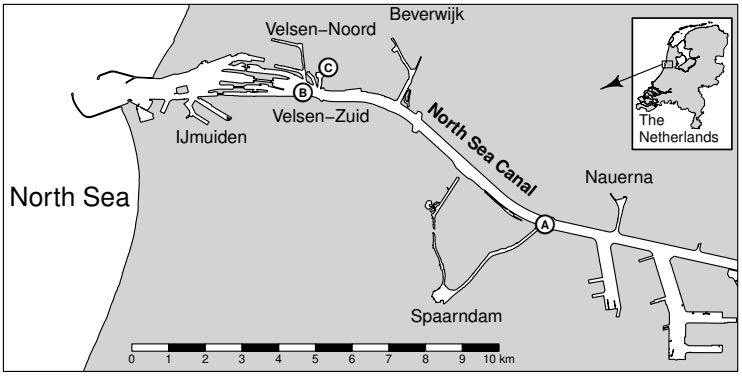


Fig. 4.1: The location in the North Sea Canal in The Netherlands where the water temperature is monitored upstream (A) and downstream (B) of the power plant (C). Map based upon material from `http://www.openstreetmap.org` (© OpenStreetMap contributors) made available under the Open Database License.

4.2.5 Extrapolation to lower effect levels

For toxic stress SSDs usually consist of chronic no-observed-effect concentration values (Aldenberg and Slob, 1993; Straalen and Denneman, 1989). In the present study, 50% effect levels were used, as no-observable-effect levels are rarely published for the effects of a temperature increase. It is desirable to extrapolate our results to lower effect levels, for more conservative risk assessment of temperature changes, as is applied for toxic stress. Sullivan et al. (2000) reported a linear relationship between the 10% lethal temperatures (LT10) and the LT50 for salmonids. We found a similar relationship based upon lethal temperature curves of 13 species (Al-Habbib and Grainger, 1977; Kellogg et al., 1984; Otto, 1973; Paul, 1980; Rantin and Petersen, 1985; Woo and Fung, 1980; Urban, 1994) (fish and bivalves): $LT10 = 0.98 \cdot LT50 - 0.88$ ($N = 53$, $r^2 = 0.992$, see also Appendix 4.A.6). In the present study we assume that this relationship applies to all species that were included in our SSD and we use it to extrapolate all our results from 50 to 10% mortality levels to achieve a more conservative risk assessment.

4.3 Results

For 50 aquatic species studies on the effect of temperature increases were selected according to the quality criteria. From these studies the regression coefficients a_i and b_i for Eqn. 4.2 were calculated. The average values and standard deviation of these regression coefficients describe the SSD at a certain T_a (Eqns. 4.2, 4.5, and 4.6) and were calculated at: μ_a = –0.8362, σ_a = 0.1244, μ_b = 27.46 °C and σ_b = 4.961 °C (Appendix 4.A.4, Table 4.1). In addition, the constant c was calculated from the regression coefficients using Eqn. 4.7 at –0.495 °C. This constant is also required to describe the SSD at a certain T_a. The relationship between the T_a and the hazardous temperature increase (HTI) for 50% of the species (HTI50) is linear and the relation with HTI5 is nearly linear (Fig. 4.2A). The HTI is the equivalent of the hazardous concentration of the traditional SSDs for toxicity. The HTI50 is the exposure metric where potentially 50% of the species is affected and is not the same as the LT50 or the TTI, which are both measures of sensitivity of a single species. At low T_a the sensitivity of the species is much lower than at higher T_a. When sensitivity of the species is extrapolated from 50 to 10% mortality, both HTI5 and HTI50 drop less than 2 °C for all acclimation temperatures (Fig. 4.2A). The standard deviation of the temperature interval (TTI) by which the temperature can increase above the background temperature without killing more than 50% of the population is a function of T_a and is between 4.5 and 5 °C for acclimation temperatures between 0 and 30 °C (Fig. 4.2B). The standard deviation is minimal at 16 °C. After extrapolation from 50 to 10% mortality levels, the standard deviation decreases with less than 0.2 °C for all acclimation temperatures (Fig. 4.2B).

Based on the data derived from the literature using Eqn. 4.2, three SSDs were constructed for three relevant acclimation temperatures in The Netherlands: 5, 12.5, and 20 °C (Fig. 4.3). These SSDs can be fitted according to a normal distribution. For species from different subphyla of the animal kingdom different markers are used. As can be seen from the figure, especially the vertebrates (all fish) are most sensitive to temperature changes, as they dominate the left side of the curve.

The temperatures downstream and upstream (Fig. 4.4A) of the power plant at Velsen-Noord in the North Sea Canal show a clear seasonal effect and vary between 3 and 25 °C. The differences between upstream and downstream temperatures, calculated from these data vary up to 7 °C (Fig. 4.4B).

The upstream temperature is considered as the background or T_a. Based on this, the SSD parameters σ_{TTI} and μ_{TTI} were calculated for the selected period, using Eqns. 4.5 and 4.6 (Fig. 4.5A). Using the relative warming (ΔT) of the water at Velsen-Noord, the SSD parameters and Eqn. 4.3, the PAF is calculated (Fig. 4.5B). The PAF fluctuates with peaks up to 0.09 in the summer, when using the SSD based on 50% effect levels. When the 10% effect levels are used, the PAFs peak

up to 0.14 (Fig. 4.5B).

4.4 Discussion

The present study describes the development of an SSD for temperature induced mortality. The SSD is based on data from literature for 50 aquatic species. When the SSD is applied to a real life situation in The Netherlands, the potentially affected fraction of species fluctuates with peaks up to 0.14 in summertime.

4.4.1 Quality of the data and the SSD

The quality of the SSD depends on the quantity and quality of the underpinning data set. As the availability of suitable temperature tolerance data is low, the selection criteria were defined not too strict to ensure the inclusion of enough data for the construction of the SSD.

Test procedures and conditions used in literature for the determination of temperature tolerance vary greatly. These differences cause sometimes high variations in sensitivity and therewith reduce the reliability of the SSD. It is therefore advisable to define standardised conditions for temperature tolerance tests, as is done for toxicity tests, to improve comparability of the results and their applicability in risk assessment.

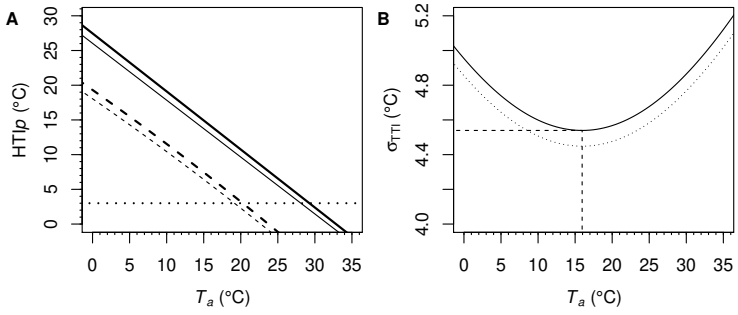


Fig. 4.2: (A) The hazardous temperature increases for 50% of the species (HTI50 = μ_{TTI}; solid lines) as calculated with Eqn. 4.5 and for 5% of the species (HTI5; dashed lines) as a function of the acclimation temperature (T_a) based on the data of 50 species. Bold lines indicate the hazardous temperature increases based on 50% mortality data, thin lines indicate hazardous temperature increases extrapolated to 10% mortality. The dotted line indicates the current EU water quality objective for cyprinid waters. (B) The standard deviation (σ_{TTI}; solid line) of the temperature tolerance interval (TTI) is a function of T_a (determined in the present study using Eqn. 4.6 and the parameters σ_a, σ_b, and c). The dotted line is the extrapolation of the standard deviation of the TTI from 50% mortality to 10% mortality. The dashed line indicates at which T_a the standard deviation of TTI is minimal.

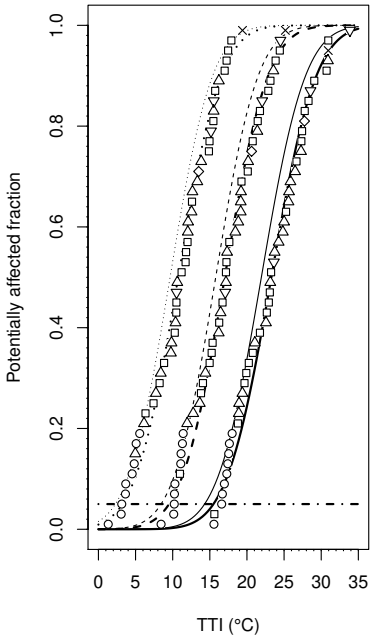


Fig. 4.3: Species sensitivity distributions for temperature tolerance intervals (TTIs) at different acclimation temperatures: 5 (solid lines), 12.5 (dashed lines) and 20 °C (dotted lines); markers (□ non-salmonid vertebrates; ○ salmonids; △ mollusca; ▽ crustacean; ◇ medusozoa; × annelida) indicate TTIs (50% mortality) for 50 individual species from different animal classes as determined in the present study based on literature data. Bold lines indicate species sensitivity distributions based on 50% mortality TTIs, thin lines indicate species sensitivity distributions extrapolated to 10% mortality. The line with alternating dashes and dots indicates the generally accepted risk level of 5% (Aldenberg and Slob, 1993; Straalen and Denneman, 1989).

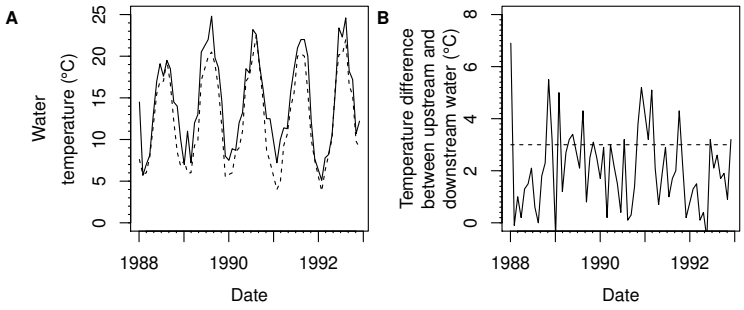


Fig. 4.4: (A) Water temperature upstream (dashed line) and downstream (solid line) of the power plant at Velsen-Noord in the North Sea Canal (data source: **http://www.waterbase.nl**) and (B) calculated temperature difference between them (solid line), together with the current European quality objective for cyprinid waters (dashed line).

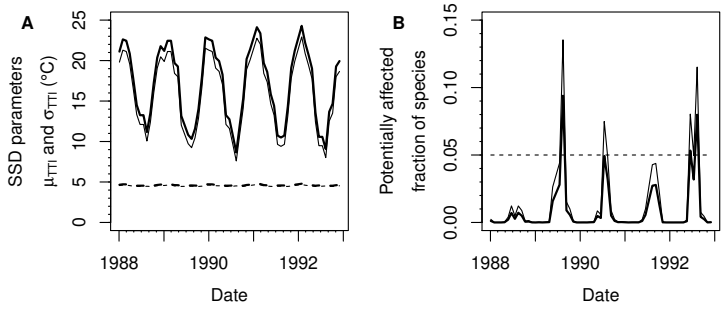


Fig. 4.5: (A) Species sensitivity distributions (SSDs) parameters σ_{TTI} (variation in species sensitivity, bold dashed based on 50% mortality; thin dashed line based on 10% mortality) and μ_{TTI} (average species sensitivity, bold solid line based on 50% mortality; thin solid line based on 10% mortality) for the situation at Velsen-Noord and (B) the potentially affected fraction (bold solid line based on 50% mortality; thin solid line based on 10% mortality) as a result of the thermal discharge at Velsen-Noord, calculated with the SSD parameters and the temperatures at the location, together with the generally accepted risk level of 5% (dashed line) (Straalen and Denneman, 1989; Aldenberg and Slob, 1993).

A subgroup of the vertebrates (mainly composed of salmonids, the family of Salmonidae) appears to the left side of the SSD (Fig. 4.3). This indicates they are more sensitive to thermal effects than the normal distribution would predict. Salmonids are known to be sensitive to temperature effects (Sullivan et al., 2000). If this subgroup of nine salmonids is omitted from the SSD, the average TTI increases and the standard deviation of TTI decreases with approximately 1.5 and 0.9 °C respectively. The goodness of fit of the normal distribution to the data improves in this case. This suggests the need of an area-specific approach for the risk assessment of a thermal discharge if specifically sensitive species are present.

The use of linear regression-derived data instead of the reported experimental data introduces extra uncertainty to the model. However, when such an SSD for T_a = 20°C is compared with one based on data from literature, there is only little difference in average sensitivity of 0.5%, and in variance of 1.3%. For this comparison, eight species had to be excluded, since no data were available at an acclimation temperature of 20 °C. This reveals an important argument to use linear regression; as one does not have to dismiss species because the method does not depend on actual acclimation temperatures.

Another advantage of linear regression is that the SSD can be interpolated and if necessary extrapolated for any T_a. All of the experiments, used to develop the SSD, were performed at acclimation temperatures between 0.5 and 36 °C. Therefore, the model is expected to be most reliable in that range. Monte Carlo simulations (unpublished results) show that the uncertainty, introduced into the SSD by the variance of the regression parameters, increases with increasing T_a. This indicates that the SSD is more accurate at the lower acclimation temperatures.

The linear model defined to describe the TTI nicely fits the data ($r^2 > 0.9$), the HTI50 inherits this linearity. It is possible, however, that certain data might deviate some from linearity. Also, the HTI50 might deviate from linearity, but based on the good correlation of the underpinning data, this deviation is not expected to be large.

4.4.2 Applicability in risk assessment

The SSD, as presented, applies to risk assessment of heat discharges on a local scale. For toxicity a PAF of 5% or less is generally considered acceptable (Straalen and Denneman, 1989; Aldenberg and Slob, 1993). When this limit is applied to the current SSD and adopted as limit for cyprinid waters in our example of the North Sea Canal, the PAF exceeds this 5% limit less frequent and at different times than the temperature increase would exceed the current European quality objective of 3 °C (Figs. 4.4 and 4.5). The important improvement of our SSD approach is that it more realistically depends on ambient temperature

while the 3 °C limit does not. The SSD method is not over-protective as it is not based on no-observable-effect levels, but the 50% mortality levels have been extrapolated to 10% to make the risk assessment more conservative. A shortcoming of the SSD applied is that it contains species that are not indigenous to the North Sea Canal. Even though temperature elevations (TTI), relative to ambient temperatures, are used to express the species sensitivity; we observed a significant (*t* test, two-tailed, two-sample unequal variance, $p < 0.001$) difference in sensitivity between vertebrates (fish) from (sub)tropical and temperate regions (Fig. 4.6). This means that selecting species from comparable natural ambient temperature would make the SSD represent the water body better. This, however, would decrease the sample size thus, the quality of the SSD. Therefore, a good balance between the applicability and quality of the SSD has to be considered, when selecting species to represent a certain water body type.

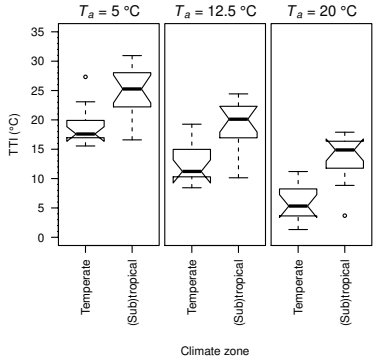


Fig. 4.6: Sensitivity towards a temperature increase (TTI) of vertebrates (all fish) from temperate regions (N = 14) and (sub)tropical regions (N = 15) at three acclimation temperatures (T_a): 5, 12.5 and 20 °C; boxes show first, second and third quartiles; whiskers show minimum and maximum values, excluding outliers which are shown as markers; notches indicate ±1.58 times the interquartile range; the difference in sensitivity towards a temperature increase between fish from the two climates is significant for all three acclimation temperatures (*t* test, two-tailed, two-sample unequal variance, $p < 0.001$).

The data set in the present study was mainly composed of fish (N = 29) and molluscs (N = 16). In addition, the set also includes data on a medusozoa (N = 1), crustacea (N = 3), and an annelida (N = 1). The data set is too small to give any indication on differences in sensitivity between subphyla. The same is the case for differences between marine and freshwater species. It is important to include more marine and freshwater species of different subphyla in case of future development of a data set with species native to moderate temperatures.

The SSD is based on upper temperature tolerances only. As in the practice of risk assessment of thermal discharges, the ambient temperature will only in-

crease and therefore not likely interfere with the lower temperature tolerance of species. The SSD also does not include that a shift in the ambient temperature may be unfavorable to some species and favorable to others. This could sometimes result in a shift in the competitive balance between species, thus in indirect changes in species composition. When including the lower tolerance limits as well, this aspect could partially be taken into account. Latour et al. (1994) used both lower and upper tolerance limits, or more accurately the species-response function for environmental variables (soil nitrogen load and moisture change), in calculating the fraction of protected species (the reverse of the potentially affected fraction). The same could be done for temperature; however, it is important to note that the temperature tolerance depends on the acclimation temperature. As lower temperature tolerance limits are less available in literature than the upper limits, it is advisable to include these when performing new experiments with species native to moderate regions.

A general critique of probabilistic models is that they are based on individual-level end-points, which may not be directly or consistently related to risks for populations (Forbes and Forbes, 1993; Forbes et al., 2001). However, the approach of protecting individual species to protect ecosystems has been shown to be protective with toxicants (Hose and Van den Brink, 2004; Van Wijngaarden et al., 2005; Selck et al., 2002). Of course the SSD-model is not mechanistic and does not incorporate understanding of the biology of the system, and below lethal temperatures sub-lethal effects such as changed behavior and especially in wintertime, some temperature dependent or triggered processes (e.g., growth, spawning, nesting, and egg-laying of fish (Daniels, 1978; Kjellman and Eloranta, 2002; Rowland, 1983) might occur. Whether our method also is applicable to protect against sub-lethal effects, possibly with the application of additional safety factors, needs to be investigated.

The variation of species sensitivity appears to be minimal at an acclimation temperature of 16.0 °C (Fig. 4.2B). This means that the slope of the SSD is steepest at that T_a and the risk level (PAF) will increase strongest at that T_a with increasing TTIs. This minimum is theoretical and depends on the selected species; it results indirectly from the linear relationship between T_a and TTI. Due to these linear relationships, there will always be an acclimation temperature at which the variation between species is minimal (Appendix 4.A.5). Whether this minimum is of any practical relevance is unclear. However, this minimum variation could be used as a worst-case-scenario in risk assessment, as it represents the steepest slope of the SSD.

The correlation between T_a and TTI (Eqn. 4.2) will result in negative values for TTI at certain acclimation temperatures that have no physical meaning. Theoretically, species with negative TTI values already show 50% or more mortality at the acclimation temperature. These species obviously could not acclimate to the given temperature. The species, selected in this study, all have positive TTI

values at acclimation temperatures below 21.3 °C. As the distribution of species is not on a logarithmic scale, TTI can become 0, so there will always be an affected fraction, even at TTI = 0. We suggest to consider the risk level at TTI = 0 as a background risk level. With the presented data set, the PAF is negligible (< 0.001) at acclimation temperatures below 16.0 °C. Struijs et al. (1997) introduce the added-risk approach for toxicants, to deal with the background risk. They consider the background risk (in our case the PAF at TTI = 0 °C) acceptable and therefore disregard it. The PAF is calculated for the remaining fraction only. This approach is applicable in our case. It will, however, complicate potential applications of the SSD, i.e. the construction and interpretation of multi-stressor PAFs for integrated risk assessment, and requires further study.

Although our SSD method is different from the current European practice of determining the risk of temperature elevation of surface water, it can be applied in a tiered approach of risk assessment. It can be used with modelled temperatures to predict risk in a certain system, but specific SSDs can also be applied to predict the risk for local situations based on actual measurements of the temperature as shown in the present study. Also, the more conservative European water quality standard could be used for initial identification of potential problem situations and as a second tier, the SSD could be applied for a more detailed quantification of the risk level. As sensitive subgroups are identified in the SSD, the next step could be to determine whether the potentially affected species are actually present or not. Since our model only contains acute mortality data, the possible sub-lethal effects and their consequences for population dynamics should be involved in the final tier. A promising aspect of the SSD approach is its potential use in integrated risk assessment for situations with multiple-stressors that currently are not yet taken into account together. In a first tier approach, effects from different stressors can be assumed to be additive, and expressed as a multi-stressor-PAF (De Zwart and Posthuma, 2005). However, some effects (e.g., toxicity and hypoxia) depend on temperature (Cairns et al., 1975; Heugens et al., 2001; Shimps et al., 2005). Heugens (2003) showed that assuming additive interaction of cadmium toxicity and temperature stress mostly results in an overestimation of effects on *Daphnia magna*, and is therefore protective. To what extent the assumption of additive interaction results in protective assessment for other toxicants, stressors and species requires further study.

4.5 Conclusions

The present study shows that the SSD method that is currently mainly applied to quantify toxic stress is very suitable to estimate the risk of thermal effects. Because the SSD method uses multiple temperature tolerance values, the model is more realistic than the traditional approach using only data for the most sensitive species. Although the model can be used for generic assessment, its

strength will be optimally used when it is based upon site relevant species for location specific assessment. The SSD-approach also is promising for the first tier in integrated risk assessment of multiple stressors (e.g., oxygen depletion, toxic stress), although it requires further study to assess the relevance of interactions between the stressors and their effects. Location-specific multi-stressor-PAFs would be a good basis for implementation of the Water Framework Directive's intention of integrating principles for protection and sustainable use of water.

Acknowledgements

The present study is based upon work performed in a knowledge investment program, funded by the Netherlands Organisation for Applied Scientific Research (TNO). The writing of this paper is partially funded by Wageningen Institute for Marine Resources and Ecosystem Studies (IMARES). TNO and Wageningen IMARES are acknowledged for their support, both financially and scientifically.

References

Al-Habbib, O. A. M., Grainger, J. N. R., 1977. The effect of constant and changing temperatures on the thermal resistance of *Lymnaea peregra* (Müller). J Therm Biol 2 (4), 191–195.

Aldenberg, T., Jaworska, J. S., Traas, T. P., 2002. Normal species sensitivity distributions and probabilistic ecological risk assessment. In: Posthuma, L., Suter II, G. W., Traas, T. P. (Eds.), Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, US, pp. 49–102.

Aldenberg, T., Slob, W., 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotox Environ Safe 25 (1), 48–63.

Beitinger, T. L., Bennett, W. A., McCauley, R. W., 2000. Temperature tolerances of North American freshwater fishes exposed to dynamic changes in temperature. Environ Biol Fish 58, 237–275.

Cairns, J., Heath, A. G., Parker, B. C., 1975. The effects of temperature upon the toxicity of chemicals to aquatic organisms. Hydrobiologia 47, 135–171.

Carter, H. H., Schubel, J. R., Wilson, R. E., Woodhead, P. M. J., 1979. Thermally induced biological effects caused by once-through cooling systems: A rationale for evaluation. Environ Manage 3, 353–368.

Daniels, R. A., 1978. Nesting behaviour of *Harpagifer bispinis* in Arthur Harbour, *Antarctic Peninsula*. J Fish Biol 12 (5), 465–474.

De Zwart, D., Posthuma, L., 2005. Complex mixture toxicity for single and multiple species: Proposed methodologies. Environ Toxicol Chem 24 (10), 2665–2676.

EC, 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. Off J Eur Comm L327, 1–72.

EC, 2006. Directive 2006/44/EC of the European Parliament and of the Council of 6 September 2006, on the quality of fresh waters needing protection or improvement in order to support fish life. Off J Eur Comm L264, 20–31.

Forbes, T. L., Forbes, V. E., 1993. A critique of the use of distribution-based extrapolation models in ecotoxicology. Funct Ecol 7, 249–254.

Forbes, V. E., Calow, P., Sibly, R. M., 2001. Are current species extrapolation models a good basis for ecological risk assessment? Environ Toxicol Chem 20 (2), 442–447.

Gaddum, J. H., 1945. Lognormal distributions. Nature 156, 463.

Hartholt, J. G., Jager, Z., 2004. Effecten van koelwater op het zoute aquatische milieu. Tech. Rep. RIKZ/2004.043, Rijksinstituut voor Kust en Zee, Den Haag, The Netherlands.

Heugens, E. H. W., 2003. Predicting effects of multiple stressors on aquatic biota. Ph.D. thesis, University of Amsterdam, Amsterdam, The Netherlands.

Heugens, E. H. W., Hendriks, A. J., Dekker, T., Straalen, N. M. v., Admiraal, W., 2001. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. Crit Rev Toxicol 31 (3), 247–284.

Hose, G. C., Van den Brink, P. J., 2004. Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. Arch Environ Con Tox 47, 511–520.

Jobling, M., 1981. Temperature tolerance and the final preferendum-rapid methods for the assessment of optimum growth temperatures. J Fish Biol 19 (4), 439–455.

Kellogg, R. L., Ligotino, R. J., Jinks, S. M., 1984. Thermal mortality prediction equations for entrainable striped bass. T Am Fish Soc 113 (6), 794–802.

Kerkum, L. C. M., Bij de Vaate, A., Bijstra, D., De Jong, S. P., Jenner, H. A., 2004. Effecten van koelwater op het zoete aquatische milieu. Tech. Rep. 2004.033, Rijkswaterstaat-Rijksinstituut voor Integraal Zoetwaterbeheer en Afvalwaterbehandeling, Lelystad, The Netherlands.

Kjellman, J., Eloranta, A., 2002. Field estimations of temperature-dependent processes: case growth of young burbot. Hydrobiologia 481, 187–192.

Latour, J. B., Reiling, R., Slooff, W., 1994. Ecological standards for eutrophication and desiccation: Perspectives for a risk assessment. Water Air Soil Poll 78, 265–277.

McErlean, A. J., Mihursky, J. A., Brinkley, H. S., 1969. Determination of upper temperature tolerance triangles for aquatic organisms. Chesap Sci 10 (3-4), 293–296.

McMahon, R. F., Matthews, M. A., Ussery, T. A., Chase, R., Clarke, M., 1995. Studies of heat tolerance of zebra mussels: Effects of temperature acclimation and chronic exposure to lethal temperatures. Tech. Rep. EL-95-9, Center for Biological Macrofouling Research, Arlington, TX, USA.

Otto, R. G., 1973. Temperature tolerance of the mosquitofish, *Gambmia affinis* (Baird and Girard). J Fish Biol 5 (5), 575–585.

Paul, J. D., 1980. Salinity-temperature relationships in the queen scallop *Chalamys opercularis*. Mar Biol 56, 295–300.

Rantin, F. T., Petersen, J. A., 1985. Thermal tolerance of South American cichlid, *Geophagus brasiliensis*. Rev Hydrobiol Trop 18 (3), 221–226.

Rowland, S. J., 1983. Spawning of the Australian freshwater fish murray cod, *Maccullochella peeli* (Mitchell), in earthen ponds. J Fish Biol 23 (5), 525–534.

Selck, H., Riemann, B., Christoffersen, K., Forbes, V. E., Gustavson, K., Hansen, B. W., Jacobsen, J. A., Kusk, O. K., Petersen, S., 2002. Comparing sensitivity of ecotoxicological effect endpoints between laboratory and field. Ecotox Environ Safe 52 (2), 97–112.

Sheahan, D., Rycroft, R., Allen, Y., Kenny, A., Mason, C., Irish, R., 2001. Contaminant status of the north sea. Tech. Rep. TR_004, Centre for Environment, Fisheries and Agriculture Science, Lowestoft, UK.

Shimps, E. L., Rice, J. A., Osborne, J. A., 2005. Hypoxia tolerance in two juvenile estuary-dependent fishes. J Exp Mar Biol Ecol 325 (2), 146–162.

Smit, M. G. D., Hendriks, A. J., Schobben, J. H. M., Karman, C. C., Schobben, H. P. M., 2001. The variation in slope of concentration-effect relationships. Ecotox Environ Safe 48 (1), 43–50.

Smit, M. G. D., Holthaus, K. I. E., Trannum, H. C., Neff, J. M., Kjeilen-Eilertsen, G., Jak, R. G., Singsaas, I., Huijbregts, M. A. J., Hendriks, A. J., 2008. Species sensitivity distributions for suspended clays, sediment burial, and grain size change in the marine environment. Environ Tox Chem 27 (4), 1006–1012.

Straalen, N. M. V., Denneman, C. A. J., 1989. Ecotoxicological evaluation of soil quality criteria. Ecotox Environ Safe 18 (3), 241–251.

Struijs, J., Van de Meent, D., Peijnenburg, W. J. G. M., Van den Hoop, M. A. G. T.,

Crommentuijn, T., 1997. Added risk approach to derive maximum permissible concentrations for heavy metals: How to take natural background levels into account. Ecotox Environ Safe 37 (2), 112–118.

Sullivan, K., Martin, D. J., Cardwell, R. D., Toll, J. E., Duke, S., 2000. An analysis of the effects of temperature on salmonids of the Pacific Northwest with implications for selecting temperature criteria. Tech. rep., Sustainable Ecosystems Institute, Portland, OR, USA.

Suter II, G. W., Traas, T. P., Posthuma, L., 2002. Issues and practices in the derivation and use of species sensitivity distributions. In: Posthuma, L., Suter II, G. W., Traas, T. P. (Eds.), Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, US, pp. 455–458.

Urban, H.-J., 1994. Upper temperature tolerance of ten bivalve species off Peru and Chile related to El Niño. Mar Ecol Prog Ser 107, 139–145.

Van Wijngaarden, R., Brock, T., Van den Brink, P., 2005. Threshold levels for effects of insecticides in freshwater ecosystems: A review. Ecotoxicology 14, 355–380.

Wheeler, J. R., Grist, E. P. M., Leung, K. M. Y., Morritt, D., Crane, M., 2002. Species sensitivity distributions: data and model choice. Mar Pollut Bull 45 (1-12), 192–202.

Woo, N. Y. S., Fung, A. C. Y., 1980. Studies on the biology of the red sea bream *Chrysophrys major*. I. temperature tolerance. Mar Ecol Prog Ser 3, 121–124.

4.A Appendix

4.A.1 The linear relationship between T_a and TTI

This section provides an example based on data for the bivalve *Mya arenaria* from Kennedy and Mihursky (1971). A linear relationship is found between the Temperature Tolerance Interval (TTI) and the T_a (Fig. 4.7).

4.A.2 Goodness of fit tests

Statistical goodness-of-fit tests show that, generally for all acclimation temperatures, the normal distribution fits the temperature sensitivity data best (Fig. 4.8)

4.A.3 Derivation of Eqn. 4.6

Here σ_{TTI}, the standard deviation of TTI for the selected species, will be expressed as a function of T_a. We start with the definition of the standard deviation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{4.8}$$

Substitution with Eqn. 4.2 and 4.5 gives:

$$\sigma_{\mathrm{TTI}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(a_i T_a + b_i - \mu_{\mathrm{TTI}})^2} \tag{4.9}$$

and consequently, using Eqn. 4.5:

$$\sigma_{\mathrm{TTI}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}((a_i-\mu_a)\,T_a + (b_i-\mu_b))^2}, \tag{4.10}$$

which can be extended to:

$$\sigma_{\mathrm{TTI}} = \left(\sqrt{\frac{1}{N-1}}\right)\times \sqrt{T_a^2\sum_{i=1}^{N}(a_i-\mu_a)^2 + 2T_a\sum_{i=1}^{N}(a_i-\mu_a)(b_i-\mu_b) + \sum_{i=1}^{N}(b_i-\mu_b)^2} \tag{4.11}$$

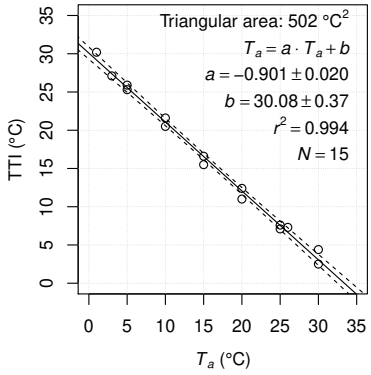


Fig. 4.7: Linear relationship between T_a and TTI for *Mya arenaria*; markers show experimental values (Kennedy and Mihursky, 1971), solid line shows fitted linear line; dashed lines are 95% confidence limits; regression parameters a and b are listed as best fit ± standard error.

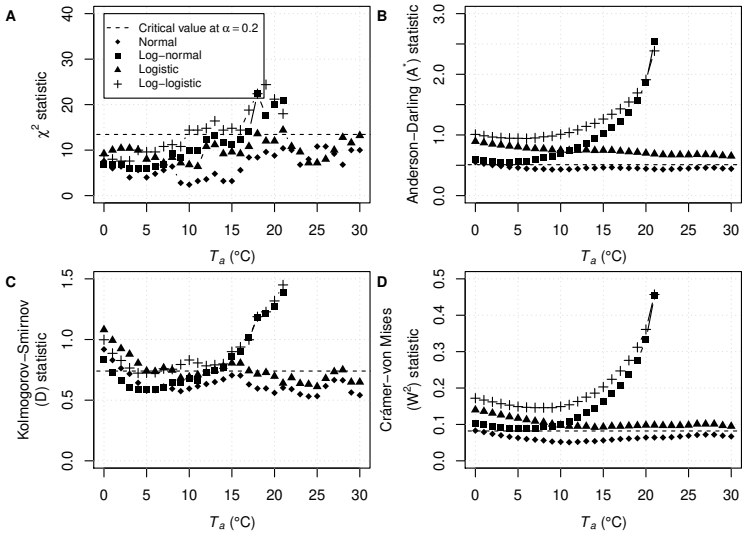


Fig. 4.8: (A) χ^2, (B) Anderson-Darling, (C) Kolmogorov-Smirnov and (D) Crámer-von Mises tests at different T_a values, with their critical value for α = 20% represented by dashed lines.

In this equation the definition of the standard deviation (Eqn. 4.8) can be recognised for a and b. Substitution gives:

$$\sigma_{TTI} = \sqrt{T_a^2\sigma_a^2 + \sigma_b^2 + \frac{2T_a}{N-1}\sum_{i=1}^{N}(a_i - \mu_a)(b_i - \mu_b)} \tag{4.12}$$

Constant c is now defined as:

$$c = \frac{2}{N-1}\sum_{i=1}^{N}(a_i - \mu_a)(b_i - \mu_b) \tag{4.13}$$

Substitution of c into Eqn. 4.13 gives:

$$\sigma_{TTI} = \sqrt{T_a^2\sigma_a^2 + \sigma_b^2 + cT_a} \tag{4.14}$$

This concludes the derivation of Eqn. 4.6.

4.A.4 Regression coefficients

Table 4.1: Overview of regression coefficients a and b and their standard deviation derived with linear regression; N is the number of data points used for linear regression and r^2 is the square of the Pearson moment correlation coefficient of the regression. The acclimation and exposure duration in hours, unless mentioned otherwise and the source of the effect data is also given.

Species	a	b (°C)	N	r^2	Acclimation duration (h)	Exposure duration (h)	Source
Annelida							
Clymenella torquata	-0.770±0.090	34.8±1.0	3	0.995	336	1 °C / 5 min	Kenny (1969)
Crustacea							
Callinectes sapidus	-1.247±0.114	40.1±2.3	16	0.993	504	48	Tagatz (1969)
Daphnia magna	-0.868±0.146	32.9±3.2	6	0.981	>336	1	Paul et al. (2004); Zeis et al. (2004)
Pacifastacus leniusculus	-0.874±0.049	28.0±0.9	17	0.997	168–504	48	Becker et al. (1975)
Medusozoa							
Chrysaora quinquecirrha	-0.950±0.082	32.5±2.2	5	0.996	1–240	24	Gatz et al. (1973)
Mollusca							
Cerastoderma edule	-0.926±0.029	28.9±0.5	5	0.999	–[11]	96	Ansell et al. (1981)
Cerastoderma glaucum	-0.843±0.079	30.7±1.4	4	0.996	–[11]	96	Ansell et al. (1981)
Cerastoderma tuberculatum	-0.848±0.048	26.8±0.9	4	0.998	–[11]	96	Ansell et al. (1981)

Continued on next page

[11] Not reported in original publication

Table 4.1 continued

Species	*a*	*b* (°C)	*N*	r^2	Accl. dur.	Exp. dur.	Source
Chlamys opercularis	-0.932±0.034	23.6±0.5	4	0.999	504–672	48	Paul (1980)
Donax semistriatus	-0.892±0.043	27.6±0.8	4	0.999	–[11]	96	Ansell et al. (1980a)
Donax trunculus	-0.853±0.045	29.1±0.8	4	0.999	–[11]	96	Ansell et al. (1980a)
Donax vittatus	-0.700±0.202	22.4±2.6	5	0.953	–[11]	96	Ansell et al. (1980a)
Gemma gemma	-0.979±0.093	35.8±1.8	6	0.994	–[11]	96	Kennedy and Mihursky (1971)
Lymnaea peregra	-0.784±0.049	31.2±0.6	3	0.999	360–720	1	Al-Habbib and Grainger (1977)
Macoma balthica	-0.888±0.044	30.3±0.9	12	0.998	–[11]	24	Kennedy and Mihursky (1971)
Mulinia lateralis	-0.867±0.033	30.0±0.6	4	0.999	–[11]	24	Kennedy and Mihursky (1971)
Mytilus edulis	-0.888±0.051	25.5±0.8	5	0.998	>336	24	Wallis (1975)
Mya arenaria	-0.901±0.076	30.1±1.4	15	0.994	–[11]	24	Kennedy and Mihursky (1971)
Tellina fabula	-0.883±0.051	23.8±0.7	4	0.998	–[11]	96	Ansell et al. (1980b)
Tellina tenuis	-0.874±0.063	28.1±1.0	8	0.996	–[11]	96	Ansell et al. (1980b)
Trichomya hirsuta	-0.779±0.107	28.9±2.5	7	0.987	>336	24	Wallis (1977)
Vertebrata							
Alosa pseudoharengus	-0.626±0.009	22.8±0.2	3	1.000	720	168	Otto et al. (1976)
Ambassis commersoni	-0.796±0.010	31.4±0.3	5	1.000	168	24	Rajaguru and Ramachandran (2001)
Chrysophrys major	-0.663±0.151	22.1±3.4	5	0.970	336	48	Woo and Fung (1980)

Continued on next page

Table 4.1 continued

Species	*a*	*b* (°C)	*N*	r^2	Accl. dur.	Exp. dur.	Source
Coregonus artedii	-0.742±0.142	20.4±2.2	5	0.979	504	167	Edsall and Colby (1970)
Esox lucius	-0.618±0.205	18.7±2.6	6	0.932	–[11]	168	Hokanson et al. (1973)
Etroplus suratensis	-0.722±0.043	31.9±1.2	5	0.998	168	24	Rajaguru and Ramachandran (2001); Rajaguru (2002)
Geophagus brasiliensis	-0.687±0.049	28.7±1.2	6	0.997	672	15 °C / h	Rantin and Petersen (1985)
Gambusia affinis	-0.786±0.101	30.6±2.3	14	0.986	>5040	168	Otto (1973)
Geotria australis	-1.044±0.015	28.3±0.3	3	1.000	504–672	96	Macey and Potter (1978)
Girella nigricans	-0.831±0.169	27.1±3.6	3	0.986	1440	72	Doudoroff (1942)
Leiostomus xanthurus	-0.610±0.050	24.8±0.8	3	0.998	96	24	Hartwell and Hoss (1979)
Monacanthus chinensis	-0.725±0.066	26.3±1.7	3	0.997	240	24–48	Menasveta (1981)
Morone chrysops	-1.075±0.160	32.7±3.2	4	0.989	–[11]	24	McCormick (1978)
Morone saxatilis	-0.650[12]	23.5[12]	6	0.950	>336	72	Cook et al. (2006)
Mugil dussumerii	-0.804±0.144	31.8±3.7	3	0.989	240	24–48	Menasveta (1981)
Noemacheilus barbatulus	-0.778±0.138	23.8±2.4	14	0.974	336	1	Elliott et al. (1994)
Oncorhynchus gorbuscha	-0.861±0.047	20.9±0.8	5	0.998	168–672	–[13]	Brett (1952)
Oncorhynchus keta	-0.888±0.022	21.4±0.3	5	1.000	168–672	–[13]	Brett (1952)
Oncorhynchus kisutch	-0.879±0.027	22.4±0.4	5	0.999	168–672	–[13]	Brett (1952)

Continued on next page

[12]Regression parameter is presented in original publication without confidence intervals
[13]Exposure duration is variable. Lethal temperatures are calculated from resistance times.

Table 4.1 continued

Species	*a*	*b* (°C)	*N*	r^2	Accl. dur.	Exp. dur.	Source
Oncorhynchus nerka	-0.873±0.084	22.0±1.3	5	0.995	168–672	–[13]	Brett (1952)
Oncorhynchus tshawytscha	-0.831±0.146	21.7±2.4	5	0.982	168–672	–[13]	Brett (1952)
Poecilia sphenops	-0.870[12]	35.3[12]	6	0.950	720	48	Hernández and Bückle (2002)
Salmo salar	-0.943±0.091	22.0±1.2	3	0.997	120	168	Bishai (1960)
Salmo trutta	-0.957±0.056	22.2±0.7	6	0.998	120	168	Bishai (1960)
Salvelinus alpinus	-0.949±0.188	20.3±2.5	31	0.965	72–144	168	Baroudy and Elliott (1994); Elliott and Klemetsen (2002)
Sebasticus marmoratus	-0.750±0.050	27.0±1.0	3	0.999	336–504	96	Kita et al. (1996)
Sciaenops ocellatus	-0.603±0.083	23.8±1.7	6	0.988	>336	5 °C / h	Procarione and King (1993)
Therapon jarbua	-0.787±0.081	33.0±2.3	5	0.994	168	24	Rajaguru and Ramachandran (2001); Rajaguru (2002)
Tilapia mossambica	-0.915±0.051	35.3±1.5	8	0.998	96	83	Allanson and Noble (1964)
Average (μ_a and μ_b)	-0.836	27.5					
Standard deviation (σ_a and σ_b)	0.124	5.0					

End of Table 4.1

4.A.5 The minimum value of σ_{TTI}

In this section the minimum of σ_{TTI} will be derived. We start by taking the derivative of Eqn. 4.6:

$$\sigma'_{TTI} = \frac{2\sigma_a^2 T_a + c}{2\sqrt{\sigma_a^2 T_a^2 + \sigma_b^2 + c T_a}} \tag{4.15}$$

One of the properties of a minimum of an equation is that the derivative is equal to 0, which means that:

$$2\sigma_a^2 T_a + c = 0 \tag{4.16}$$

and in this case also:

$$\sigma_a^2 T_a^2 + \sigma_b^2 + c T_a > 0 \tag{4.17}$$

The latter condition (Eqn. 4.17) is always met, since it is an expression to calculate the standard deviation (Eqn. 4.14). This is always a positive number, given that there is a deviation from the average. Now, Eqn. 4.16 can be rewritten as:

$$T_a = -c/\left(2\sigma_a^2\right) \tag{4.18}$$

Eqn. 4.18 gives the value of T_a at $\sigma'_{TTI} = 0$. If there is truly a minimum at $T_a = -c/\left(2\sigma_a^2\right)$, σ'_{TTI} must be positive for $T_a > -c/\left(2\sigma_a^2\right)$ and negative for $T_a < -c/\left(2\sigma_a^2\right)$. To prove this, we take the following expression:

$$T_a = -\left(c - \Delta c\right)/\left(2\sigma_a^2\right); \tag{4.19}$$

where Δc represents a deviation of the T_a where $\sigma'_{TTI} = 0$. Substitution of Eqn. 4.19 into Eqn. 4.15 gives:

$$\sigma'_{TTI} = \frac{\Delta c}{2\sqrt{\sigma_a^2 T_a^2 + \sigma_b^2 + c T_a}} \tag{4.20}$$

It was already shown that the divisor is always positive, since it is an expression for a standard deviation. Hence, σ'_{TTI} is negative for any negative deviations ($\Delta c < 0$) from Eqn. 4.18 and vice versa, proving that Eqn. 4.18 is the minimum of Eqn. 4.6. The minimum of Eqn. 4.6 is now given by substituting Eqn. 4.18 into

Eqn. 4.6:

$$\sigma_{TTI,min} = \sqrt{-c^2/(4\sigma_a^2) + \sigma_b^2} \tag{4.21}$$

This concludes the derivation of the minimum value of σ_{TTI}.

4.A.6 Extrapolation to lower effect levels

In this section the method and some consequences of the extrapolation from 50% to 10% mortality levels are described. Sullivan et al. (2000) already found a correlation between the 10% and 50% mortality level (LT10 and LT50 respectively) for salmonids. In this section, the following relationship is assumed between LT10 and LT50:

$$LT10 = \varepsilon \cdot LT50 + \xi; \tag{4.22}$$

where ε and ξ are extrapolation factors. With linear regression, the values were calculated at: $\varepsilon = 0.98$ and $\xi = -0.88$ ($N = 53$, $r^2 = 0.992$; based on lethal temperature curves of 14 species). It is now assumed that this extrapolation is valid for all species in the SSD. For the TTI, distinction between 50% and 10% effect level can also be made:

$$TTI50 = LT50 - T_a \tag{4.23}$$

and consequently:

$$TTI10 = (\varepsilon - 1)\, T_a + \varepsilon \cdot TTI50 + \xi \tag{4.24}$$

The SSD parameters μ and σ can now also be extrapolated to the 10% effect level:

$$\begin{aligned} \mu_{TTI10} &= \\ (\varepsilon - 1)\, T_a + \varepsilon \cdot \mu_{TTI50} + \xi &= \\ (\varepsilon\,(\mu_a + 1) - 1)\, T_a + \varepsilon \cdot \mu_b + \xi & \end{aligned} \tag{4.25}$$

$$\sigma_{TTI10} = \varepsilon \cdot \sigma_{TTI50} = \varepsilon\sqrt{T_a^2\sigma_a^2 + \sigma_b^2 + c\,T_a} \tag{4.26}$$

The HTI5 (the 5% hazardous temperature interval, comparable with the 5% hacardous concentration, HC5) value can also be extrapolated to the 10% effect

level:

$$\mathrm{HTI5}_{\mathrm{TTI10}} = (\varepsilon - 1)\, T_a + \varepsilon \cdot \mathrm{HTI5}_{\mathrm{TTI50}} + \xi \tag{4.27}$$

More generally, the HTIp value can be expressed as:

$$\mathrm{HTI}p_{\mathrm{TTI10}} = \mu_{\mathrm{TTI10}} - k_s \sigma_{\mathrm{TTI10}}; \tag{4.28}$$

where k_s is the extrapolation factor as discussed by Aldenberg and Jaworska (2000).

$$\begin{aligned} \mathrm{HTI}p_{\mathrm{TTI10}} &= (\varepsilon - 1)\, T_a + \varepsilon \cdot (\mu_{\mathrm{TTI50}} - k_s \sigma_{\mathrm{TTI50}}) + \xi = \\ (\varepsilon\,(\mu_a + 1) - 1)\, T_a &+ \varepsilon \left(\mu_b - k_s \sqrt{T_a^2 \sigma_a^2 + \sigma_b^2 + c\, T_a} \right) + \xi \end{aligned} \tag{4.29}$$

References in Appendix 4.A

Al-Habbib, O. A. M., Grainger, J. N. R., 1977. The effect of constant and changing temperatures on the thermal resistance of *Lymnaea peregra* (Müller). J Therm Biol 2 (4), 191–195.

Aldenberg, T., Jaworska, J. S., 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotox Environ Safe 46 (1), 1–18.

Allanson, B. R., Noble, R. G., 1964. The tolerance of *Tilapia mossambica* (Peters) to high temperature. T Am Fish Soc 93 (4), 323–332.

Ansell, A. D., Barnett, P. R. O., Bodoy, A., Massé, H., 1980a. Upper temperature tolerances of some European molluscs II. *Donax vittatus*, *D. semistriatus* and *D. trunculus*. Mar Biol 58, 41–46.

Ansell, A. D., Barnett, P. R. O., Bodoy, A., Massé, H., 1980b. Upper temperature tolerances of some European molluscs. I. *Tellina fabula* and *T. tenuis*. Mar Biol 58 (1), 33–39.

Ansell, A. D., Barnett, P. R. O., Bodoy, A., Massé, H., 1981. Upper temperature tolerances of some European molluscs. Mar Biol 65, 177–183.

Baroudy, E., Elliott, J. M., 1994. The critical thermal limits for juvenile Arctic charr *Salvelinus alpinus*. J Fish Biol 45 (6), 1041–1053.

Becker, C. D., Genoway, R. G., Merrill, J. A., 1975. Resistance of a northwestern crayfish, *Pacifastacus leniusculus* (Dana), to elevated temperatures. T Am Fish Soc 104 (2), 374–387.

Bishai, H. M., 1960. Upper lethal temperatures for larval salmonids. J Cons 25 (2), 129–133.

Brett, J. R., 1952. Temperature tolerance in young pacific salmon, genus *Oncorhynchus*. J Fish Res Board Can 9 (6), 265–323.

Cook, A. M., Duston, J., Bradford, R. G., 2006. Thermal tolerance of a northern population of striped bass *Morone saxatilis*. J Fish Biol 69 (5), 1482–1490.

Doudoroff, P., 1942. The resistance and acclimatization of marine fishes to temperature changes. I. experiments with *Girella nigricans* (Ayres). Biol Bull 83 (2), 219–244.

Edsall, T. A., Colby, P. J., 1970. Temperature tolerance of young-of-the-year cisco, *Coregonus artedii*. T Am Fish Soc 99 (3), 526–531.

Elliott, J. M., Elliott, J. A., Allonby, J. D., 1994. The critical thermal limits for the stone loach, *Noemacheilus barbatulus*, from three populations in north-west England. Freshwater Biol 32 (3), 593–601.

Elliott, J. M., Klemetsen, A., 2002. The upper critical thermal limits for alevins of Arctic charr from a norwegian lake north of the Arctic circle. J Fish Biol 60 (5), 1338–1341.

Gatz, A. J., Kennedy, V. S., Mihursky, J. A., 1973. Effects of temperature on activity and mortality of the scyphozoan medusa, *Chrysaora quinquecirrha*. Chesap Sci 14 (3), 171–180.

Hartwell, S. I., Hoss, D. E., 1979. Thermal shock resistance of spot (*Leiostomus xanthurus*) after acclimation to constant or cycling temperature. T Am Fish Soc 108 (4), 397–400.

Hernández, M., Bückle, L. F., 2002. Temperature tolerance polygon of *Poecilia sphenops* valenciennes (pisces: Poeciliidae). J Therm Biol 27 (1), 1–5.

Hokanson, K. E. F., McCormick, J. H., Jones, B. R., 1973. Temperature requirements for embryos and larvae of the northern pike, *Esox lucius* (Linnaeus). T Am Fish Soc 102 (1), 89–100.

Kennedy, V. S., Mihursky, J. A., 1971. Upper temperature tolerances of some estuarine bivalves. Chesap Sci 12 (4), 193–204.

Kenny, R., 1969. Temperature tolerance of the polychaete worms *Diopatra cuprea* and *Clymenella torquata*. Mar Biol 4, 219–223.

Kita, J., Tsuchida, S., Setoguma, T., 1996. Temperature preference and tolerance, and oxygen consumption of the marbled rockfish, *Sebastiscus marmoratus*. Mar Biol 125, 467–471.

Macey, D. J., Potter, I. C., 1978. Lethal temperatures of ammocoetes of the southern hemisphere lamprey, *Geotria australis* Gray. Environ Biol Fish 3, 241–243.

McCormick, J. H., 1978. Effects of temperature on hatching success and survival of larvae in the white Bass. Prog Fish Cult 40 (4), 133–137.

Menasveta, P., 1981. Lethal temperature of marine fishes of the Gulf of Thailand. J Fish Biol 18 (5), 603–607.

Otto, R. G., 1973. Temperature tolerance of the mosquitofish, *Gambmia affinis* (Baird and Girard). J Fish Biol 5 (5), 575–585.

Otto, R. G., Kitchel, M. A., O'Hara Rice, J., 1976. Lethal and preferred temperatures of the alewife (*Alosa pseudoharengus*) in lake michigan. T Am Fish Soc 105 (1), 96–106.

Paul, J. D., 1980. Salinity-temperature relationships in the queen scallop *Chalamys operculatis*. Mar Biol 56, 295–300.

Paul, R. J., Lamkemeyer, T., Maurer, J., Pinkhaus, O., Pirow, R., Seidl, M., Zeis, B., 2004. Thermal acclimation in the microcrustacean (*Daphnia*): a survey of behavioural, physiological and biochemical mechanisms. J Therm Biol 29 (7-8), 655–662.

Procarione, L. S., King, T. L., 1993. Upper and lower temperature tolerance limits for juvenile red drums from Texas and South Carolina. J Aquat Anim Health 5 (3), 208–212.

Rajaguru, S., 2002. Critical thermal maximum of seven estuarine fishes. J Therm Biol 27 (2), 125–128.

Rajaguru, S., Ramachandran, S., 2001. Temperature tolerance of some estuarine fishes. J Therm Biol 26 (1), 41–45.

Rantin, F. T., Petersen, J. A., 1985. Thermal tolerance of South American cichlid, *Geophagus brasiliensis*. Rev Hydrobiol Trop 18 (3), 221–226.

Sullivan, K., Martin, D. J., Cardwell, R. D., Toll, J. E., Duke, S., 2000. An analysis of the effects of temperature on salmonids of the Pacific Northwest with implications for selecting temperature criteria. Tech. rep., Sustainable Ecosystems Institute, Portland, OR, USA.

Tagatz, M. E., 1969. Some relations of temperature acclimation and salinity to thermal tolerance of the blue crab, *Callinectes sapidus*. T Am Fish Soc 98 (4), 713–716.

Wallis, R. L., 1975. Thermal tolerance of *Mytilus edulis* of Eastern Australia. Mar Biol 30, 183–191.

Wallis, R. L., 1977. Some aspects of the thermal tolerance of *Trichomya hirsuta* (mollusca: Bivalvia) of Eastern Australia. Mar Biol 43, 217–223.

Woo, N. Y. S., Fung, A. C. Y., 1980. Studies on the biology of the red sea bream *Chrysophrys major*. I. temperature tolerance. Mar Ecol Prog Ser 3, 121–124.

Zeis, B., Maurer, J., Pinkhaus, O., Bongartz, E., Paul, R. J., 2004. A swimming activity assay shows that the thermal tolerance of *Daphnia magna* is influenced by temperature acclimation. Can J Zool 82 (10), 1605–1613.

Photo: Steve Geelhoed

Chapter 5

The toxic exposure of flamingos to *per*- and polyfluoroalkyl substances (PFAS) from firefighting foam applications in Bonaire

Reprinted from Marine Pollution Bulletin 124(1), Pepijn de Vries, Diana M.E. Slijkerman, Christiaan J.A.F. Kwadijk, Michiel J.J. Kotterman, Leo Posthuma, Dick de Zwart, Albertinka J. Murk and Edwin M. Foekema, (pages 102–111, copyright (2017)), with permission from Elsevier

Abstract

In 2010 an oil terminal next to nature reservation Saliña Goto (Bonaire) caught fire. Firefighting resulted in elevated per- and polyfluoroalkyl substances (PFAS) concentrations in the salt lake. Within months flamingo abundance in Goto dropped to near complete absence. After statistical analysis, rainfall was deemed an unlikely cause for this decline. Toxicological effects on abundance of prey are likely the main cause for the flamingo absence. This reduced PFAS exposure via food and thus risk towards flamingos during the first years after the fires. Although the sediment is still polluted with persistent PFAS, flamingos returned, and started to feed on organisms with PFAS levels that exceed safety thresh-

olds, placing the birds and other wildlife at risk. Monitoring bird populations is advised to assess potential toxic effects on birds and their offspring. This case suggests that applying persistent chemicals to reduce incident impacts may be more harmful than the incident itself.

5.1 Introduction

Washington Slagbaai National Park (WSNP) is a nature reserve wetland that encompasses about 25% of the Caribbean island of Bonaire. It is designated an Important Bird and Biodiversity Area (IBA, code AN009, Fig. 5.1) by Birdlife International. The park includes Saliña Goto and Saliña Slagbaai. They are protected under the Ramsar Convention (Ramsar, 2016) based on their fundamental ecological functions and their economic, cultural, scientific and recreational value. These and other saliñas (high saline inland bays) in the park are important foraging habitats for many birds, such as the Caribbean flamingo (*Phoenicopterus ruber*) (Wells and Debrot, 2008).

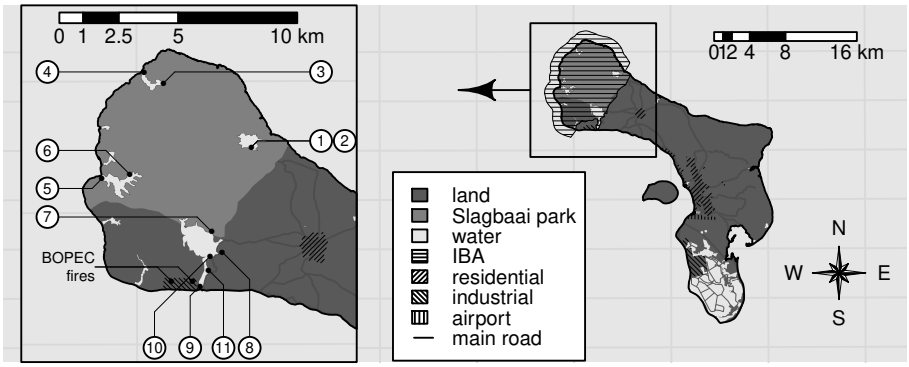


Fig. 5.1: Map of the island Bonaire (right), and a detail of Saliña Goto and the selected monitoring locations (left). Location 1 and 2 are at Saliña Matijs; 3 and 4 are at Saliña Bartol; 5 and 6 are at Saliña Slagbaai; Locations 7, 8, 9, 10 and 11 are at Saliña Goto. Important Bird Area (IBA) is only shown in the overview on the right and not in the detailed map on the left. The map is based on data from **www.openstreetmap.org** combined with data from **dcbd.services.geodesk.nl/geoserver/web**.

Bonaire is of global importance for water bird populations, including the Caribbean flamingo. Wells and Debrot (2008) estimated that the number of Caribbean flamingos fluctuates between 1,500 and 7,000 breeding individuals, mostly with a multi-year estimated average of approximately 5,000. Next to breeding, the Caribbean flamingo uses Bonairean saliñas to forage (Rooth, 1965; Wells and Debrot, 2008).

The BOnaire PEtroleum Corporation (BOPEC) facility (Fig. 5.1) is situated in close proximity of this highly valued nature area, especially Saliña Goto. BOPEC runs a fuel oil storage and shipment terminal owned by the Venezuelan oil company Petróleos de Venezuela S.A. (PDVSA). A channel connects Saliña Goto with the sea via a partially permeable structure. It passes the facility at a distance of 30-50 m and it allows limited tidal influences to the lake (Buitrago et al., 2010). Daily operations of BOPEC started in 1975, and are still ongoing. The Caribbean flamingo commonly forages in Saliña Goto despite BOPEC's operations.

During a period of heavy rains and a thunderstorm, thunderbolts set two petrochemical storage tanks on fire on the 8th of September, 2010. One tank contained approximately 14,500 m^3 crude oil and another tank approximately 22,300 m^3 naphta (Joustra et al., 2011). During the multi-day fire six types of firefighting foams with a total estimated volume of approximately 145,000 L were applied (Joustra et al., 2011).

The following observations of immediate effects were reported by Mooij et al. (2011). The fires caused a few days of emissions of vast amounts of black smoke and soot which were deposited all over Bonaire, but mainly in the north including Washington Slagbaai Park. Mooij et al. (2011) reported declining soot deposition with increasing distances to the fires. In the following rain period soot was flushed from plants and soil into small water bodies, streams and saliñas. Park rangers reported that the flamingo foraging behaviour changed in the days after the fires. Concerns grew about direct and indirect ecotoxicological impact of the fires and firefighting foams, because of reports of dead prey species (brineflies), altered foraging behaviour and dropped flamingo counts after the events (Mooij et al., 2011).

An initial quick theoretical hazard and risk assessment was performed by RIVM (Dutch National Institute for Public Health and the Environment) during and shortly after the fires ended (Mooij et al., 2011). Based on expert knowledge on oil fires and firefighting foam constituents, several compounds were expected to be emitted: polycyclic aromatic hydrocarbons (PAHs, from burned oil); per- and polyfluoroalkyl substances (PFAS) as active foam forming agents; and metals due to the corrosion of facility building materials. The assessment by RIVM indicated that specifically the persistent hazardous chemicals (PFAS) in the firefighting foams formed the major concern.

Indeed PAHs and PFAS were found in various matrices all over Bonaire and were particularly high within the national park. PFAS levels in sediment and water sampled days after the fires were inversely related with distance from the BOPEC site (Mooij et al., 2011). The levels of especially PFAS gave rise to concerns for adverse environmental impact (Mooij et al., 2011). PAH levels correlated with PFAS concentrations, but were not found at levels posing concern for an environmental risk (Mooij et al., 2011). The soot deposition did not result in elevated concentrations of heavy metals (Mooij et al., 2011).

Over a period of approximately four months after the fires, a further reduction in flamingo abundance at Saliña Goto was reported by local observers (Mooij et al., 2011). For multiple years flamingos were not or only incidentally seen.

The absence of the flamingos in Saliña Goto seemed to be triggered by the fire-incident. However, in the same period the island of Bonaire experienced unusually intensive rainfall events. A correlation between rainfall, water level and flamingo abundance has been suggested before at lagoons in the GalÃąpagos Islands (Vargas et al., 2008). Ecological explanations given for these observations were changes in prey density (dilution) and distribution (Tripp and Collazo, 2003) and too deep water for flamingos to use (Kushlan, 1978). Also on Bonaire, effects of rainy events on flamingo abundance have been reported. Before, it has been shown that the number of water birds drops dramatically in rich saliñas (Goto, Slagbaai and Matijs) and increases enormously in other northern saliñas in Bonaire, when through excessive precipitation the water level exceeds a certain limit (Simal et al., 2011). The exact movements of the flamingos within the island of Bonaire and to-and-from mainland Venezuela are however poorly known (Wells and Debrot, 2008). Flamingo migration is affected by shifting food availability and availability of breeding sites (Baldassarre and Arengo, 2000; Del Hoyo et al., 1992; Elphick et al., 2001; Sprunt, 1975). In addition, juvenile birds tend to follow their parents, and copy their routes and feeding places (BirdLife International, 2016).

This study aims to assess the likelihood that the consequences of the fires including PFAS exposure are responsible for the long term flamingo decline in Saliña Goto. Therefore, an environmental risk assessment was performed, taking into account toxicological, environmental and ecological conditions and their dynamic evolution over time since the fires.

5.2 Method and materials

The study consisted of various methods for post-incident data collection, ranging from sampling and subsequent chemical analyses to collations of existing pre- and post-incident data, such as long-term flamingo counts. Earlier studies have shown that post-incident data collection never is ideally complete regarding ecological, chemical and other baseline data (Posthuma et al., 2014).

5.2.1 Sampling locations

Sampling took place in the years 2010, 2012, 2013 and 2015, facilitated by local authorities and area managers and rangers of Slagbaai National Park Bonaire (STINAPA). Overall, sampling sites at the impacted Saliña Goto were selected in a distance gradient from the BOPEC in the west, up to the north eastern border of the Saliña (Fig. 5.1). In addition, three reference saliñas were selected

and sampled, based on the observation that these were not or far less impacted by PFAS (Mooij et al., 2011). These reference saliñas were also selected because flamingos are also known to forage there. Table 5.1 summarises selected locations, sample type and endpoints analysed over a six year research period.

5.2.2 Sampling of sediment, water and biota for chemical analyses

Water and sediment samples were taken in duplicate in 2010 and 2012; while in 2015 sediment samples were not replicated. Water samples were taken prior to sediment samples, to avoid suspension of particulate matter into the water column. Water samples were taken by submersing a 1 L plastic beaker, attached to a 2 m pole (to avoid suspension of particulate matter by wading through the water) to approximately mid water depth, and subsequently filling two glass 1 L bottles. After water sampling, two 250 mL sediment samples were taken by carefully scraping a 1-2 cm sediment layer off of the bottom surface. Collected sediments were transferred to a 250 mL glass container (plastic in 2015). All samples were transported in a cool box to the local lab and then stored in a refrigerator (4 °C), until transport to the Netherlands for analysis.

Depending on the presence and catchability of fish, *Cyprinodon* sp. were caught using a macrofauna net at five locations in October 2012 and at one location in May 2013 (Table 5.1). Artemia (brineshrimp) were collected in December 2015 using a plankton net. Fish and Artemia were stored frozen in zip-lock bags until analysis for PFAS levels.

5.2.3 Sampling of benthic community and quantifying abundance

As a measure of flamingo food availability, the benthic community of the saliñas was sampled twice: in October 2012 and May 2013 (Table 1). In October 2012 a macrofauna net with a width of 30 cm and mesh size of 2 mm was used. A selection of locations was sampled in May 2013, and a macrofauna net with a width of 40 cm and mesh size of 1 mm was used.

In order to sample the benthic community, the net was dragged approximately 2 cm into the sediment over a distance of 5 m. After washing out the finest clay and sand particles, the benthic animal samples were stored in a polyethylene container, and preserved with a solution of 6-10% buffered formaldehyde in local water.

In the lab the samples were rinsed with seawater over a sieve with a mesh size of 0.5 mm. Biota specimens were sorted and by means of a stereomicroscope identified up to the highest taxonomic level possible, but at least to class. In case the sample was too large to handle due to large amounts of debris

Table 5.1: Sampling overview (what is sampled when, where and for what analysis) and distances to the fires.

		September 2010		October 2012				May 2013		December 2015	
		chemistry		chemistry			benthos	chemistry	benthos	chemistry	
Location	Distance to fires (km)	sediment	water	sediment	water	biota[14]		biota[14]		sediment	biota[15]
1 Matijs	6.4		X	X	X	X	X			X	
2 Matijs	6.4			X	X		X				
3 Bartol	8.4			X	X		X		X	X	
4 Bartol	9.0			X	X	X	X				
5 Slagbaai	5.5			X	X	X	X		X		
6 Slagbaai	5.0		X	X	X		X		X		
7 Goto	2.5	X	X	X	X	X	X		X	X	
8 Goto	1.9	X	X	X	X		X		X	X	
9 Goto	0.8	X	X	X	X	X	X	X	X[16]	X	
10 Goto	1.6			X	X		X			X	X
11 Goto	1.2			X	X						

[14] Fish.
[15] Artemia.
[16] Is sampled, but not successfully.

and organic matter, the sample was homogenised and subsampled by a factor of 2 (location 2 (Matijs), location 6 (Slagbaai), location 10 (Goto) or 32 (location 9 (Goto)) and then analysed. The corresponding sub factor was applied to correct for subsampling.

5.2.4 Chemical analyses

PFAS analyses were performed at two laboratories. The 2010 samples were analysed at Ducares, 2012 and 2013 samples at Wageningen Marine Research. Both laboratories hold ISO/IEC 17025 accreditation, and procedures followed similar analytical principles. Detailed description on the analytical methods is presented in Appendix 5.A.1. PFPA, PFHxA, PFHpA, PFNA, PFDcA, PFDoA, PFTrA, PFBS, PFHxS, PFOS and PFDS (see Appendix 5.A.3 Table 5.5 for full names of chemicals) were analysed in all samples. Additional PFAS substances (PFBA, PFOA, PFUnA, PFTeDA, PFHxDA, PFODA and PFHpS; see Supplemental Info for full names of chemicals) were analysed in specific samples, in specific matrices, and in specific sampling years.

5.2.5 First tier ecotoxicological risk assessment

Although a large set of PFAS was measured, the first tier risk assessment focused on perfluorooctanyl sulphonic acid (PFOS), because it is one of the major PFAS type components in the fire-fighting foams used (Mooij et al., 2011). Furthermore, PFOS is listed as a priority substance under various international regulations (e.g., OSPAR (OSPAR, 2013); and the European Water Framework Directive: EC (2011)). Consequently, exposure and effect levels relevant for the interpretation of exposure data from this study are available for this particular substance.

Also, environmental quality standards (EQSs) are available for PFOS as proposed by De Zwart et al. (2012) and listed in Table 5.2 (Anonymous, 2011; Bodar et al., 2011; De Zwart et al., 2012; Moermond et al., 2010; Mooij et al., 2011). These EQS-values are based on bioconcentration and ecotoxicity data of sufficient study quality and were derived according to the EU Technical Guidance Document (ECHA, 2003) on the derivation of environmental quality standards. Ecotoxicological impacts may occur due to direct exposure via water or sediment, as well as indirect exposure via the food-chain. The EQS derivation process eventually yielded an EQS for both direct (via e.g. water) and secondary poisoning (via food), abbreviated as $EQS_{eco,water}$ and $EQS_{sp,water}$ respectively. These EQSs are also available for sediments ($EQS_{eco,sediment}$ and $EQS_{sp,sediment}$).

In practice, such first tier risk assessment utilises the concept of risk quotient (RQ). An RQ is simply a measured concentration divided by its corresponding

Table 5.2: Environmental quality standards (EQS) as used for the first tier risk assessment. Different EQS values are derived to either protect against direct ecological effects (eco) and secondary poisoning (sp). The assessment factor (AF) applied in the EQS is also listed. EQS values for sediments are standardised for an organic content of 10%.

EQS type	Value	Reference
eco, water	23 ng/L	De Zwart et al. (2012); Anonymous (2011); Moermond et al. (2010)
sp, water	2.6 ng/L	De Zwart et al. (2012); Moermond et al. (2010)
eco, sediment	10 μg/kg d.w.	De Zwart et al. (2012); Bodar et al. (2011)
sp, sediment	3.2 μg/kg d.w.	De Zwart et al. (2012); Bodar et al. (2011)

EQS value. RQ values lower than one are considered sufficiently safe for the ecotoxicity endpoint under consideration. Values equal to or larger than one indicate that effects cannot be not excluded. Whether or not effects occur strongly depends on local conditions influencing the bioavailability of the compounds; other compounds that are present; and the vulnerability of the ecosystem under investigation. In this study sediment EQS values were standardised to a 10% organic content. This standardisation is used to correct for differences in bioavailability of compounds which is affected by the organic content. When the organic content was not measured it was assumed to be 10%.

5.2.6 Rainfall data

Historical daily time series of precipitation levels (mm/day) were obtained from the weather station at Flamingo Airport on Bonaire (World Meteorology Organisation identifier 78990, downloaded from `www.tutiempo.net` on October 4th, 2016). This station is situated roughly 15 km southeast of Saliña Goto at an elevation of 6 m. The average precipitation in the 14 days prior to each date on which flamingo birds were counted was determined (Appendix 5.A.2, Fig. 5.9). By taking the average precipitation of the 14 days before bird counting events it is assumed that the spatiotemporal variation in rain fall and its potential effect on water level among saliñas is smoothened.

5.2.7 Flamingo count data

Bird counts were, and still are, conducted every middle of the month at fixed observation points across the island at flamingo breeding and foraging sites. Counting is conducted at a minimal disturbance distance during the morning. Two persons count independently and the two counts are averaged. Bird counting data have been reported since June 1981, and resulted in an extensive database

of flamingo numbers.

Not all northern locations (Washington Slagbaai) are fully covered by the data set, starting at the year 1981. Therefore, the period starting at 19 November 1996 up until 12 August 2016 was selected from the data set in the present study. This resulted in 203 monthly records between November 1996 and August 2016 for each location. Flamingo abundance was determined for the locations Goto (site: Goto east) and all other northern saliña locations combined (excluding Saliña Goto).

5.2.8 Assessment of flamingo population development based on bird counts

The development of bird counts over time is analysed with a Generalised Additive Model (GAM, Wood (2006)). A GAM fits a smoothing function of potentially relevant explanatory variables (in this case time, season and precipitation) to the census data in order to describe the number of birds in relation to these variables.

Rainfall data and flamingo counts are known to co-vary on long term and seasonal time scales (Vargas et al., 2008). The precipitation data were used in the GAM analysis of bird counts as rainfall extremes (i.e., surplus or shortage of rainfall). The model was split into two stages: firstly, yearly and seasonal patterns were filtered out of the precipitation data with a GAM model (Appendix 5.A.2, Figs. 5.9 and 5.10); secondly, residuals from the first model were used as a proxy for precipitation extremes and form the input for the main model explaining bird counts. The GAM analysis was used to analyse fluctuations of bird counts in time, but also to evaluate the effect of surplus rainfall.

All GAM analyses were implemented using the Mixed GAM Computation Vehicle (mgcv) package (Wood, 2011) in R (R Core Team, 2016). First, the precipitation data was fourth root transformed and a Gaussian distribution family with identity link (Wood, 2006) was applied. The transformation was applied as none of the standard link functions or distribution families would properly describe a fit to the precipitation data. Nth root transformation has also been suggested by Stidd (1970) in order to adjust for skewness in precipitation data. The model is formulated as:

$$I\left(\mu_{precip\_transformed}\right) \sim s\left(year\right) + s_{cc}\left(month\right) \tag{5.1}$$

Where I is the identity link function and s is the spline smoothing function and s_{cc} is a cyclic cubic regression spline. Here, the first smoother is a function of the year and the second is a function of the month. Furthermore, $\mu_{precip\_transformed}$ represents the response as the fourth root transformed precipitation. Residuals from this model (i.e., the difference between the model fitted predictions and the

observed transformed precipitation) are used as input for the GAM model of bird counts:

$$r_{precip}(year, month) = observed\_precip(year, month)^{1/4} - predicted\_precip\_transformed(year, month) \quad (5.2)$$

The main model was then fitted to describe the statistical association between the bird counts and the potentially explanatory variables: year, month (using a cyclic smoother) and precipitation residuals obtained from the first step. With this GAM a quasi-Poisson distribution family and a log link is assumed for the count data. This model was fitted separately to both the flamingo count data of Saliña Goto and that of the other northern locations (excluding Saliña Goto) and is formulated as:

$$log(\mu_{bird\_counts}) \sim s(year) + s_{cc}(month) + s(r_{precip}) \quad (5.3)$$

5.3 Results

5.3.1 PFAS levels in water and sediment

Concentrations of each individual PFAS measured in water and sediment are shown in Fig. 5.2 and Fig. 5.3 and listed in Appendix 5.A.3 (Tables 5.7 and 5.8). Water concentrations of PFAS at sampling sites that were ≥ 5 km away from the fires (Table 5.1) were mostly below the limit of quantification (with the highest limit of 5 ng/L for PFODA at location 1, Saliña Matijs), or slightly above that in only some of the samples (where 3.5 ng/L was the highest concentration of PFOA at location 2, Saliña Matijs, see Appendix 5.A.3, Tables 5.7 and 5.8). PFAS concentrations at Saliña Goto were generally higher than those at more distant (reference) locations.

The most notable PFAS concentrations in a reference saliña occurred at sampling location 6 (Saliña Slagbaai) where the water concentration of PFOS was 8.4 ng/L and PFPA water concentration was 7.0 ng/L in 2010 (Fig. 5.2 and Appendix 5.A.3, Tables 5.7 and 5.8). In later years, these concentrations were similar or have decreased (Fig. 5.2 and Appendix 5.A.3, Tables 5.7 and 5.8). Sediment concentrations were all below the limit of quantification for the reference locations in all sampling years, except for samples taken at location number 3 (Bartol) where the concentration in 2015 was slightly above the limit of quantification (Fig. 5.3).

In 2010 the water column at Saliña Goto, concentrations of PFPA, PFPHxA, PFBS, PFHxS and PFOS were mostly above 100 ng/L, where PFHxS concentrations were highest. Highest PFHxS concentrations were found in the northern

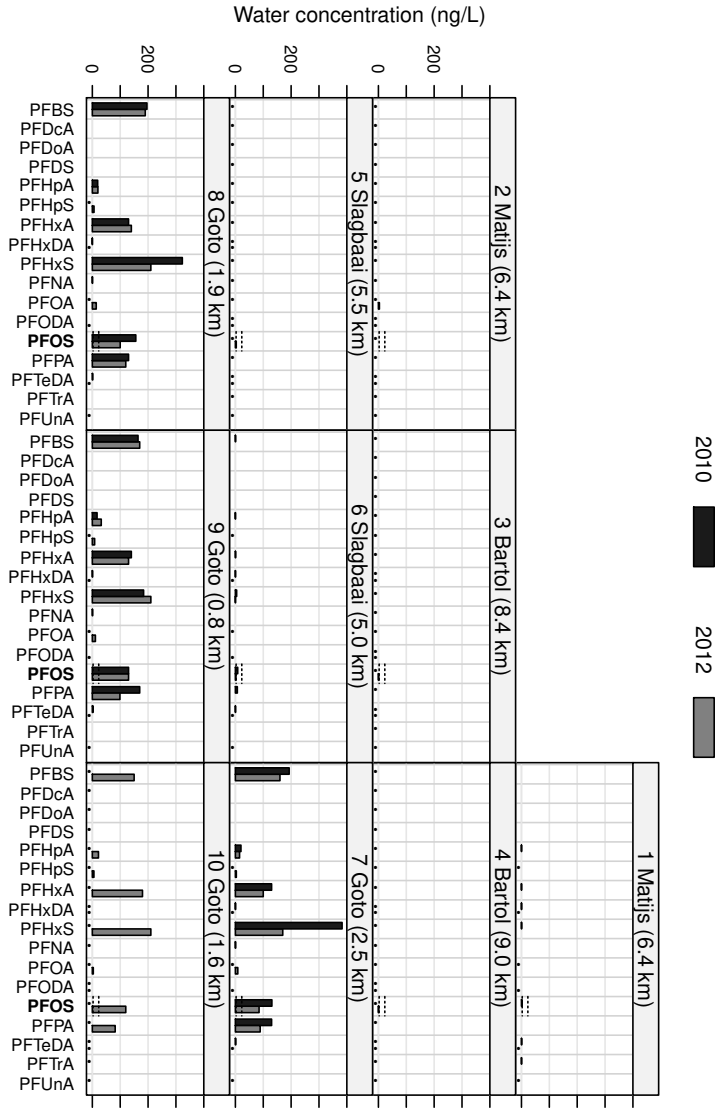


Fig. 5.2: PFAS concentrations in water samples collected in the years 2010 and 2012. Concentrations below the limit of quantification are not shown (see Appendix 5.A.3, Tables 5.7 and 5.8 for these limits). Substances that are not measured at a specific location in a specific year are marked with a dot below the origin. The environmental quality standards used in this study ($EQS_{eco,water}$ and $EQS_{sp,water}$, Table 5.2) for PFOS are shown as dotted horizontal lines. Above each panel between parentheses is the distance of the sampling locations to the fires.

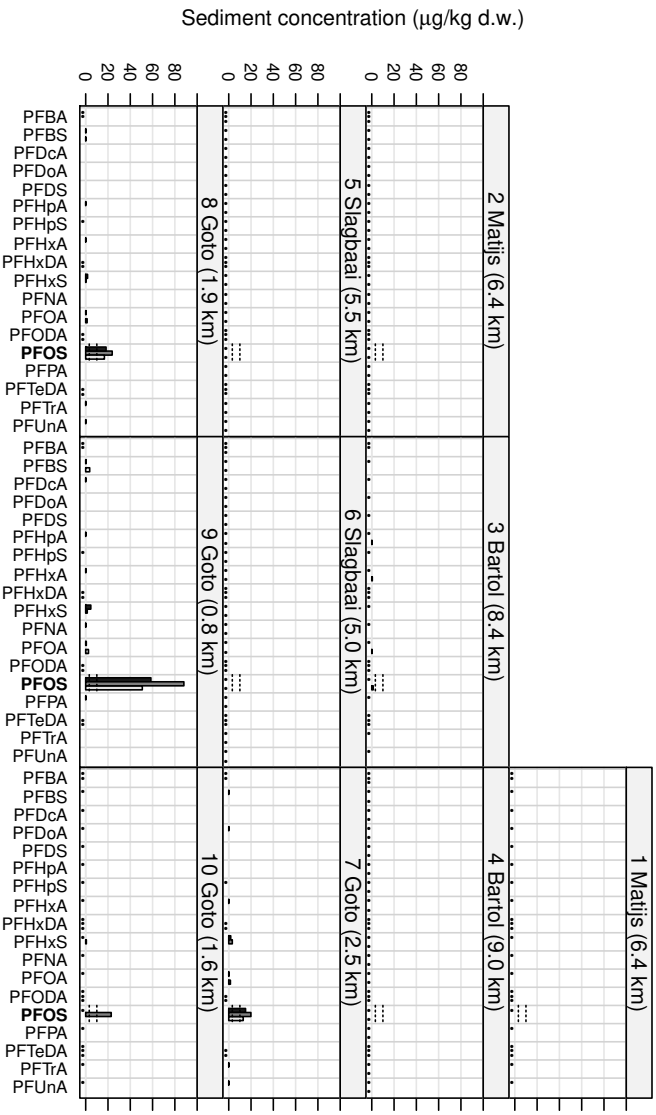


Fig. 5.3: PFAS concentrations in sediment samples collected in the years 2010 and 2012. Concentrations below the limit of quantification are not shown in this plot (see Appendix 5.A.3, Tables 5.7 and 5.8 for these limits). Substances that are not measured at a specific location in a specific year are marked with a dot below the origin. The environmental quality standards used in this study ($EQS_{eco,sediment}$ and $EQS_{sp,sediment}$, Table 5.2) for PFOS are shown as dotted horizontal lines. Above each panel between parentheses is the distance of the sampling locations to the fires.

parts of Saliña Goto, whereas the reverse is true for PFOS, concentrations were highest at the southern locations (nearest the site of the fire). In sediments, most PFAS concentrations were below limit of quantification at Saliña Goto, except for PFOS, which was found at levels ranging from 7.2 up to 783 μg/kg dry weight.

5.3.2 First tier risk assessment

Logically, the Risk Quotients (RQ) for PFOS showed similar patterns as the concentration measurements (Table 5.3). For the more distant reference locations, only two samples had PFOS concentrations that exceeded their compound-specific EQS_{eco} and EQS_{sp} (Table 5.2), with an RQ of 1.2 and 3.2 for location 5 (Saliña Slagbaai) in water in 2012, and location 6 (Saliña Slagbaai) in water in 2010, respectively. At Saliña Goto the EQS-values were exceeded for the two exposure routes and endpoints (i.e., RQ > 1), which indicates that the exposure conditions in the Saliña are not safe, i.e., direct effects and effects of secondary poisoning cannot be excluded. Highest RQ-values were found in samples from 2012 at Saliña Goto 9 in sediment with a value of 8.8 for EQS_{eco} and in 2010 at Saliña Goto location 8 with a value of 60 in water for EQS_{sp}. The RQ for secondary poisoning is obviously higher than that based upon the direct exposure via water, triggering concerns for insufficient safety for food-chain mediated effects more than for direct effects.

Risk levels for PFOS are elevated in the earliest samples and persisted for the period covered by the sampling (2010 up to 2015).

5.3.3 PFAS levels in biota

PFOS concentrations in fish at Saliña Goto in 2012 and 2013 ranged from 72 to 450 μg/kg wet weight while in Artemia averaged at 60 μg/kg wet weight in 2015 (Table 5.4). PFOS levels in fish at reference locations were at least a twentyfold lower ranging between 0.7 and 3.3 μg PFOS/kg wet weight. Other PFAS levels in biota were generally lower and weakly correlate to PFOS levels (Table 5.4).

5.3.4 Abundance of benthic organisms

Overall, mainly larvae were found from various undefined species and crustaceans, but not much other (*epi*)benthic life was found at Saliña Matijs in 2012 (Fig. 5.4). In the Saliña Bartol sampling locations mainly oligochaetes were found in densities up to 63 individuals per m^2). In Saliña Slagbaai brinefly was the dominant species in 2012 and Artemia in 2013. Both species were found in high numbers, especially in 2013 with 71 brinefly larvae and 726 Artemia per m^2). Number of individuals differed largely between years and locations, and number of individuals at reference locations Bartol 3 (2013) and Slagbaai 5 (2012) are low.

Table 5.3: Risk quotients (RQs) for each sampled location and year in either sediment or in water. Risk quotients equal to or greater than 1 indicate a risk and are printed in bold face. B.d. = below detection; u.r. = unclear result (i.e., the detection limit is greater than the EQS); sed. = sediment; wat. = water.

EQS type	Year	1 Matijs	2 Matijs	3 Bar-tol	4 Bar-tol	5 Slag-baai	6 Slag-baai	7 Goto	8 Goto	9 Goto	10 Goto	11 Goto
sed., eco	2010							**1.5**	**1.8**	**5.8**		
sed., eco	2012[17]	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	**2.3**	**2.8**	**16.8**	**2.5**	**1.0**
sed., eco	2015	b.d.		0.1				**1.3**	**1.7**	**5.1**	u.r.	
sed., sp	2010							**4.7**	**5.7**	**18.3**		
sed., sp	2012[17]	b.d.	b.d.	b.d.	b.d.	b.d.	b.d.	**7.1**	**8.7**	**52.4**	**7.8**	**3**
sed., sp	2015	b.d.		0.4				**4.1**	**5.2**	**15.9**	u.r.	
wat. eco	2010	0.1					0.4	**5.7**	**6.8**	**5.7**		
wat., eco	2012	b.d.	b.d.	0.03	0.03	0.1	0.1	**3.7**	**4.3**	**5.7**	**5.2**	**5.2**
wat., sp	2010	0.7					**3.2**	**50.4**	**60**	**50**		
wat., sp	2012	b.d.	b.d.	0.3	0.3	1.2	0.6	**32.7**	**38.5**	**50**	**46.2**	**46.2**

[17] standardised to 10% organic content and excludes DS in water (salts). These corrections are not applied to other years.

Table 5.4: Concentrations of PFASs in biota at the sampled locations (in μg/kg w.w., unless mentioned otherwise). Note that most concentrations are based on wet weight. For Artemia, the concentrations is also determined based on dry weight.

Species type	Year	Location	Multiple[18]	PFBA	PFHxA	PFHpA	PFOA	PFNA	PFBS	PFHxS	PFHpS	PFOS
Fish	2012	1 Matijs					4.9	0.9	< 0.5	< 0.5	< 0.5	1.5
Fish	2012	4 Bartol					< 0.2	< 0.2	< 0.2	< 0.2	< 0.2	0.7
Fish	2012	5 Slagbaai					11	2.4	< 0.09	< 0.1	< 0.1	3.3
Fish	2012	7 Goto					0.3	0.08	0.4	5	0.5	72
Fish	2012	9 Goto					120	29	< 1.0	18	1.5	450
Fish	2013	9 Goto					64	11	< 0.4	2.3	0.6	160
Artemia	2015	10 Goto	< 0.8[19]	< 4.5[19]	18[19]	4.5[19]	15[19]	< 0.8[19]	< 0.8[19]	< 0.8[19]	< 0.8[19]	451[19]
Artemia	2015	10 Goto	< 0.1	< 0.6	2.4	0.6	2.0	< 0.1	< 0.1	< 0.1	< 0.1	60

[18] PFPeA, PFDcA, PFUnA, PFDoA, PFTrA, PFTeA and PFDS
[19] Concentration in μg/kg d.w.

Compared to the reference locations, in general less benthic life was found at the locations in Saliña Goto. At Goto 7 only 3 individual brinefly larvea per m^2 were found in 2013, while in 2012 the samples did not contain a single macroinvertebrate. In samples of Goto 8 no macroinvertebrate were found, and only 1.3 crustaceans per m^2 were found at Goto 10 in 2012. Goto 9 samples showed higher numbers up to 64 unspecified larvea per m^2 and 21.3 brinefly per m^2, with a subsampling factor 32. A cyprinidae fish was also caught (not shown in the figure) in 2012.

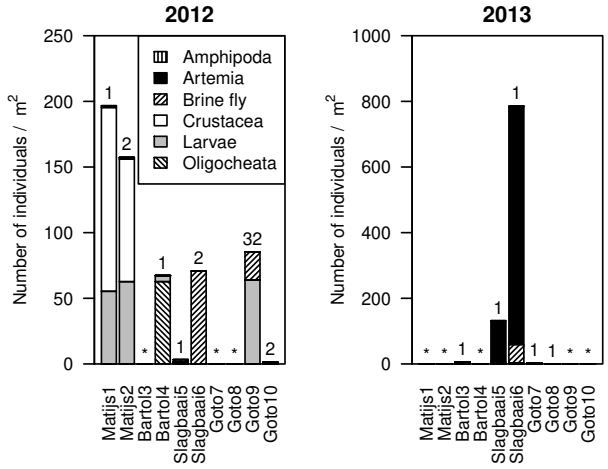


Fig. 5.4: Benthic invertebrate species density (number of individuals per m^2) found in each of the benthic samples. Numbers above each bar indicate the subsampling factor (i.e., the factor by which the volume of the main sample was subbed before further analysis). Asterisks indicate locations that were not sampled or could not be analysed. Note that the y-scales are different for 2012 (left) compared to 2013 (right).

5.3.5 Flamingo abundance

The number of flamingos in Saliña Goto, and the northern Saliñas excluding Goto fluctuated over time (Fig. 5.5 and Fig. 5.6). Flamingo numbers in Saliña Goto steadily fluctuate around 400 birds until the fires, after which they were reduced to nil or a few dozen at most up to 2015. Then flamingo numbers rose sharply, to values exceeding the long-year observation average. In other northern saliñas number of birds also steadily fluctuated around 250 birds, with a slight increase of counted numbers after the fires.

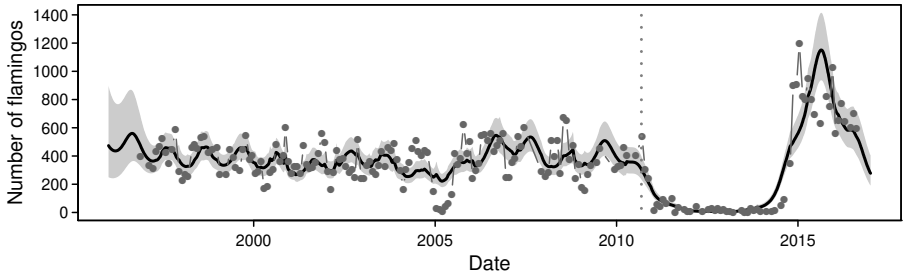


Fig. 5.5: Observed number of birds (dark grey dots) at Saliña Goto and the smoothed Generalised Additive Model trend (black line with 95% confidence intervals shown as grey bands). Vertical dashed line indicates the moment of the fires at the oil terminal.

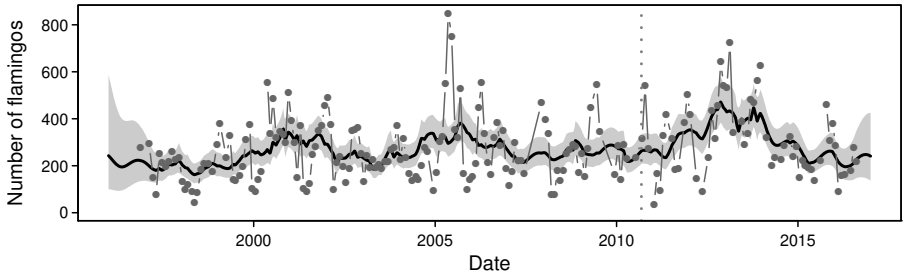


Fig. 5.6: Number of birds (dark grey dots) at northern sites (excluding Saliña Goto) and the smoothed Generalised Additive Model trend (black line with 95% confidence intervals shown as grey bands). Vertical dashed line indicates the moment of the fires at the oil terminal.

5.3.6 Flamingo abundance in relation to rainfall (GAM model)

Bonaire experienced heavy rainfall in the period during and shortly after the fires (Appendix 5.A.2, Figs. 5.9 and 5.9), followed by an extensive period of drought. The question is whether the heavy rains could have caused the decline in Flamingo numbers. The GAM fit to the bird count data at Saliña Goto (data and model in Fig. 5.5 and Fig. 5.7) showed that the sharp decline after the fires could not be explained by the general seasonal variation in counts, according to the results of the parameter estimation. Nor could the sharp decline be attributed to the excessive rainfall after the fires, as the explanatory variable representing the surplus of rainfall shows a nearly horizontal line with wide confidence intervals (on the right in Fig.e 5.7). An increase of modelled flamingo numbers in Saliña Goto was observed since fall 2014, including a peak in numbers in 2015. As the surplus rainfall did not explain any of the observed variation (Fig. 5.5, Fig. 5.7, Fig. 5.6 and Fig. 5.8), a slightly better fitting model was obtained when this explanatory variable is left out of the model. Results of this model are very similar to the figures shown and therefore not elaborated here.

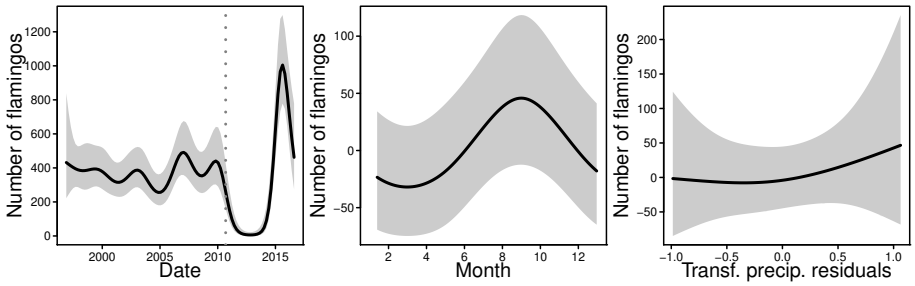


Fig. 5.7: Generalised Additive Modelling results split per explanatory variable for Saliña Goto: overall time on the left, the month (in the middle fitted as a cyclical smoother) and the transformed precipitation residuals (i.e., representing the surplus of rainfall in the two week period before the bird count event). Vertical dashed line indicates the moment of the fires at the oil terminal.

The same model fit to the data of all northern locations (excluding Saliña Goto, Fig. 5.6) also showed no indications of association between the surplus rainfall and abundances. An increase of bird counts was found after the fires (Fig. 5.8 on the left) which levels off to numbers observed previously starting around the year 2013.

The data analyses imply that there is no indication that declining flamingo numbers at Saliña Goto after the fires are related to the excessive rainfall that occurred.

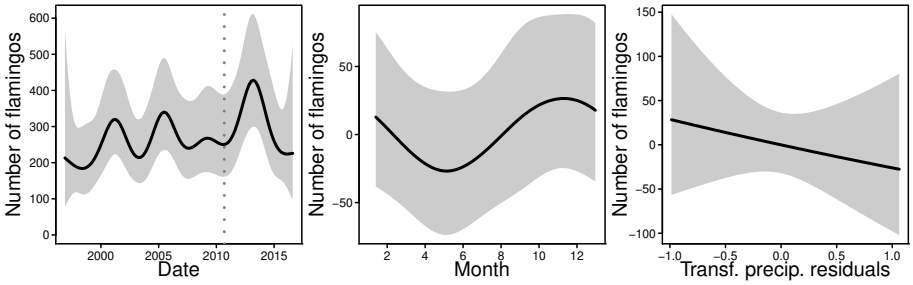


Fig. 5.8: Generalised Additive Modelling results split per explanatory variable for the northern reference locations (excluding Saliña Goto): overall time on the left, the month (in the middle fitted as a cyclical smoother) and the transformed precipitation residuals (i.e., representing the surplus of rainfall in the two week period before the bird count event). Vertical dashed line indicates the moment of the fires at the oil terminal.

5.4 Discussion

The present study aimed to assess the likelihood that the consequences of the fires, including PFAS exposure, are responsible for the flamingo decline in Saliña Goto. The possible role of the ecological factors rainfall and benthic conditions were evaluated as well as the risk posed by PFOS (a major component of the foams used in firefighting) found in sediments, water and biota (food species). The results of Saliña Goto were compared with those from several reference sites.

5.4.1 Does rainfall surplus explain flamingo absence?

Water depth in various lakes and lagoons is known to be an important predictor of the abundance of flamingos (Bucher et al., 2000; Espinoza et al., 2000; Pirela, 2000). In earlier studies rainfall and lagoon water levels were shown to be highly correlated (Vargas et al., 2008) and predictive for flamingo abundance, the predictive power was only 22-50% and varied among lagoons (Vargas et al., 2008). In the present study, a relation between surplus rain and flamingo abundance could not be established. Historical dynamics in surplus rainfall did not indicate prolonged absence of flamingos before. Therefore the evidence suggests that (an) other steering factor(s) must have contributed to the absence of flamingos in Goto in the period after the heavy rains in 2010.

However, the rain could have had an indirect effect, namely transporting deposited PFOS from the surrounding catchment area to the Saliña, and afterwards diluting the PFOS concentrations in combination with the tidal currents. Furthermore, the heavy rainfall and rain water discharge from the catchment area could have influenced the physical and ecological conditions of Saliña Goto

(Buitrago et al., 2010; Simal et al., 2011). Via stratification due to rainfall, possibly resulting in local hypoxic conditions, the circumstances for flamingo prey organisms could have worsened, resulting in increased vulnerability towards PFOS toxicity. These indirect effects on flamingo abundance are not covered by the GAM analysis which is parameterised with direct effects of rainfall surplus.

5.4.2 Water and sediment quality

Water and sediment quality were evaluated with a first tier risk assessment using generic EQSs to study direct effects on the flamingo birds exposed via their prey species, and indirect effects via ecotoxicological effects on the prey species thereby reducing the flamingo food availability.

The sediment water distribution coefficients (Kd) of PFAS increase with chain length, and PFOS has a relatively high Kd compared to other PFAS such as PFPA, PFPHxA, PFBS, PFHxS (Kwadijk et al., 2010). This might explain the higher PFOS sediment concentrations, and lower water concentrations observed (Fig. 5.2 and Fig. 5.3 and Tables 5.7 and 5.8 in the Appendix 5.A.3) compared to the other PFAS. Therefore, PFOS exposure is expected to be one of the most relevant of all the PFAS.

Sediment concentrations of the year 2012 were standardised to a 10% organic content, in order to compare the levels with the same standardisation used in the derivation of the EQS. The organic content was, however, not measured for the samples taken in the other sampling years. The organic content in the sediment from 2012 ranged from 6% up to 12% with one outlier of 30% (Appendix 5.A.3, Table 5.6). Standardisation to 10% organic content therefore resulted in concentrations that ranged from 17% lower up to 67% higher compared to non-standardised concentrations, or 66% lower for the outlier.

The first tier risk assessment for PFOS represents realistic worst-case conditions in a standardised, generic exposure-impact scenario. Thus, samples with an RQ¡1 can be considered safe. In the situation of Saliña Goto, however, more PFAS were present for which the EQS was not available. Additive impact of these substances cannot be excluded. Therefore, our analysis might underestimate the overall impact of PFAS.

The exceedance of EQS for both direct effects observed in both water and sediments at Saliña Goto indicates that effects on one or more species cannot be excluded, including adverse effects on the benthic, highly exposed, food species of the flamingos. The exceedance of the EQS for secondary poisoning in the same Saliña indicates that food for flamingos is contaminated with PFOS to such an extent, that effects on the birds via food once this is present again can still not excluded.

A problem with the generic EQSs is that they are derived from the perspective of European surface water conditions. Although they are legally applicable to

the saliñas, due to the political status of Bonaire as a special municipality of the Netherlands, these conditions are different from those in the saliñas. Saliñas are defined by highly saline and tropical conditions and a highly specific food-chain. Currently, no specific information is available and first-tier risk assessment cannot be easily adapted to these conditions. However, extreme abiotic conditions in saliñas can presumably pose multiple stresses upon organisms, which can increase their sensitivity (Bednarska et al., 2013; Heugens et al., 2001). Aspects listed above suggest that the limitations in the data set are not expected to result in an overestimation of ecological risk.

5.4.3 Was the benthic community affected by PFOS?

There was a clear observation that the fires caused effects on flamingo resources, as dead brine flies were reported directly after the fires (Mooij et al., 2011). A month after the BOPEC fires, Jorcin and Caglierani Casanova (2011) sampled the macrobenthos community across some saliñas. In all saliñas, macrofauna was found, however, in Goto at three out of the seven sampled locations no benthic life was found, three locations with limited numbers (101 up to 152 individuals per m^2), and 1 location with moderate numbers (up to 2323 individuals per m^2 Ephydra (brinefly; part of flamingo diet; Rooth (1965)). All locations of the Saliñas Slagbaai, Bartol and Matijs had a considerable abundance of insects, crustaceans, molluscs and protozoa (total abundance varying 222-254,495 individuals m^2) compared to Saliña Goto. We have no records on the food conditions in Saliña Goto shortly before the fires. However, given that reported flamingo numbers in that period were in the normal range, it can be speculated that food conditions before the fires were better than at the first moment of sampling.

Although the sampling of the benthic community was limited, and should therefore be considered as indicative for the situation, poor benthic life was observed in Saliña Goto compared to reference saliñas two and three years after the fires. An exception is Saliña Goto 9 in the year 2012, where the abundance is relatively high. The uncertainty in this abundance is also high, due to a large subsampling factor (32) applied to that sample. Based on the limited benthic data, large differentiation in abundance among and within saliñas became apparent. Saliña Matijs and Saliña Slagbaai showed highest number of benthic organisms, but also intra-Saliña variance (Fig. 5.4). Saliña Bartol has variable abundance, and densities are in general lower than all other reference locations. These findings relate well to those of Simal (2010); Simal et al. (2011) who reported varying bird diversity among the saliñas, indicating that reference saliñas probably differ in attractiveness. Benthic prey density might be an explanatory factor in this observation.

It is known that PFOS can affect insects via e.g. the moulting cycle, reproduction or survival (e.g. Mommaerts et al. (2011)). Sub lethal effects to damselflies

life cycle was found at levels 10 μg PFOS/L and metamorphosis was indicated to be the most sensitive endpoint (Bots et al., 2010; Van Gossum et al., 2009). This might explain the observation in Saliña Goto that insects and crustaceans have been present at very low densities for some time after the fires. PFOS concentrations in Saliña Goto in water samples were 100 times lower than effect levels in studies of Van Gossum et al. (2009) and Bots et al. (2010). Still, taking into account the long term exposure since the fires, the cumulative stress of additional PFAS and the specific brine circumstances, and the relatively high PFOS levels in the sediment it is possible that contaminant exposure has affected the benthic community.

Based on the first tier risk assessment, there are sufficient arguments to assume that PFOS exposure could have altered the ecological situation especially in Saliña Goto. For the other measured PFAS and various PAHs, the water and sediment concentrations did not exceed the generic protective EQSs (when available, Mooij et al. (2011)), their additive or other mixture effects, however, could not be taken into account. Also several other compounds will have been released during the fire, additionally, compounds present in the firefighting foams may have been transmitted via the soot to the nearby saliñas and added to the toxicity. But not all their identities, concentrations and potential effects are known.

5.4.4 Flamingo-specific refined risk assessment

Measured levels of PFOS in Artemia and fish suggest that PFOS was taken up in biota in 2013 and 2015, and that exposure of flamingos through the food chain occured. In turn, this may lead to adverse effects which we will assess here specifically for flamingos in Saliña Goto.

Newsted et al. (2005) reported avian Toxicity Reference Values (TRV) for PFOS, based on no-observed adverse effect level (NOAEL) obtained on growth and reproduction studies with quail. A TRV is defined as a daily dose of a chemical expressed in milligrams of chemical per kilogram of body weight per day (mg/kg bw/day) and represents a dose associated with NOAEL or lowest-observed adverse effect level (LOAEL). The avian TRV expressed as an average daily intake (ADI) is estimated to be 0.021 mg PFOS/kg body weight/day (Newsted et al., 2005). Dietary intake at or below the TRV is not expected to pose significant risks to avian populations.

The concentration of PFOS in brine shrimp (Artemia) from Saliña Goto in 2015 was 0.451 mg PFOS/kg dry weight (Table 5.4). Artemia is a food source for flamingos (Rooth, 1965) and a flamingo needs about 270 g dry weight of food/day. Assuming foraging flamingos in Saliña Goto solely eat 270 g of Artemia sp. (equalling roughly 135,000 brine shrimps per day (Rooth, 1965)), this would result in an estimated daily intake of 0.12 mg PFOS/day/flamingo. The mean

body weight of Caribbean flamingo is 2.99 ± 0.47 kg, with a range of 2.4-4.2 kg (Lindemann et al., 2016). This results in an estimated daily intake of 0.04 (ranging between 0.029 and 0.051) mg PFOS/kg flamingo/day. This exceeds the TRV for birds by up to approximately a factor 2. Therefore, health effects cannot be excluded. There are indications that they incidentally consume sediment (Rooth, 1965). When large amounts of sediment would be ingested, this would decrease the exposure to PFOS as sediment levels were lower than levels in Artemia. Exposure can be higher when diet is primarily composed of fish, in which the levels were found to be higher (Table 5.4).

Chronic exposure to low levels of various organic contaminants are known to affect eggs shell thickness (Gilbertson and Reynolds, 2010), reproductive behaviour and sexual preference of adult birds (Frederick and Jayasena, 2011), malformation of chicks (Grasman et al., 1998) and long term breeding probability, hatching and fledging probabilities of chicks (Goutte et al., 2014). Effects on these reproductive endpoints will influence wild population success of birds. Since these chronic effects of PFOS are not studied extensively in birds and additionally, PFOS can be transferred to the eggs as well (Newsted et al., 2005), a carefully designed monitoring program should fill in the knowledge gaps on actual risk for flamingos and other birds now foraging again in Saliña Goto, including the potential effects on their offspring and thus the development of populations.

5.5 Conclusions and recommendations

A direct effect of surplus rainfall on flamingo abundance in Saliña Goto was not found in the present study. Indirect effects of persistent, toxic firefighting foam via disappearance of food species are a likely explanation for the observed changes in flamingo counts. Social behaviour patterns may have strengthened and prolonged the clear and long term abandoning of Saliña Goto. This set of phenomena may have in fact protected the birds from toxic risk posed by PFOS exposures during their initial four years of absence.

Now that the birds have returned, indicating a restored food supply, they will be exposed to PFOS levels in their feed. Based on a first tier risk assessment toxic effects of PFOS via the food chain cannot be excluded. A confirmation of true exposure requires additional PFOS analysis in food and in flamingo blood. This poses practical problems, given the protected status of the species. The present research reveals that management of short-term incidents can have chronic toxic consequences when persistent firefighting foam constituents are emitted into the environment. It also shows the importance of incident preparedness industrial activities near vulnerable areas.

Acknowledgements

The following people and organisations are acknowledged for their contribution to the work presented in this paper. Wageningen Marine Research: Joel Couperus, Babeth van der Weide, Maadjieda Tjon-Atsoi, Evert van Barneveld and Erik Meesters. RIVM: Charles Bodar, Arthur de Groot, Monique van der Aa. Bonaire Government (openbaar lichaam): Frank van Slobbe. STINAPA Park rangers. Ducares: Peter Boshuis. Geke Lianne den Otter is acknowledged for proofreading the manuscript.

The Dutch Ministry of Economic Affairs and Ministry of Infrastructure and the Environment are acknowledged for funding the research presented in this paper under grant and project codes BO-11-011.05-21, M/609022/10/BO and M/609224.

References

Anonymous, 2011. Perfluorooctane sulphonate (PFOS) PFOS EQS dossier 2011.

Baldassarre, G., Arengo, F., 2000. A review of the ecology and conservation of Caribbean flamingos in Yucatan, Mexico. Waterbirds 23 (Special Publication 1), 70–79.

Bednarska, A., Jevtić, D., Laskowski, R., 2013. More ecological era: incorporating natural environmental factors and animal behavior. Integr Environ Assess Manag 9 (3), e39–e46.

BirdLife International, 2016. Species factsheet: *Phoenicopterus ruber*. `http://www.birdlife.org`.

Bodar, C., Lijzen, J., Moermond, C., Peijnenburg, W., Smit, E., Verbruggen, E., Janssen, M., 2011. Advies risicogrenzen grond en grondwater voor PFOS. Tech. Rep. 601050002/2011, RIVM, Bilthoven, The Netherlands.

Bots, J., De Bruyn, L., Snijkers, T., Van den Branden, B., Van Gossum, H., 2010. Exposure to perfluorooctane sulfonic acid (PFOS) adversely affects the life cycle of the damselfly *Enallagma cyathigerum*. Environ Pollut 158, 901–905.

Bucher, E. H., Chani, J. M., Echevarria, A. L., 2000. Andean flamigos breeding at Laguna Brava, la Rioja, Argentina. Waterbirds 23 (Special publication 1), 119–120.

Buitrago, J., Rada, M., Barroeta, M., Fajardo, E., Rada, E., Narváez, J., Simal, F., Monente, J., Capelo, J., Narváez, J., 2010. Physico-chemical indicator's monitoring for water quality in the saliñas of northwest Bonaire, n.a. - a base line study. Tech. rep., Estación de Investigaciones Marinas de Margarita to STINAPA, Bonaire.

De Zwart, D., De Groot, A., Kotterman, M., Van der Aa, M., Aalbers, T., Slijkerman,

D., Bodar, C., 2012. Follow-up study on the chemical status of Lake Goto, Bonaire - measurements and risk assessment. Tech. Rep. 609224001/2012, RIVM, Bilthoven, The Netherlands.

Del Hoyo, J., Elliott, A., Sargatal, J. (Eds.), 1992. Handbook of Birds of the World. Vol. 1. Lynx Edicions, Barcelona.

EC, 2011. Proposal for a Directive of the European Parliament and of the Council amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. COM(2011) 876 final 2011/0429 (COD).

ECHA, 2003. Technical guidance document on risk assessment in support of Commission Directive 93/67/EEC on risk assessment for new notified substances, Commission regulation (EC) No 1488/94 on risk assessment for existing substances and of Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Tech. Rep. EUR 20418 EN/2, European Chemicals Agency, Ispra, Italy.

Elphick, C., Dunning Jr., J., Sibley, D. (Eds.), 2001. National Audubon Society: The Sibley Guide to Bird Life and Behavior. Alfred A. Knopf, New York.

Espinoza, F., Parra, L., Aranguren, J., Martino, A., Quijada, M., Pirela, D., Rivero, R., Gutierrez, T., Jimenez, N., Leal, S., Leon, E., 2000. Numbers and distribution of the Caribbean flamingo in Venezuela. Waterbirds 23 (Special publication 1), 80–86.

Frederick, P., Jayasena, N., 2011. Altered pairing behaviour and reproductive success in white ibises exposed to environmentally relevant concentrations of methylmercury. P Roy Soc Lond B Bio 278 (1713), 1851–1857.

Gilbertson, M., Reynolds, L. M., 2010. Hexachlorobenzene (HCB) in the eggs of common terns in Hamilton Harbour, Ontario. Bull Environ Contam Toxicol 7, 371–373.

Goutte, A., Barbraud, C., Meillère, A., Weimerskirch, H., Chastel, O., 2014. Demographic consequences of heavy metals and persistent organic pollutants in a vulnerable long-lived bird, the wandering albatross. P Roy Soc Lond B Bio 281 (1787).

Grasman, K. A., Scanlon, P. F., Fox, G. A., 1998. Reproductive and physiological effects of environmental contaminants in fish eating birds of the great lakes: a review of historical trends. Environ Monitor Assess 53, 117–145.

Heugens, E., Hendriks, A., Dekker, T., Van Straalen, N., Admiraal, W., 2001. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. Crit Rev Toxicol 31 (3), 247–284.

Jorcin, Caglierani Casanova, 2011. Zoobentonic community of the saliñas of the

Slagbaai Washington national park - isle of Bonaire - Netherlands Antilles. Tech. rep., HidroLógica Asesoría y Consultoría Ambiental Ltda., Botucatu, São Paulo, Brasil.

Joustra, T., Brouwer-Korf, A., Mertens, F., Visser, J., 2011. Tank fires BOPEC Bonaire, 8-11 September 2010. Tech. Rep. M2010IN0908-02, The Dutch Safety Board, The Hague, The Netherlands.

Kushlan, J. A., 1978. Feeding ecology of wading birds. In: Sprunt, A., Ogden, J., Winckler, S. (Eds.), Wading birds. Audubon Society, New York, pp. 249–297.

Kwadijk, C., Korytar, P., Koelmans, A., 2010. Distribution of perfluorinated compounds in aquatic systems in The Netherlands. Environ Sci Technol 44 (10), 3746–3751.

Lindemann, D., Carpenter, J., KuKanich, B., 2016. Pharmacokinetics of a single dose of oral and subcutaneous Meloxicam in Caribbean flamingos. J Avian Med Surg 30 (1), 14–22.

Moermond, C., Verbruggen, E., Smit, C., 2010. Environmental risk limits for PFOS - a proposal for water quality standards in accordance with the Water Framework Directive. Tech. Rep. 601714013/2010, RIVM, Bilthoven, The Netherlands.

Mommaerts, V., Hagenaars, A., Meyer, J., De Coen, W., Swevers, L., Mosallanejad, H., Smagghe, G., 2011. Impact of a perfluorinated organic compound PFOS on the terrestrial pollinator *Bombus terrestris* (Insecta, Hymenoptera). Ecotoxicology 20 (2), 447–456.

Mooij, M., Van de Meent, D., Bodar, C., Boshuis, E., De Groot, A., De Zwart, D., Hoffer, S., Janssen, P., De Groot, G., Peijnenburg, W., Verbruggen, E., 2011. Compound depositions from the BOPEC fires on Bonaire - measurements and risk assessment. Tech. Rep. 609022067/2011, RIVM, Bilthoven, The Netherlands.

Newsted, J., Jones, P., Coady, K., Giesy, J., 2005. Avian toxicity reference values for perfluorooctane sulfonate. Environ Sci Technol 39 (23), 9357–9362.

OSPAR, 2013. OSPAR list of chemicals for priority action (revised 2013). Tech. Rep. 2004-12, OSPAR.

Pirela, D., 2000. Monitoring Caribbean flamingos at Los Olivitos Wildlife Refuge, Western Venezuela. Waterbirds 23 (Special publication 1), 103–108.

Posthuma, L., Wahlstrom, E., Nijenhuis, R., Dijkens, C., De Zwart, D., Van de Meent, D., Hollander, A., Brand, E., Den Hollander, H., Van Middelaar, J., Van Dijk, S., Hall, E., Hoffer, S., 2014. The flash environmental assessment tool: Worldwide first aid for chemical accidents response, pro action, prevention and preparedness. Environ Int 72, 140–156.

R Core Team, 2016. R: A Language and Environment for Statistical Computing.

R Foundation for Statistical Computing, Vienna, Austria.

Ramsar, 2016. The list of wetlands international importance - September 2016. **http://www.ramsar.org/sites/default/files/documents/library/sitelist.pdf**.

Rooth, J., 1965. The flamingos on Bonaire: habitat, diet and reproduction of *Pheonicopterus ruber ruber*. Tech. Rep. Publication 41, The Natural Science Study Group of Suriname and the Dutch Antilles.

Simal, F., 2010. Monitoring program for waterbirds inhabiting the saliñas located on the northwest of Bonaire, Netherlands Antilles. Tech. Rep. Year report 2009, STINAPA.

Simal, F., Holian, P., Albers, E., Wolfs, E., 2011. Monitoring program for waterbirds inhabiting the saliñas located on the northwest of Bonaire, Netherlands Antilles. Tech. Rep. Year report 2010, STINAPA.

Sprunt, A., 1975. The Caribbean. In: Kear, J., Duplaix-Hall, N. (Eds.), Flamingos. T. & A. D. Poyser, Berkhamsted, UK, pp. 65–74.

Stidd, C., 1970. Nth root normal distribution of precipitation. Water Resour Manag 6, 1095–1103.

Tripp, K., Collazo, J., 2003. Density and distribution of water boatmen and brine shrimp at a major shorebird wintering area in Puerto Rico. Wetl Ecol Manag 11, 331–341.

Van Gossum, H., Bots, J., Snijkers, T., Meyer, J., Van Wassenbergh, S., De Coen, W., De Bruyn, L., 2009. Behaviour of damselfly larvae (*Enallagma cyathigerum*) (Insecta, Odonata) after long-term exposure to PFOS. Environ Pollut 157, 1332–1336.

Vargas, F., Barlow, S., Hart, T., Jimenez-Uzcátegui, G., Chavez, J., Naranjo, S., Macdonald, D., 2008. Effects of climate variation on the abundance and distribution of flamingos in the Galápagos Islands. J Zool 276 (3), 252–265.

Wells, J., Debrot, A., 2008. Bonaire. In: Wege, D., Anadón-Irizarry, V., Vincenty, M. (Eds.), Important Bird Areas in the Caribbean: key sites for conservation. BirdLife International, Cambridge, UK, pp. 95–102.

Wood, S. (Ed.), 2006. Generalized Additive models: An Introduction with R. CRC Press, Boca Raton, FL, USA.

Wood, S., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J Roy Stat Soc B 73 (1), 3–36.

5.A Appendix

5.A.1 Detailed methodology for chemical analyses

The batch of sediment and water samples of 2010 was analysed by Ducares (TNO). Analyses were performed in three parts: an extraction part which was different for each matrix; a clean-up step; and determination by LC-MS/MS which was the same for all samples. Sediments were extracted by shaking 5 gram sediment for 30 minutes using 20 mL of a mixture of THF/water (50:50) after the addition of 50 μL acetic acid. The extract was centrifuged and 10 mL of supernatant was transferred to a new tube and concentrated until dry and subsequently resubstituted in 10 mL water which was then ready for clean-up. Water samples were prepared for clean up by adding 50 μL acetic acid to 100 mL sample.

Samples were cleaned up using OASIS WAX-SPE and Supelclean ENVI-Carb SPE columns. Extracts were brought on the WAX column after which the column was washed with acetatebuffer (pH 4) and THF/Acetonitril (20:80). After washing the column was placed on top of an ENVI-Carb SPE column and eluted using 0.5% ammonia in methanol. The cleaned extract was dried under a stream of nitrogen and reconstituted in an acetatebuffer/methanol mixture and transferred to a sample vial for analysis by LC-MS.

Analyses were performed using a Waters Ultima Pt LC-MS in ESI negative mode. Source temperature was 100 °C and desolvation temperature 250 °C. Separation was performed using a Waters Aquity column using a gradient with 10 mM Ammoniumacetate pH 4 as solvent A and a mixture of Acetonitril/Methanol (20:80) as solvent B with a flow of 0.3 mL/min. Quantification was performed by isotope dilution.

PFASs in water, sediment and biota samples of 2012, 2013 and 2015 were extracted and analysed according to the procedure described by Kwadijk et al. (2010) (see reference list of main text). In short, for water, 1 L water samples were extracted using 1 g OASIS HLB SPE cartridges (Waters), dried using sodium sulphate and cleaned up using 50 mg ENVIcarb (Sigma Aldrich, Zwijndrecht, Netherlands). Samples were concentrated to 0.7 mL and then transferred to GC-vials and stored at 4 °C until analysis.

For sediment and biota samples 2-5 grams of sample was extracted by shaking with acetonitrile (LGC, Wessel, Germany) and subsequently dried using sodium sulphate followed by a clean-up using ENVIcarb (Sigma-Aldrich, the Netherlands). Samples were concentrated to 0.7 mL and then transferred to GC-vials and stored at 4 °C until analysis. For all PFAS analyses 50 ng of 13C4-PFOS in 350 μL of methanol and 50 ng of 13C4-PFOA in 350 μL of methanol were used as internal standards.

PFAS analysis was carried out using a Thermo Finnigan (Waltham, United States) Surveyor Autosampler and HPLC coupled with a Thermo Finnigan LCQ

advantage Ion-Trap MS with electrospray (ESI–MS/MS) for quantification and detection. Separation was performed on a 100 x 2.10mm (5 μm) Fluophase RP column (Thermo) using ammonium formate and formic acid in acetonitrile as mobile phase A and ammonium formate and formic acid in demi water as mobile phase B. Carboxylates and sulfonates were analysed in separate runs due to a high difference in sensitivity for the capillary temperature. For the carboxylates a capillary temperature of 200 °C was set, for the sulfonates this was set to 300 °C.

Internal reference samples were determined with every set of samples with satisfactory results (within 2s), no PFAS were detected in the blanks and all calibration curves had an $R^2 \geq 0.995$. LOQ in the final extract was 0.3 ng/mL.

The effect of high salt contents on the analysis was tested by using internal reference material (IRM) for water and sediment with and without added salt. No effect on determined levels was observed, not even with salt concentrations twice the levels of the Bonaire saliña samples.

Dry weights were determined gravimetrically by weight loss (104 °C, 3 hours). Dry weight in sediment samples was corrected for dry weight of water samples from the year 2012; due to the high salt content the mass of salt in the adhering water was significant.

Dry weight was also determined by washing sediment twice with a relative large volume of demineralised water (followed by centrifugation). The dry weight was comparable with the values obtained by correcting for calculated adhering salt.

Organic carbon was determined gravimetrically as loss on ignition (550°C, 22 h) in sediment samples previously washed twice with demineralised water to remove the salt.

5.A.2 Precipitation model

Figs. 5.9 and 5.10 show the GAM model fit to the precipitation model, from which the residuals are used in the model of the bird counts, as described in the main text.

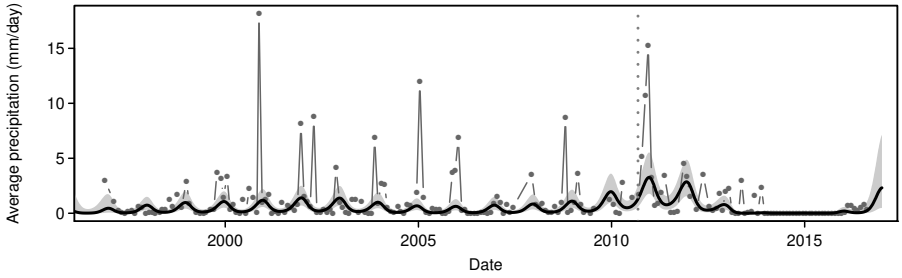


Fig. 5.9: Average precipitation in the two weeks prior to each bird counting events (grey dots) and the smoothed Generalised Additive Model trend (black line with 95% confidence bands shown as grey bands). Vertical dashed line indicates the moment of the fires at the oil terminal.

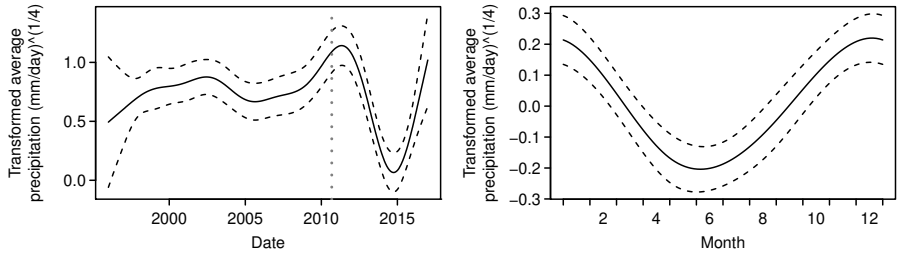


Fig. 5.10: Generalised Additive Modelling results of the transformed precipitation data (Fig. 5.9) split per explanatory variable: overall time on the left and the month on the right. Vertical dashed line indicates the moment of the fires at the oil terminal.

5.A.3 Additional tables

Tables with additional information used in the main text. Table 5.5 lists all relevant substances and their acronyms. Table 5.6 lists the organic content measured in sediment samples. Tables 5.7 and 5.8 list concentrations as measured in each of the water and sediment samples.

Table 5.5: List of full PFAS chemical name and their corresponding abbreviation as used in this Chapter.

Abbreviation	Full chemical name	Abbreviation	Full chemical name
PFBA	Perfluoro-n-butanoic acid	PFTrA	Perfluoro-n-tridecanoic acid
PFPA	Perfluoro-n-pentanoic acid	PFTeDA	Perfluoro-n-tetradecanoic acid
PFHxA	Perfluoro-n-hexanoic acid	PFHxDA	Perfluoro-n-hexadecanoic acid
PFHpA	Perfluoro-n-heptanoic acid	PFODA	Perfluoro-n-octadecanoic acid
PFOA	Perfluoro-n-octanoic acid	PFBS	Perfluoro-1-butanesulfonate
PFNA	Perfluoro-n-nonanoic acid	PFHxS	Perfluoro-1-hexanesulfonate
PFDcA	Perfluoro-n-decanoic acid	PFHpS	Perfluoro-1-heptanesulfonate
PFUnA	Perfluoro-n-undecanoic acid	PFOS	Perfluoro-1-octanesulfonate
PFDoA	Perfluoro-n-dodecanoic acid	PFDS	Perfluoro-1-decanesulfonate

Table 5.6: Organic content measured in sediment samples of 2012.

Location	Sample	Year	Organic content (%)
1 Matijs	sediment	2012	7.2
3 Bartol	sediment	2012	10
7 Goto	sediment	2012	8.2
8 Goto	sediment	2012	7.4
9 Goto	sediment	2012	29.8
10 Goto	sediment	2012	5.6
11 Goto	sediment	2012	12

Table 5.7: Results of chemical analyses of sediment and water samples, part A (see Table 5.8 for part B). Concentrations below the limit of detection are reported as less than that limit. See Table 5.5 for full chemical name.

Location	Year	Sample	Unit	PFBA	PFPA	PFHxA	PFHpA	PFOA	PFNA	PFDcA	PFUnA	PFDoA
7 Goto	2010	sed.	μg/kg d.w.		< 0.1	0.1	< 0.05	0.33	< 0.1	< 0.05	0.09	0.06
8 Goto	2010	sed.	μg/kg d.w.		< 0.1	0.17	0.05	0.29	< 0.1	< 0.05	0.08	< 0.05
9 Goto	2010	sed.	μg/kg d.w.		0.12	0.18	0.11	0.57	0.14	0.07	< 0.05	< 0.05
1 Matijs	2010	water	ng/L		< 0.1	0.14	0.12		< 0.1	< 0.1		< 0.1
6 Slagbaai	2010	water	ng/L		7	0.7	0.12		< 0.1	< 0.1		< 0.1
7 Goto	2010	water	ng/L		130	130	20		0.61	< 0.1		< 0.1
8 Goto	2010	water	ng/L		130	130	19.2		0.56	< 0.1		< 0.1
9 Goto	2010	water	ng/L		170	140	17		0.98	< 0.1		< 0.1
1 Matijs	2012	sed.	μg/kg d.w.[20]		< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4
10 Goto	2012	sed.	μg/kg d.w.[20]		< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4
11 Goto	2012	sed.	μg/kg d.w.[20]		< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4
2 Matijs	2012	sed.	μg/kg d.w.[20]		< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4
3 Bartol	2012	sed.	μg/kg d.w.[20]		< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6
4 Bartol	2012	sed.	μg/kg d.w.[20]		< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6	< 0.6
5 Slagbaai	2012	sed.	μg/kg d.w.[20]		< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4	< 0.4
6 Slagbaai	2012	sed.	μg/kg d.w.[20]		< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5
7 Goto	2012	sed.	μg/kg d.w.[20]		< 0.2	< 0.2	< 0.2	< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
8 Goto	2012	sed.	μg/kg d.w.[20]		< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5
9 Goto	2012	sed.	μg/kg d.w.[20]		< 1.9	< 1.9	< 1.9	< 1.9	< 1.9	< 1.9	< 1.9	< 1.9
1 Matijs	2012	water	ng/L		< 3.1	< 3.1	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3

Continued on next page

[20]Excluding salt content.

Table 5.7 continued

Location	Year	Sample	Unit	PFBA	PFPA	PFHxA	PFHpA	PFOA	PFNA	PFDcA	PFUnA	PFDoA
10 Goto	2012	water	ng/L		82	180	22	3.5	< 0.3	< 0.3	< 0.3	< 0.3
11 Goto	2012	water	ng/L		82	160	11	6.4	< 0.3	< 0.3	< 0.3	< 0.3
2 Matijs	2012	water	ng/L		< 3.2	< 3.2	< 0.3	3.5	< 0.3	< 0.3	< 0.3	< 0.3
3 Bartol	2012	water	ng/L		< 3.0	< 3.0	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
4 Bartol	2012	water	ng/L		< 3.0	< 3.0	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
5 Slagbaai	2012	water	ng/L		< 3.0	< 3.0	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
6 Slagbaai	2012	water	ng/L		< 2.9	< 2.9	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
7 Goto	2012	water	ng/L		89	100	15	8.7	< 0.3	< 0.3	< 0.3	< 0.3
8 Goto	2012	water	ng/L		120	140	20	14	< 0.3	< 0.3	< 0.3	< 0.3
9 Goto	2012	water	ng/L		99	130	32	11	< 0.3	< 0.3	< 0.3	< 0.3
1 Matijs	2015	sed.	µg/kg d.w.	< 1.6	< 0.3	< 0.6	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
3 Bartol	2015	sed.	µg/kg d.w.	< 1.3	< 0.3	0.5	0.3	0.3	< 0.3	< 0.3	< 0.3	< 0.3
7 Goto	2015	sed.	µg/kg d.w.	< 0.9	< 0.1	< 0.2	< 0.1	1.5	< 0.1	< 0.1	< 0.1	< 0.1
8 Goto	2015	sed.	µg/kg d.w.	< 0.6	< 0.1	< 0.2	< 0.1	1.3	< 0.1	< 0.1	< 0.1	< 0.1
9 Goto	2015	sed.	µg/kg d.w.	< 2.5	< 0.5	< 1.0	< 0.5	2.5	< 0.5	< 0.5	< 0.5	< 0.5
10 Goto	2015	sed.	µg/kg d.w.	< 1.3	< 0.2	< 0.7	< 0.2	< 1.7	< 0.2	< 0.2	< 0.2	< 0.2

End of Table 5.7

Table 5.8: Results of chemical analyses of sediment and water samples, part B (see Table 5.7 for part A). Concentrations below the limit of detection are reported as less than that limit. See Table 5.5 for full chemical name.

Location	Year	Sample	Unit	PFTrA	PFTeDA	PFHxDA	PFODA	PFBS	PFHxS	PFHpS	PFOS	PFDS
7 Goto	2010	sed.	µg/kg d.w.	0.1	< 0.2	< 0.2	< 5	0.12	1.8		15.1	< 0.1
8 Goto	2010	sed.	µg/kg d.w.	0.05	< 0.2	< 0.2	< 5	0.16	1.7		18.3	< 0.1
9 Goto	2010	sed.	µg/kg d.w.	< 0.05	< 0.2	< 0.2	< 5	0.31	4.6		58.5	< 0.1
1 Matijs	2010	water	ng/L	0.31	0.65	0.34	< 5	< 0.25	0.43		1.9	< 0.25
6 Slagbaai	2010	water	ng/L	< 0.1	0.46	0.38	< 5	0.56	4.6		8.4	< 0.25
7 Goto	2010	water	ng/L	< 0.1	1.1	0.43	< 5	193	383		131	< 0.25
8 Goto	2010	water	ng/L	< 0.1	1.3	0.13	< 5	196	323		156	< 0.25
9 Goto	2010	water	ng/L	< 0.1	2.6	0.72	< 5	164	184		130	< 0.25
1 Matijs	2012	sed.	µg/kg d.w.[21]	< 0.4				< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
10 Goto	2012	sed.	µg/kg d.w.[21]	< 0.4				< 0.2	0.4	< 0.2	14	< 0.2
11 Goto	2012	sed.	µg/kg d.w.[21]	< 0.4				< 0.2	7.3	< 0.2	11.5	< 0.2
2 Matijs	2012	sed.	µg/kg d.w.[21]	< 0.4				< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
3 Bartol	2012	sed.	µg/kg d.w.[21]	< 0.6				< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
4 Bartol	2012	sed.	µg/kg d.w.[21]	< 0.6				< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
5 Slagbaai	2012	sed.	µg/kg d.w.[21]	< 0.4				< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
6 Slagbaai	2012	sed.	µg/kg d.w.[21]	< 0.5				< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
7 Goto	2012	sed.	µg/kg d.w.[21]	< 0.2				< 0.2	3	< 0.2	18.5	< 0.2
8 Goto	2012	sed.	µg/kg d.w.[21]	< 0.5				< 0.3	0.3	< 0.3	20.6	< 0.3
9 Goto	2012	sed.	µg/kg d.w.[21]	< 1.9				< 1.3	7.7	< 1.3	499.6	< 1.3

Continued on next page

[21] Excluding salt content.

Table 5.8 continued

Location	Year	Sample	Unit	PFTrA	PFTeDA	PFHxDA	PFODA	PFBS	PFHxS	PFHpS	PFOS	PFDS
1 Matijs	2012	water	ng/L	< 3.1				< 0.3	< 0.3	< 0.3	< 0.3	< 0.3
10 Goto	2012	water	ng/L	< 2.9				150	210	5.7	120	< 0.3
11 Goto	2012	water	ng/L	< 3.0				160	200	5.3	120	< 0.3
2 Matijs	2012	water	ng/L	< 3.2				< 0.3	< 0.3	< 0.3	< 0.3	< 0.4
3 Bartol	2012	water	ng/L	< 3.0				< 0.3	< 0.3	< 0.3	0.7	< 0.3
4 Bartol	2012	water	ng/L	< 3.0				< 0.3	< 0.3	< 0.3	0.7	< 0.3
5 Slagbaai	2012	water	ng/L	< 3.0				< 0.3	< 0.3	< 0.3	3.2	< 0.3
6 Slagbaai	2012	water	ng/L	< 2.9				< 0.3	0.2	< 0.3	1.5	< 0.3
7 Goto	2012	water	ng/L	< 2.9				160	170	3.4	85	< 0.3
8 Goto	2012	water	ng/L	< 2.9				190	210	6.1	100	< 0.3
9 Goto	2012	water	ng/L	< 2.9				170	210	8.3	130	< 0.3
1 Matijs	2015	sed.	µg/kg d.w.	< 0.3				< 0.3	< 0.3	< 0.3	< 1.6	< 0.3
3 Bartol	2015	sed.	µg/kg d.w.	< 0.3				< 0.2	< 0.3	< 0.3	1.3	< 0.3
7 Goto	2015	sed.	µg/kg d.w.	< 0.1				< 1.9	< 0.1	< 0.1	13	< 0.1
8 Goto	2015	sed.	µg/kg d.w.	< 0.1				0.1	< 0.1	< 0.1	16.8	< 0.1
9 Goto	2015	sed.	µg/kg d.w.	< 0.5				3.6	< 0.5	< 0.5	50.8	< 0.5
10 Goto	2015	sed.	µg/kg d.w.	< 0.2				< 3.0	< 0.2	< 0.2	< 31.7	< 0.3

End of Table 5.8

Chapter 6

General discussion

6.1 Introduction

Ecotoxicological risk assessment evaluates the likelihood and magnitude of adverse effects experienced by an ecosystem during or after an exposure to one or more toxicants. As such, ecotoxicological risk assessment is used to evaluate and manage the quality of environmental compartments. Bioassay data are used in ecotoxicological risk assessment to quantify hazard. With decades worth of available bioassay data, hazard quantification often relies on existing data. Guidelines, such as those of the European Chemicals Agency (ECHA, 2011b), prescribe selection criteria for the inclusion of such bioassay data in hazard quantification. Most guidelines consider bioassay data to be acceptable for use in risk assessment, when it is of sufficient quality. In this thesis it is evaluated how such guidelines, for selection of existing *in vivo* bioassay data in risk assessment, can be improved. Table 6.1 lists and explains the most important concepts that are used throughout this chapter.

First of all, reliability is one of the quality aspects addressed by guidelines. Often bioassay data is considered reliable when the experiment that generated the data is reproducible. As experiments are rarely repeated, reproducibility is assumed to be high when an experiment is performed using a standardised protocol and it is well documented. As this is no guarantee for reliable and error free bioassay data, **Chapter 2** examined whether trustworthiness of data can be evaluated by applying Benford's Law.

Table 6.1: Terminology of important concepts and their definitions and interpretation as used in this thesis.

Concept	Definition
Risk assessment	"A process of evaluation, including the identification of the attendant uncertainties, of the likelihood and severity of an adverse effect(s) / event(s) occurring to man or the environment following exposure under defined conditions to a risk source(s)." (EC, 2000b)
Hazard quantification	Process in which the hazard (the potency to cause harm) level is quantified. The result can be hazard indicators such as environmental quality standards or the 5% hazardous concentration (HC5).
Bioassay	An (laboratory) experiment in which a test species is exposed to different levels of a stressor (usually a toxicant). This thesis focuses on *in vivo* bioassays.
Bioassay data	Data produced by bioassays. This thesis focuses on (no) effect concentrations (such as 50% effect concentrations (EC50) and No Observed Effect Concentrations (NOEC)) as they dominate databases with existing bioassay data.
Objective / Context	The objective for which or the context within which bioassay data is used in risk assessment and hazard quantification. (for example: deriving water quality standards for the EU Water Framework Directive; or evaluating the impact of fire fighting foams to flamingo birds in a tropical salt lake).
Quality	Commonly defined as the fitness for purpose. In risk assessment reliability and relevance are important aspects of bioassay data quality. It is argued here that certainty is also an important aspect.
Reliability	Aspect of quality that should reflect the acceptability of bioassay data. Usually evaluated with qualitative scoring systems focusing on the level of standardisation and documentation (which should indicate the reproducibility). Reliable data should also be free of (intentional) errors.

Continued on next page

Table 6.1 continued

Concept	Definition
Relevance	Relevance of bioassay data indicates that the experiment matches well with the situation for which the risk needs to be assessed (which depends on the context). This includes relevant species, concentrations, exposure routes, and other conditions (such as temperature). Highly relevant data allow for accurate hazard quantification.
Certainty	Highly certain bioassay data and hazard quantification indicate high precision and high accuracy of the estimated value.
Precision	Precise bioassay data and hazard quantification indicates little variance around the estimated value.
Accuracy	Accurate bioassay data and hazard quantification indicates that the estimated value is close to the 'true' value. This is the case when relevant data is used.
Variance	Deviation of estimated values from its expected mean value.

End of Table 6.1

Once the reliability of the data has been determined, choices have to be made about the level of standardisation that is required to reach the optimal selection. This was examined with the case studies described in **Chapters 3** and **4**. In **Chapter 3** bioassays are evaluated in which species are exposed to elevated carbon dioxide levels. It showed that selecting bioassays with a high level of standardisation does improve the reliability of hazard quantification, but does not automatically improve the precision of the quantification. **Chapter 4** studied bioassays where species are exposed to elevated temperatures. These bioassays were in general more precise and reliable as they enjoy a higher level of standardisation.

High precision indicates that the hazard level has little variation around its estimate. Accurate hazard quantification indicates that hazard levels closely reflect the conditions in the field for which the risk needs to be addressed. Accuracy therefore largely depends on the context in which bioassay data is used. Therefore, in **Chapter 5** risk is assessed for a real field situation and hazard quantification is tied to the context in which it is used. **Chapter 5** evaluates risks of toxic components in firefighting foams to flamingo birds. After a large petrochemical fire, these toxicants entered the salt lake where the birds forage. Most available bioassay data is not highly representative for the tropical salt lake, hampering an accurate risk assessment. But lack of relevant bioassay data is not necessarily the main issue in this study as (internal) exposure levels are also uncertain.

The guidelines by ECHA (2011b) focus on quality of the data in the selection procedure, where quality is defined as the reliability and relevance of data. However, it can be argued that the accuracy and precision (i.e., the certainty) of data should be considered as an aspect of quality; as is the case for geographic data (ISO, 2013). Whereas ECHA guides the analysis of certainty in the risk assessment process (ECHA, 2012), this thesis proposes to consider certainty of bioassay data at an earlier stage of the risk assessment cycle, namely in the data selection process, as will be explained in this chapter.

This chapter will address the aspects of quality (reliability, certainty and relevance), by reviewing literature and using the case studies presented in the previous chapters. Firstly, inherent quality will be addressed, which includes quality aspects (reliability and certainty) of the data that are not affected by external factors. Secondly, the relevance of the data; where the context will be examined in which the data will be used and to which conditions, used for generating the data, it should relate. The relevance of each of the quality aspects to the data selection process will be discussed and concluded with a future perspective.

6.2 Inherent quality of data

6.2.1 Reliability of data

Traditionally reliability of toxicological bioassay data are evaluated using a scoring system described by Klimisch et al. (1997). Depending on whether standardised/accepted protocols are used and how well the test is documented, data is classified into one of four categories (reliable without restrictions; reliable with restrictions; unreliable; or not assignable) (Klimisch et al., 1997). Reliability of bioassays is vital in hazard and risk assessment, where an increasing amount of bioassays of poor (technical) quality are produced and claims based on such assays appear to be false (Harris and Sumpter, 2015). Although the generic reliability scoring approach of Klimisch et al. (1997) was a major step forwards, the provided guidance allows for multiple interpretations which results in inconsistent reliability scoring among scientists (Kase et al., 2016; Segal et al., 2015). Numerous recent studies expand and improve upon Klimisch et al. (1997) by, for instance, refining its scoring criteria and their definition; or addressing additional aspects such as experimental setup and statistical design; or focusing on specific studies (such as *in vitro* studies) and substances (e.g., pesticides or nanoparticles); or applying the scoring criteria in a stepwise approach (Beasley et al., 2015; Isigonis et al., 2015; Lynch et al., 2016; Moermond et al., 2016; Segal et al., 2015).

Another issue with reliability scoring systems is that they do not score the reliability of data directly but rather aspects that could affect the acceptability and certainty of data. When a reliability score indicates high reliability it is probably safe to use the data when the relevance of the data is also considered. However, when a reliability score is low, it does not mean that the data is necessarily unreliable (e.g., data that are generated not using quality management systems such as Good Laboratory Practice (GLP) are not necessarily unreliable). Moreover, the reliability scores also don't guarantee that the data are free of (un)intentional errors. For screening for anomalies in the data itself, rather than scoring the conditions under which they were generated and reported, Benford's Law can be a useful instrument (**Chapter 2**). This is important as the anomalies could result from errors or fraud (data manipulation), in particular for cases where stakes are high and interests conflict.

A major shortcoming of the current scoring systems is that the scores don't quantify the reliability of bioassay data. This would be desirable, as it would allow risk assessors to determine how much overall reliability of hazard quantification improves and compare this with the reduction in certainty due to the smaller sample size after selecting reliable bioassay data only. This is illustrated in **Chapter 3**, where the uncertainty increased in the hazard indicators (such as the 5% and 50% hazardous concentration: HC5 and HC50), due to a smaller data set used, after selecting only the reliable carbon dioxide (CO_2)

effect concentrations. It remains to be studied whether the improvement of reliability (by selecting only the utmost reliable data) outweighs the reduction of certainty in this particular case, but also in general. In reliability assessments, this also requires a sharper distinction between acceptability and certainty, as also pointed out by Moermond et al. (2016).

6.2.2 Standardisation of protocols

Standardisation of protocols improves reproducibility and thus certainty of test results. Such standardisation is promoted by quality assurance frameworks such as Good Laboratory Practise (GLP, OECD (1998)) which is required for acceptance of tests within specific legislative frameworks (Férard and Férard, 2013). The use of GLP and standardisation of test protocols however can also create biases in available toxicity data and hazard assessments (Forbes et al., 2016). Hendriks et al. (2013) showed that the variability in effect concentrations among species increases when more species have been tested (Fig 6.1, right panel). When larger number of non-standardised tests are included the uncertainty may also increase. However, it is hypothesised that this is because when fewer species have been tested, they are more likely all standard test species (tested under standard protocols), with less variance in sensitivity. It is no surprise that, due to the bias discussed above, the usefulness of a limited focus on standardisation and GLP is under debate (Borgert et al., 2016; McCarty et al., 2012; Moermond et al., 2017). An important shortcoming of standardisation is that it tends to focus on a limited number of specific environmental compartments and conditions (mainly freshwater, in temperate climatic regions). This causes problems when a hazard level is required for situations that strongly deviate from these standard conditions, as was shown in **Chapter 5** for a Caribbean saline lake and its sediments. How influential such problems are and to which extent the data can be extrapolated will be discussed in more detail in the following sections, also in relation to certainty of data.

6.2.3 Certainty of data

In risk assessment, there are several sources of uncertainty. ECHA points out that uncertainty can originate from the scenario specification, model definition and parameter estimation (ECHA, 2012). Uncertainty in scenario specification can originate from an incomplete or incorrect description of the system, for which the risk needs to be assessed (e.g., overlooking important exposure pathways). Uncertainty in the model definition can occur, for instance, when a model is used outside its specified domain, or when not or incorrectly including model structures (such as correlation between parameters). Parameter estimates (such as hazard indicators like the HC5), that are derived for a specific scenario with a certain model, can be uncertain amongst others due to measurement and

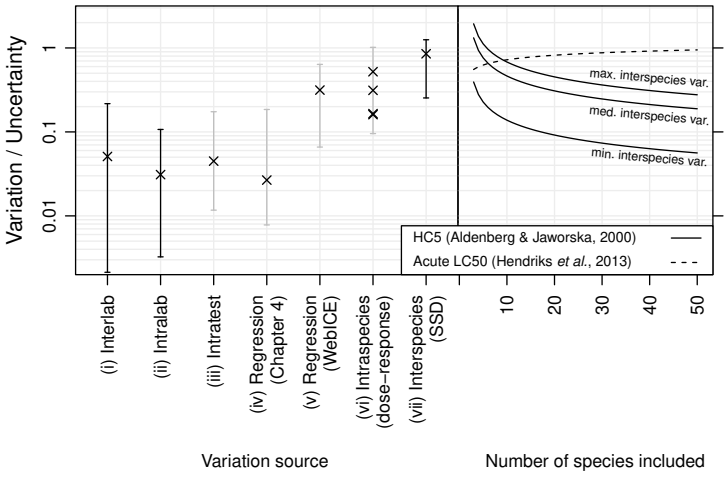


Fig. 6.1: This figure compares variance/uncertainty resulting from several specific causes which allows to place variance/uncertainty resulting from other causes into context. Left panel: variation in $\log_{10}$-transformed (no) effect concentrations with different causes expressed as standard deviations unless mentioned otherwise. More details for this figure are presented in Appendix 6.A. Black error bars show minimum and maximum, grey ones the 5% and 95% percentiles. (i, *inter*lab) Variation of EC50 and LC50 between different laboratory (Gaudet-Hull et al., 1994; Grothe and Kimerle, 1985). (ii, *intra*lab) Variation of repeated tests at the same laboratory (Gaudet-Hull et al., 1994; Grothe and Kimerle, 1985). (iii, *intra*test) Median proximate standard deviation of $\log_{10}$ EC50 and $\log_{10}$ LC50 values within single tests. (iv) Scaled variation in TTIs as determined with linear regression in **Chapter 4**. (v) Median variation of effect concentrations extrapolated using *inter*species correlation estimates (Dyer et al., 2008). (vi) Median slopes of dose-response curves for different species groups algae, crustaceans, molluscs and fish (from top to bottom, where the latter two overlap) (Smit et al., 2001). (vii) Median slope of Species Sensitivity Distributions (Harbers et al., 2006). Right panel: proxy of a standard deviation in $\log_{10}$ HC5 values in relation to the number of tested species (Aldenberg and Jaworska, 2000). Uncertainty in HC5 depends on the slope of the SSD and is calculated for the minimum, median and maximum slope (Harbers et al., 2006) shown as the bottom, middle and top line respectively. Dashed line shows the relation between number of tested species and the standard deviation in acute $\log_{10}$ LC50 (Hendriks et al., 2013).

sampling errors, bioassay data selection and extrapolation. The first two sources of uncertainty (scenario specification and model definition) depend on the context in which the risk assessment takes place and will be discussed to some extent in the section 6.4 on relevance and context. The uncertainty in bioassay data (on which most hazard indicators are based) is discussed here and compared with (natural) variability of bioassay data (i.e., (no) effect concentrations) also addressing its role in the data selection procedure.

Fig 6.1 shows the relative contribution to bioassay variance in (no) effect levels caused by different sources. It will be used throughout this chapter to compare with additional information about (un)certainty in bioassay data. This allows them to be placed in context and understand the magnitude of the uncertainty. Appendix 6.A provides information on the sources and approaches used to construct Fig 6.1.

Part of the natural variation, i.e., variation of sensitivity between individuals of a species (*intra*species) and that between species (*inter*species), is shown in the left panel of Fig 6.1 (labelled 'vi' based on Smit et al. (2001) and 'vii' based on Harbers et al. (2006), respectively). The *inter*species variation is often referred to as the Species Sensitivity Distribution (SSD) (Posthuma et al., 2002).

Inherent accuracy versus precision

In accurate data systematic errors (i.e., errors that are directed in a particular direction) are minimised. Inaccuracy is mostly caused by a mismatch between experimental setup and conditions and the target environment. This type of inaccuracy is not inherent and is therefore discussed in section 6.4 (Relevance and context). Inaccuracy that is not affected by such external factors generally (i.e., as part of the inherent quality) occurs when laboratory equipment is not handled or calibrated correctly (Flanagan et al., 2007) or required test conditions are not correctly maintained. Particularly for standardised tests this error is expected to be small (precision, i.e. caused by errors that are not directed in a particular direction). In addition, these errors are generally indistinguishable from random errors. For instance, the *inter*laboratory variance shown in Fig 6.1 (labelled 'i' in left panel) includes both random and systematic errors.

Uncertainty of bioassay data

An indication of the uncertainty and variability of bioassay data and its sources is shown in the left panel of Fig 6.1. Uncertainty is shown as the standard deviation of $\log_{10}$-transformed (no) effect concentration (or a proxy of this). As the data is $\log_{10}$-transformed, it will give an indication of how many orders of magnitude within the effect concentrations vary. Three standard deviations roughly span 90% of the statistical population. If, e.g., the standard deviation of

the $\log_{10}$-transformd data equals one, this means that the lower and upper 90% confidence limit is roughly a factor of 10^3 apart.

In Fig 6.1 (left panel) several sources of uncertainty are distinguished, amongst which: i) *inter*laboratory (same test performed by different laboratories); ii) *intra*laboratory (same test repeated by the same laboratory); iii) and *intra*test variability (uncertainty within a single test). *Inter*- and *intra*laboratory variation is shown for two ring tests only (as many ring tests don't provide the raw data required to construct Fig 6.1) (Gaudet-Hull et al., 1994; Grothe and Kimerle, 1985). The variability between and within laboratories is very similar to variation within single tests. This can be expected for bioassays that have been standardised or are in the process of being standardised. *Inter*- and *intra*laboratory variability can be expected to be larger for non-standardised tests. Thus uncertainty introduced when including less standardised (reliable) tests will increase the uncertainty in risk assessment, but with how much is unknown. Partly because there are no ring tests performed with tests that are not yet (close to being) standardised. But even when such ring tests are available this will not be very informative for other non-standardised tests as the uncertainty will also depend on the (statistical power resulting from the) experimental setup. In **Chapter 3** it is shown that a limited experimental setup results in uncertainties of CO_2 effect concentrations. These are, in this case, in the same range as *inter*species variation.

Uncertainty can also result from (linear) regression or from other interpolation techniques as was the case for temperature induced mortality (**Chapter 4**). **Chapter 4** reports the temperature induced mortality in relation to the acclimatisation temperature, which is obtained through linear regression based on experimental data for different acclimatisation temperatures. The uncertainty resulting from the regression (0.008 up to 0.2) is in this case (iv) in the same range as *intra*test (iii) variation (Fig 6.1). The *Inter*species Correlation Estimate (ICE) model also uses linear regression but in this case to extrapolate an effect concentration from a surrogate to a target species (Dyer et al., 2008). In the latter case the correlation is much weaker and estimates therefore more uncertain (vi) (ranging from 0.07 up to 0.7) in the range of *intra*species variation (vi) (Fig 6.1). Uncertainty from regression techniques used for *inter*- and extrapolation should be considered when quantifying hazard, in particular when this uncertainty is large.

Consequences of uncertainty for bioassay selection

The left panel of Fig 6.1 shows that uncertainty in bioassay results (iii) can be in the same range as *intra*- (vi) and *inter*species variation (vii) in sensitivity. Therefore, this uncertainty should be taken into account in hazard and risk assessment, starting with the bioassay selection process. In this process it has to be decided whether or not to omit uncertain bioassay data from the hazard

quantification, in order to optimise the certainty of the risk assessment in which it is used. As stated before, reducing the amount of data used in hazard quantification can also increase the uncertainty in the 5% hazardous concentration (HC5, a commonly used probabilistic hazard indicator (Aldenberg and Slob, 1993; Straalen and Denneman, 1989)).

The right panel in Fig 6.1 shows how the uncertainty in HC5 decreases with increasing data availability. Especially, when less then ten species are tested, the uncertainty rapidly increases with decreasing bioassay availability (Fig 6.1, right panel). Although in such cases (of low data availability) an extrapolation factor approach is preferred over the SSD approach (ECHA, 2008), the principle problem remains that omitting data (for whichever reason) reduces certainty of the hazard estimate. It is thus important that the certainty gained by omitting uncertain bioassay data should outweigh the certainty lost due to lower data availability (Dowse et al., 2013). This is also illustrated in **Chapter 3**, where the most reliable data were selected. Although these data by itself are indeed more reliable and more certain, but hazard quantification was more uncertain due to the reduction of the bioassay sample size. Fig 6.1 assists in comparing bioassay uncertainty (e.g., iii), natural variability (e.g., vi and vii) and uncertainty in hazard quantification (right panel). In a realistic data set selection process, more specific information on data uncertainty and its propagation in hazard quantification is required. In addition, its consequences for risk assessment need to be addressed.

6.3 Test species and conditions

Bioassay data relevance is for certain aspects independent on the context in which it is used. Which concentrations and exposure routes are relevant mostly depends on the type of chemical and its chemico-physical properties. While the decision which test conditions and which test species (and their condition and life-stage) are relevant depends on the context for which the risk assessment is needed. This section discusses the importance of test conditions and species selection for their use in hazard quantification and subsequent risk assessment. Their relevance in context of their use is discussed in the next section (6.4). In this section exaples of some important test conditions and their influence on effect concentrations are discussed.

Test temperature for instance can affect the bioavailability and toxicokinetics of substances but also the fitness (and therewith the sensitivity) of test species (**Chapter 4**). Fitness of test species at specific test temperatures will depend on their geographical origin (e.g., tropical, temperate or arctic; **Chapter 4**). Differences between sensitivity of species from different climatic regions are not expected (Camus et al., 2015) or small at best (Wang et al., 2014), even when effects on bioavailability are ignored. Heugens et al. (2001) reviewed the effects of temperature on toxicity of a wide range of toxicants (metals, pesticides and

natural toxicants) to specific species. They showed that an increase from 15°C to 20°C can lower toxic effect concentrations by a factor of 1.2 up to a factor of 12 (0.079 and 1.07 respectively on a $\log_{10}$-scale, for comparison with Fig 6.1), in addition to thermal stress the test species possibly endures. Effects of temperature can thus be in the same range as intertest variation, but also in the range of *inter*species variation (Fig 6.1).

Similarly, salinity can also affect bioavailability and sensitivity of test organisms. Based on the review by Heugens et al. (2001), it can be derived that a 50% increase in salinity can affect LC50 concentration by a factor ranging from a slight decrease (a factor of 0.99) to a considerable increase (a factor of 1.7) (depending on test substance and species). On an absolute $\log_{10}$-scale this is between 6×10^{-4} up to 0.23, which is in the range of *intra*species variation or lower (Fig 6.1). Obviously, species will also differ in their tolerance for salinity or temperature. Whether freshwater and saltwater species differ in sensitivity for toxicity remains unclear, as long as the real (internal) exposure concentration is not determined. Several studies found significant differences for some of their studied substances (Hutchinson et al., 1998a; Leung et al., 2001; Wheeler et al., 2002). However, differences between freshwater and saltwater species are generally less than a factor of 10 (1.0 on a $\log_{10}$-scale, for comparison with Fig 6.1) and are not structural (De Zwart, 2002; Wheeler et al., 2014).

Biotic ligand models describe how availability of substances to toxicity target sites are affected by abiotic conditions. They are therefore good examples of how abiotic conditions, such as pH, water hardness and dissolved organic carbons can affect bioavailability of toxicants (metals in particular) and therewith their effect. De Schamphelaere and Janssen (2002), for instance, showed that natural variations of pH, CO_3^{2-}, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-} and SO_4^{2-} result in a standard deviation of $\log_{10}$-transformed *Daphnia magna* copper effect concentrations of 0.23 (total copper) and 0.28 (free cupper ion), which is in the range of *intra*species variation or could even be in the range of *inter*species variation (Fig 6.1).

Exposure duration also affects the effects observed in bioassays. Kenaga (1982) showed that there is on average a 10.7 times difference between acute effect concentrations (LC50) and chronic effect concentrations (maximum acceptable toxicant concentration, MATC). This is 1.0 on a $\log_{10}$-scale and is in the range of *inter*species variation (Fig 6.1). The differences between acute and chronic (no) effect concentrations is mainly caused by both accumulation of the substance (increasing the internal exposure level) and accumulation of effects over time (e.g., small effects on survival rates may only become detectable after a longer exposure duration). However, the factor as determined by Kenaga (1982) cannot be attributed to purely the exposure duration. Using MATC, LOEC or NOEC, as done in that study, introduces additional uncertainty as they depend on the selected test concentrations and applied statistics. Which is part of the reason why its use also criticised (Jager, 2012) and can hamper hazard quantification

(**Chapter 3**). Unfortunately, in some cases these type of end-points are the only data available and are continued to be produced (Fox and Landis, 2016).

The state (i.e., life stage and conditions) of test species can also have a substantial effect. (Hutchinson et al., 1998b) showed that there can be considerable differences in sensitivity of life stages. The difference of log_{10}-transformed embryo versus larvae EC50 ranges up to 0.31 (compare with Fig 6.1). For larvae versus juveniles the maximum difference is 0.96 (fish) and 2.5 (invertebrates) (Hutchinson et al., 1998b). The maximum difference is even larger for juveniles versus adults 1.1 (fish) and 3.3 (invertebrates) (Hutchinson et al., 1998b). In general the younger and smaller life stages are more sensitive than of the older and larger stages. These differences can be less than the *intra*species variation, but can also be in the range of *inter*species variation (Fig 6.1). These differences are probably largely due to the changes of body surface area to volume ratios and other physiological changes during development that may affect the internal levels (Mohammed, 2013).

Of course specific traits of the selected test species will also affect the outcome of bioassays. Some (taxonomically) related group of species can be more sensitive for specific stressors. For instance, calcifying organisms are sensitive towards elevated CO_2 levels (**Chapter 3**). Salmonids are more sensitive towards temperature related mortality than most other (fish) species (**Chapter 4**). In general, species with specific receptors are more sensitive towards toxicants with a mode of action that targets these receptors. There is a bias in availability of bioassay data towards those generated with standardised protocols and test species. The selection of bioassay data and its test species are therefore not random and may not always reflect the target ecosystem (for which the risk is assessed) best. It is estimated that this selection bias can cause errors in hazard quantification by a factor of 20 or even more (1.3 on a log_{10}-scale for comparison with Fig 6.1) (Fox, 2015).

The variance in bioassay results as a consequence of varying test conditions, species selection and its state (life stage and condition) can be considerable. It is often in the same range as the *intra*species variance or even *inter*species variance. Selecting tests performed under specific conditions and with specific test species can thus considerably affect hazard quantification in both directions. It is important to select tests that are relevant for the situation that is under examination (to improve accuracy). An optimum should be sought between selecting enough bioassays of sufficient relevance for the target situation in risk assessment.

Species internal levels of a toxicant (where they cause their effect) will depend on temperature, pH, size, life stage, exposure duration, exposure route, toxicokinetics and other aspects. Therefore, the variance is expected to be reduced when effect concentrations are expressed as internal levels, rather than external concentrations.

6.4 Relevance and context

There can be different objectives for hazard quantification: deriving environmental quality standards; estimating impacts of specific incidents; comparing different scenarios/situations for existing or planned activities. These examples will be discussed with a focus on data selection and consequences of decisions.

Environmental quality standards are usually derived within legislative frameworks (e.g., the Water Framework Directive, EC (2000a)). Such standards require to be generically protective with a strong emphasis on the reliability (often interpreted as acceptability) of the underpinning data. As such, these standards usually will have to rely on limited data produced under standard conditions (GLP). Using standardised conditions will provide data with narrow applicability in general, as explained above. However, when such conditions are relevant for the legislative framework in which they are used, the consequences for the accuracy of hazard quantification are limited. Having to rely on a limited data, will affect the certainty of the hazard. By applying large extrapolation factors (sometimes referred to as safety or assessment factors), the derived quality standards are protective but not necessarily very realistic.

Chapter 5 presents a case where situation-specific assessment of the hazard is performed for the impacts of firefighting agents. Existing environmental quality standards were used to address the risks of substances from firefighting foams to the organisms in the water column and sediments. Including more recent and non-standard data, or less reliable data might improve the certainty of the hazard quantification. However, as non of the tests are performed with species and conditions that are relevant to the specific environmental conditions of the tropical saltwater lake, accuracy of the assessment will not improve.

The objective of the study was not only to assess the risk posed to the water column and sediments, but most importantly flamingo birds. An important part of the uncertainty of risk assessment is caused by uncertainties in the estimates of exposure levels, rather than that of the toxicity. Using specific bird toxicity data and intake estimates, toxic exposure of flamingo birds could not be excluded, with the potency to cause reproductive effects by toxic *in ovo* exposure of flamingo embryos. Finally, uncertainty also arises due to the lack of a baseline study. Other than bird density there are no field observations on the ecosystem state prior to the fire incident.

It is clear that the relevance of bioassay data used in hazard quantification does not only depend on the hazard quantification itself, but on the full risk assessment process as described in the general introduction (Chapter 1). It is important to keep in mind what the objectives are. Does a generic risk assessment suffice or should it be more realistic and relate to a specific scenario (as was the case in **Chapter 5**)? When using existing bioassay data it has to be translated to the situation in which one is interested, especially when the risk

assessment needs to be realistic and specific. An important aspect there is how to extrapolate the dosage used in the bioassay to (internal) exposure levels that are relevant to the field.

Comparison of scenarios/situations can be illustrated by the 'risk based approach', which was introduced by OSPAR for the management of offshore produced water discharges (a waste product of oil and gas extraction) (OSPAR, 2008, 2012a,b). This approach uses risk assessment (based on concentrations measured in the effluent and on existing bioassay data) to direct monitoring effort and stimulate risk reducing measures. As this purpose requires the relative comparison of the risk at two different times, it is important that the risk (and thus hazard) is precise. In such relative comparisons, systematic errors (inaccuracy) are not necessarily a problem, as long as such errors are known and unidirectional for the situations that are compared (OSPAR, 2012a).

Data requirements may also depend on the desired nature of the risk assessment, which can be more qualitative (e.g., an extrapolation factor approach; Craig (2006)) or more quantitative (e.g., a species sensitivity distribution approach; Posthuma et al. (2002)). These approaches are associated with the objectives, where environmental standards are often qualitative and relative comparisons are often quantitative. Furthermore, data availability can drive the quantitative nature of the hazard assessment (ECHA, 2008). In addition, bioassay data can be applied in more complex models, in order to assess population level effects of a specific species (e.g., Smit et al. (2006)). In such complex models highly specific data (produced for a specific species under specific conditions) are required (often at the cost of certainty and reliability). How specific (accurate) bioassay data needs to be also depends on whether the intended risk assessment is site specific (e.g., **Chapter 5**) or more generic (deriving environmental quality standards).

6.5 Conclusions

When deciding which bioassay data to include in hazard quantification and thus in risk assessment, not only reliability (in terms of reproducibility) and relevance should be considered, but also the (un)certainty of the data and its propagation in hazard quantification should be part of the data selection procedure. Focusing on reliability (using scoring systems) only can result in false representation of validity of the data set and the resulting hazard indicator. This is because selecting data can create a bias, affecting the accuracy of the resulting hazard indicator. The precision of the hazard data will be affected too.

Rather than strictly following guidelines (ECHA, 2011b,a) and narrowing down the available data via selection, it should be made more explicit, specific and critical why specific selection criteria are applied in the first place. In addition, it should be evaluated whether data selection achieved the desired

goals. For example when data is selected that is more relevant for our specific risk assessment case, a more accurate indication of the risk is achieved. However, by reducing the amount of data used, the answer will be less precise (right panel in Fig 6.1).

It is therefore proposed to select bioassay data iteratively in a risk assessment framework, rather than selecting data based on fixed selection criteria and leave it at that (Fig 6.2). Based on the objectives of the risk assessment and the availability of data, it should be decided which selection criteria are desirable and what this selection should achieve (e.g., highest certainty, accuracy and/or reliability; Fig 6.2). After the selection it should be evaluated whether the desired goals have been achieved (Fig 6.2). Furthermore, it should be recognised that some goals have conflicting interests, and cannot both be achieved simultaneously, as shown in this thesis. If, after several iterations, the selection criteria do not result in a usable set of data, new data should be generated via experiments or models.

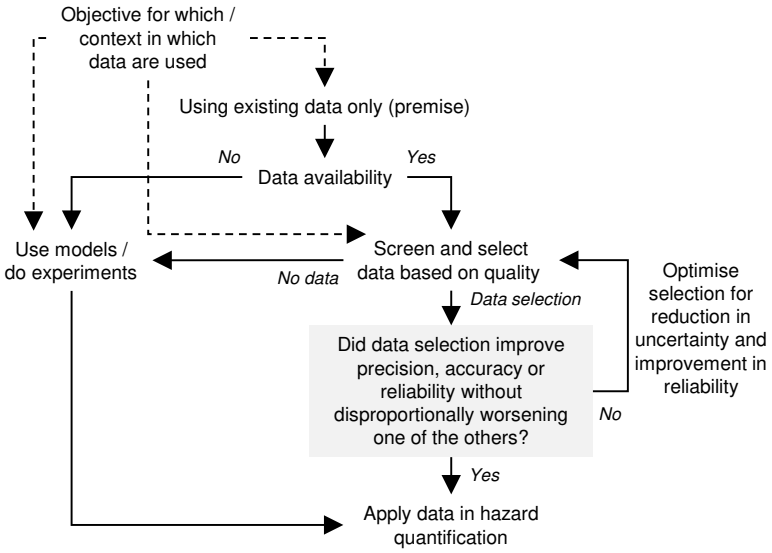


Fig. 6.2: A proposed framework for iteratively selecting data for hazard quantification.

6.6 Future perspectives

In order to compare each of the quality aspects (precision, accuracy and reliability) means are required to quantitatively compare these aspects. The improve-

ment of reliability is in particular difficult to quantify and compare with changes in certainty, resulting from the data selection procedure. This requires further research.

Another path forward is, rather than selecting data based on criteria, including all data and weighing them based on aspects such as relevance, reliability or credibility (Ågerstrand and Beronius, 2016; Mihaich et al., 2017; Semenzin et al., 2015). Weight of evidence approaches still need some maturing and guidance in the field of environmental risk assessment in order to be successfully applied (Ågerstrand and Beronius, 2016).

Focusing on available / existing data, by imposing guidelines for selecting and using the data, can slow down developments of noval approaches in bioassays and risk assessment. There already are more informative approaches to express toxic effects, rather than a simple (no) effect concentration (such as an EC50 or a NOEC). Approaches that take into account temporal aspects: how is a toxicant distributed over a species internal compartments over time via the relevant exposure routes (toxicokinetics, bioaccumulation and biomagnification), and how do they affect process rates (e.g. survival rates) of that species. These aspects should be considered when designing a new experiment. Tools to analyse these type of experiments are available (such as DEBtox, Jager and Zimmer (2012); Kooijman et al. (1996)). But when relying on (non-ideal) existing data, standardised tools to translate field relevant concentrations to internal concentrations and how these affect species that are relevant for the risk assessment are currently not available.

References

Ågerstrand, M., Beronius, A., 2016. Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. Environ Int 92-93, 590–596.

Aldenberg, T., Jaworska, J. S., 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotox Environ Safe 46 (1), 1–18.

Aldenberg, T., Slob, W., 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotox Environ Safe 25 (1), 48–63.

Beasley, A., Belanger, S. E., Otter, R. R., 2015. Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way. Environ Toxicol Chem 34 (6), 1436–1442.

Borgert, C. J., Becker, R. A., Carlton, B. D., Hanson, M., Kwiatkowski, P. L., Sue Marty, M., McCarty, L. S., Quill, T. F., Solomon, K., Van Der Kraak, G.,

Witorsch, R. J., Don Yi, K., 2016. Does GLP enhance the quality of toxicological evidence for regulatory decisions? Toxicol Sci 151 (2), 206–213.

Camus, L., Brooks, S., Geraudie, P., Hjorth, M., Nahrgang, J., Olsen, G., Smit, M., 2015. Comparison of produced water toxicity to arctic and temperate species. Ecotox Envir Safe 113 (Supplement C), 248–258.

Craig, P., 2006. Uncertainty factors in ectotoxicological risk: uncertainty factors in ectotoxicological risk management. 4th Edinburgh risk conference, Edinburgh.

De Schamphelaere, K. A. C., Janssen, C. R., 2002. A biotic ligand model predicting acute copper toxicity for *Daphnia magna*: The effects of calcium, magnesium, sodium, potassium, and pH. Environ Sci Technol 36 (1), 48–54.

De Zwart, D., 2002. Observed regularities in species sensitivity distributions for aquatic species. In: Posthuma, L., Suter II, G. W., Traas, T. P. (Eds.), Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, US, pp. 133–154.

Dowse, R., Tang, D., Palmer, C. G., Kefford, B. J., 2013. Risk assessment using the species sensitivity distribution method: Data quality versus data quantity. Environ Toxicol Chem 32 (6), 1360–1369.

Dyer, S. D., Versteeg, D. J., Belanger, S. E., Chaney, J. G., Raimondo, S., Barron, M. G., 2008. Comparison of species sensitivity distributions derived from interspecies correlation models to distributions used to derive water quality criteria. Environ Sci Technol 42 (8), 3076–3083.

EC, 2000a. 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. Off J Eur Comm L327, 1–72.

EC, 2000b. First report on the harmonisation of risk assessment procedures. part 1: The report of the scientific steering committee's working group on harmonisation of risk assessment procedures in the scientific committees advising the European Commission in the area of human and environmental health 26-27 October 2000.

ECHA, 2008. Guidance on information requirements and chemical safety assessment chapter R.10: Characterisation of dose [concentration]-response for environment.

ECHA, 2011a. Guidance on information requirements and chemical safety assessment chapter R.3: Information gathering. version 1.1.

ECHA, 2011b. Guidance on information requirements and chemical safety assessment chapter R.4: Evaluation of available information. version 1.1.

ECHA, 2012. Guidance on information requirements and chemical safety assess-

ment chapter R.19: Uncertainty analysis. version 2.1.

Férard, G., Férard, J.-F., 2013. Quality assurance in ecotoxicology: The why's and how's. In: Férard, J.-F., Blaise, C. (Eds.), Encyclopedia of Aquatic Ecotoxicology. Springer Netherlands, Dordrecht, pp. 927–944.

Flanagan, R., Taylor, A., Watson, I. D., Whelpton, R., 2007. Fundamentals of analytical toxicology. Wiley, Chichester, UK.

Forbes, V., Hall, T., Suter, G. W., 2016. The challenge: Bias is creeping into the science behind risk assessments and undermining its use and credibility. Environ Toxicol Chem 35 (5), 1068–1068.

Fox, D. R., 2015. Selection bias correction for species sensitivity distribution modeling and hazardous concentration estimation. Environ Toxicol Chem 34 (11).

Fox, D. R., Landis, W. G., 2016. Comment on *ET&C* perspectives, november 2015 – a holistic view. Environ Toxicol Chem 35 (6), 1337–1339.

Gaudet-Hull, A. M., Rayburn, J. R., Bantle, J. A., Burton, D. T., Turley, S. D., Dawson, D. A., Dumont, J. N., Finch, R. A., Maurice, M. A., Fort, D. J., Linder, G., Buchwalter, D., 1994. Fetax interlaboratory validation study: Phase II testing. Environ Toxicol Chem 13 (10), 1629–1637.

Grothe, D. R., Kimerle, R. A., 1985. Inter- and intralaboratory variability in *Daphnia magna* effluent toxicity test results. Environ Toxicol Chem 4 (2), 189–192.

Harbers, J. V., Huijbregts, M. A. J., Posthuma, L., Van de Meent, D., 2006. Estimating the impact of high-production-volume chemicals on remote ecosystems by toxic pressure calculation. Environ Sci Technol 40 (5), 1573–1580.

Harris, C. A., Sumpter, J. P., 2015. Could the quality of published ecotoxicological research be better? Environ Sci Technol 49 (16), 9495–9496.

Hendriks, A. J., Awkerman, J. A., De Zwart, D., Huijbregts, M. A., 2013. Sensitivity of species to chemicals: Dose–response characteristics for various test types (LC50, LR50 and LD50) and modes of action. Ecotox Environ Safe 97, 10–16.

Heugens, E. H. W., Hendriks, A. J., Dekker, T., Van Straalen, N. M., Admiraal, W., 2001. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. Crit Rev Toxicol 31 (3), 247–284.

Hutchinson, T. H., Scholz, N., Guhl, W., 1998a. Analysis of the Ecetoc aquatic toxicity (EAT) database IV – comparative toxicity of chemical substances to freshwater versus saltwater organisms. Chemosphere 36 (1), 143–153.

Hutchinson, T. H., Solbe, J., Kloepper-Sams, P. J., 1998b. Analysis of the Ecetoc aquatic toxicity (EAT) database III – comparative toxicity of chemical substances to different life stages of aquatic organisms. Chemosphere 36 (1), 129–

142.

Isigonis, P., Ciffroy, P., Zabeo, A., Semenzin, E., Critto, A., Giove, S., Marcomini, A., 2015. A multi-criteria decision analysis based methodology for quantitatively scoring the reliability and relevance of ecotoxicological data. Sci Total Environ 538, 102–116.

ISO, 2013. ISO 19157:2013 geographic information – data quality.

Jager, T., 2012. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. Environ Toxicol Chem 31 (2), 228–229.

Jager, T., Zimmer, E. I., 2012. Simplified dynamic energy budget model for analysing ecotoxicity data. Ecological Modelling 225 (Supplement C), 74–81.

Kase, R., Korkaric, M., Werner, I., Ågerstrand, M., 2016. Criteria for reporting and evaluating ecotoxicity data (CRED): comparison and perception of the Klimisch and CRED methods for evaluating reliability and relevance of ecotoxicity studies. Env Sci Eur 28 (1), 7.

Kenaga, E. E., 1982. Predictability of chronic toxicity from acute toxicity of chemicals in fish and aquatic invertebrates. Environ Toxicol Chem 1 (4), 347–358.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharm 25 (1), 1–5.

Kooijman, S. A. L. M., Bedaux, J. J. M., Slob, W., 1996. No-effect concentration as a basis for ecological risk assessment. Risk Anal 16 (4), 445–447.

Leung, K. M., Morritt, D., Wheeler, J. R., Whitehouse, P., Sorokin, N., Toy, R., Holt, M., Crane, M., 2001. Can saltwater toxicity be predicted from freshwater data? Mar Pollut Bull 42 (11), 1007–1013.

Lynch, H. N., Goodman, J. E., Tabony, J. A., Rhomberg, L. R., 2016. Systematic comparison of study quality criteria. Regul Toxicol Pharm 76, 187–198.

McCarty, L. S., Borgert, C. J., Mihaich, E. M., 2012. Information quality in regulatory decision making: Peer review versus good laboratory practice. Environ Health Perspect 120 (2), 927–934.

Mihaich, E., Capdevielle, M., Urbach-Ross, D., Slezak, B., 2017. Hypothesis-driven weight-of-evidence analysis of endocrine disruption potential: a case study with triclosan. Crit Rev Toxicol 47 (4), 263–285.

Moermond, C., Beasley, A., Breton, R., Junghans, M., Laskowski, R., Solomon, K., Zahner, H., 2017. Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches. Integr Environ Assess Manag 13 (4), 640–651.

Moermond, C. T. A., Kase, R., Korkaric, M., Ågerstrand, M., 2016. CRED: Criteria

for reporting and evaluating ecotoxicity data. Environ Toxicol Chem 35 (5), 1297–1309.

Mohammed, A., 2013. Why are early life stages of aquatic organisms more sensitive to toxicants than adults? In: Gowder, S. (Ed.), New Insights into Toxicity and Drug Testing. InTech, Rijeka, Ch. 3.

OECD, 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. No. 1 OECD Principles of Good Laboratory Practice (as revised in 1997). Organisation for Economic Co-operation and Development, Paris.

OSPAR, 2008. Meeting of the Offshore Industry Committee (OIC): Summary record. 2008, report: OIC 08/13/1-E.

OSPAR, 2012a. OSPAR guidelines in support of recommendation 2012/5 for a risk-based approach to the management of produced water discharges from offshore installations (OSPAR agreement: 2012-7).

OSPAR, 2012b. OSPAR recommendation 2012/5 for a risk-based approach to the management of produced water discharges from offshore installations.

Posthuma, L., Suter II, G. W., Traas, T. P., 2002. Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, US.

Segal, D., Makris, S. L., Kraft, A. D., Bale, A. S., Fox, J., Gilbert, M., Bergfelt, D. R., Raffaele, K. C., Blain, R. B., Fedak, K. M., Selgrade, M. K., Crofton, K. M., 2015. Evaluation of the ToxRTool's ability to rate the reliability of toxicological data for human health hazard assessments. Regul Toxicol Pharm 72 (1), 94–101.

Semenzin, E., Lanzellotto, E., Hristozov, D., Critto, A., Zabeo, A., Giubilato, E., Marcomini, A., 2015. Species sensitivity weighted distribution for ecological risk assessment of engineered nanomaterials: The *n*-TiO_2 case study. Environ Toxicol Chem 34 (11), 2644–2659.

Smit, M. G. D., Hendriks, A. J., Schobben, J. H. M., Karman, C. C., Schobben, H. P. M., 2001. The variation in slope of concentration–effect relationships. Ecotox Environ Safe 48 (1), 43–50.

Smit, M. G. D., Kater, B. J., Jak, R. G., Van den Heuvel-Greve, M. J., 2006. Translating bioassay results to field population responses using a Leslie-matrix model for the marine amphipod *Corophium volutator*. Ecol Model 196 (3), 515–526.

Straalen, N. M. V., Denneman, C. A. J., 1989. Ecotoxicological evaluation of soil quality criteria. Ecotox Environ Safe 18 (3), 241–251.

Wang, Z., Kwok, K. W. H., Lui, G. C. S., Zhou, G.-J., Lee, J.-S., Lam, M. H. W., Leung, K. M. Y., 2014. The difference between temperate and tropical saltwater species' acute sensitivity to chemicals is relatively small. Chemosphere

105 (Supplement C), 31–43.

Wheeler, J. R., Leung, K. M. Y., Morritt, D., Sorokin, N., Rogers, H., Toy, R., Holt, M., Whitehouse, P., Crane, M., 2002. Freshwater to saltwater toxicity extrapolation using species sensitivity distributions. Environ Toxicol Chem 21 (11), 2459–2467.

Wheeler, J. R., Maynard, S. K., Crane, M., 2014. Are acute and chronic saltwater fish studies required for plant protection and biocidal product active substance risk assessment? Environ Toxicol Chem 33 (3), 703–707.

6.A Appendix

6.A.1 Details for Fig 6.1

Fig. 6.1 in the main text is based upon the following information and approaches.

Both *inter*- and *intra*laboratory variation ((i) and (ii) in Fig. 6.1) are based on acute EC50 and LC50 data from ring tests with effluent on *Daphnia magna* and sodium cyclamate on frog embryos (Gaudet-Hull et al., 1994; Grothe and Kimerle, 1985). Although more ring tests have been published, these two contain raw data, from which the $\log_{10}$-transformed standard deviations could be calculated.

*Intra*test variation ((iii) in Fig. 6.1) is based on EC50 and LC50 data in both saltwater and freshwater with species labelled as 'standard', extracted from the US EPA ECOTOX database[22]. The standard deviation is in this case calculated as $\log_{10}\left(\frac{\text{UCL}}{\text{LCL}}\right)/(2\times 1.96)$, where UCL and LCL are the reported 95% upper and lower confidence limits of the effect concentration (EC50 or LC50) respectively.

Uncertainty in Temperature Tolerance Intervals ($\sigma_{\text{TTI,regression}}$, resulting from linear regression) ((iv) in Fig. 6.1) cannot be compared directly with that of toxicity data, because the values are not $\log_{10}$-transformed and represent a different quantity (temperature rather than concentration). Therefore, the $\sigma_{\text{TTI,regression}}$ is scaled as follows: $\sigma_{\text{TTI,regression}} \times \sigma_{\text{tox}}/\sigma_{\text{TTI}}$; where σ_{tox} is the median slope of toxicant SSDs from (Harbers et al., 2006) and σ_{TTI} is the slope of the temperature effect SSD from **Chapter 4**. Uncertainty is shown for an acclimatisation temperature (T_a) of 5°C, for which the uncertainty was largest.

*Inter*species correlation estimates ((v) in Fig. 6.1) are based on an SSD data generated with the webICE model[23], using *Daphnia magna* as a surrogate species with a hypothetical effect concentration of 100 μg/L. This model presents the 95% confidence intervals for the estimated effect concentrations, based on linear regression. Like before, these confidence intervals are transformed into standard deviations using $\log_{10}\left(\frac{\text{UCL}}{\text{LCL}}\right)/(2\times 1.96)$.

The slope of dose-response curves ((vi) in Fig. 6.1) are reported for effect concentrations transformed with natural logarithms, in the original publication. The standard deviation is converted to 10-based logarithms by dividing reported values by ln(10).

The slopes of SSDs ((vii) in Fig. 6.1) are reported as shape parameters β for log-logistic distributions. They are converted to standard deviations by multiplying them with $\pi/\sqrt{3}$.

The uncertainty in HC5 (solid line in right panel of Fig. 6.1) cannot be properly expressed as a standard deviation as it is skewed. For comparability with the other uncertainties, a proxy of the standard deviation is obtained by dividing the

[22] `ftp://ftp.epa.gov/pub/ecotox/ecotox_ascii_06_15_2017.exe`, accessed on 14 September 2017

[23] `https://www3.epa.gov/webice/iceSSDSpecies.html?filename=as`, accessed on 15 September 2017

two-sided 68.2% confidence interval (calculated with the method described by Aldenberg and Jaworska (2000)) by two.

References in Appendix 6.A

Aldenberg, T., Jaworska, J. S., 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotox Environ Safe 46 (1), 1–18.

Gaudet-Hull, A. M., Rayburn, J. R., Bantle, J. A., Burton, D. T., Turley, S. D., Dawson, D. A., Dumont, J. N., Finch, R. A., Maurice, M. A., Fort, D. J., Linder, G., Buchwalter, D., 1994. Fetax interlaboratory validation study: Phase II testing. Environ Toxicol Chem 13 (10), 1629–1637.

Grothe, D. R., Kimerle, R. A., 1985. Inter- and intralaboratory variability in *Daphnia magna* effluent toxicity test results. Environ Toxicol Chem 4 (2), 189–192.

Harbers, J. V., Huijbregts, M. A. J., Posthuma, L., Van de Meent, D., 2006. Estimating the impact of high-production-volume chemicals on remote ecosystems by toxic pressure calculation. Environ Sci Technol 40 (5), 1573–1580.

Glossary of acronyms and abbreviations

An overview of frequently used or otherwise important acronyms and abbreviations and their meaning are presented here for your convenience.

CAS	Chemical Abstract Service; a division of the American Chemical Society that provides unique, unmistakable identifiers for chemical substances
CO_2	Carbon dioxide; a trace gas in the Earth's atmosphere
EC50	Median effect concentration; the exposure concentration of a substance which induces a response halfway between the baseline and maximum after a specified exposure time
ECHA	European Chemicals Agency; the driving force among regulatory authorities in implementing the EU's chemicals legislation for the benefit of human health and the environment
ECOTOX	The ECOTOXicology database; a source for locating single chemical toxicity data for aquatic life, terrestrial plants and wildlife, created and maintained by the US EPA
EQS	Environmental Quality Standard; A threshold concentration below which effects are not likely to occur in the environment
GLP	Good Laboratory Practice; a quality system of management controls for research laboratories and organisations to try to ensure the uniformity, consistency, reliability, reproducibility, quality, and integrity of chemical (including pharmaceuticals) non-clinical safety tests; from physio-chemical properties through acute to chronic toxicity tests
HC5	5% Hazardous concentration; the concentration at which no more than 5% of specific species are exposed above their effect levels (usually based on chronic NOECs)

HC50	50% Hazardous concentration; the concentration at which no more than 50% of specific species are exposed above their effect levels (usually based on chronic NOECs)
HTI5	5% Hazardous temperature interval; a temperature increase above a specified acclimation temperature that causes at least 50% mortality for potentially 5% of the species
HTI50	50% Hazardous temperature interval; a temperature increase above a specified acclimation temperature that causes at least 50% mortality for potentially 50% of the species
LC50	Median lethal concentration; the exposure concentration of a substance which induces mortality halfway between the baseline and maximum mortality after a specified exposure time
LOEC	Lowest Observed Effect Concentration; the lowest test concentration of a substance at which significant (adverse) effects are observed within a specified exposure time
LT10	10% lethal temperature; the temperature which induces 10% mortality between the baseline and maximum mortality after a specified exposure time
LT50	Median lethal temperature; the temperature which induces mortality halfway between the baseline and maximum mortality after a specified exposure time
NOEC	No Observed Effect Concentration; the highest test concentration of a substance at which no significant (adverse) effects are observed within a specified exposure time
OECD	Organisation for Economic Co-operation and Development; organisation that promotes policies that will improve the economic and social well-being of people around the world
OSPAR	Oslo Paris convention; the mechanism by which fifteen Governments of the western coasts and catchments of Europe, together with the European Community, cooperate to protect the marine environment of the North-East Atlantic
PAF	Potentially Affected Fraction; ecological risk indicator giving the fraction of species that are potentially affected at a specified level
pCO_2	the partial pressure of carbon dioxide (CO_2)
PNEC	Predicted No-Effect Concentration; the concentration of a substance below which exposure to a substance is not expected to cause adverse effects to species in the environment
PFAS	*Per-* and polyfluoroalkyl substances. A diverse group of compounds resistant to heat, water and oil. Present in, amongst others, specific firefighting foams
PFOS	Perfluorooctanyl sulphonic acid. A specific PFAS

REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals; EU regulation whose aims are to ensure a high level of protection of human health and the environment from the risks that can be posed by chemicals, the promotion of alternative test methods, the free circulation of substances on the internal market and enhancing competitiveness and innovation
SSD	Species Sensitivity Distribution; represents the variation in sensitivity of species to a stressor by a statistical or empirical distribution function of responses for a sample of species
TTI	Temperature Tolerance Interval; the interval by which a temperature increase above the acclimation temperature causes 50% mortality after a specified exposure time
US EPA	United States Environmental Protection Agency; United States organisation that aims at the protection of human health and the environment

Summary

Ecotoxicological risk deals with the adverse effects experienced by organisms in an ecosystem caused by exposure to one or more toxicants. Before risk can be assessed the toxicant's hazard needs to be quantified. Hazard is the potency of a substance to cause harm to the environment. It is commonly based on existing *in vivo* bioassay data. These assays are laboratory experiments in which species are exposed to different concentrations. Effects caused by the toxicant and experienced by test species, such as reduced survival or growth, are recorded. Results from such experiments are traditionally summarised in so-called (no) effect concentrations, where 50% effect concentrations (EC50) and no observed effect concentrations (NOEC) are common descriptors. These are concentrations at which either a specific or respectively no effect is observed and are used to quantify the hazard of a substance.

In ecotoxicological risk assessment it is common practise to use databases with vast amounts of *in vivo* bioassay data. Data from such databases need to be screened and evaluated for adequacy before they can be used in ecotoxicological risk assessment. European guidelines are in place for the risk assessment process as well as for the data selection and screening process. Current guidelines consider bioassay data to be adequate when they are reliable and relevant for the purpose for which they are used. Using several case studies, this thesis examines how these guidelines for bioassay data selection, which focus on specific aspects of reliability and relevance, can be improved.

Currently, several approaches are available to classify the reliability of bioassay data qualitatively. This process of classification, such as using the Klimisch score, is time-consuming and focuses on the application of standardised protocols and the documentation of the study that generated the data. The presence of irregularities and (un)intentional errors, however, is not addressed. **Chapter 2** shows that Benford's Law, based on the occurrence of the data's first digits following a logarithmic scale, can be applied to bioassay data for identifying irregularities. This approach is already successfully applied in accounting. Benford's Law can be used as a reliability indicator, in addition to existing reliability classifications. The law can be used to efficiently trace irregularities in large

data sets of interpolated (no) effect concentrations such as EC50s (possibly the result of data manipulation), without having to evaluate the source of each individual record. Application of the law to systems in which large amounts of toxicity data are registered (e.g., European Commission Regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals) can therefore be valuable.

In some cases, protocols of bioassays remain yet to be standardised, resulting in little availability of reliable data. In **Chapter 3** it is studied how lack of availability of reliable data affects hazard quantification, focusing on bioassays with elevated carbon dioxide (CO_2) levels. The environmental impact of elevated CO_2 levels has become of more interest in recent years. This in relation to globally rising CO_2 levels and related considerations of geological CO_2 storage as a mitigating measure. In **Chapter 3** effect data from literature were collected in order to quantify the hazard of elevated CO_2 levels to marine species. It became evident that information currently available from the literature is mostly insufficient for a quantitative approach. Most studies focus on effects of expected future CO_2 levels, testing only one or two elevated concentrations. A full dose-response relationship, a uniform measure of exposure, and standardised test protocols are essential for quantifying the hazard of elevated CO_2 levels reliably. When only the most reliable CO_2 bioassay data are selected, the estimated hazard level does not change much, but the uncertainty in the estimated hazard level increases considerably as the size of the data pool is reduced. It is therefore argued that the certainty in bioassay data and consequences for certainty of the hazard level estimate needs to be an integral part of the data selection process and not only later when the risk is being evaluated.

Other non-toxic stressors have been studied with higher level of standardisation and certainty, such as thermal stress. In **Chapter 4** temperature sensitivity data is based on bioassays for 50 different aquatic species. These data are used to quantify hazard of heat discharges that can be used to assess risk of situations that are localised in space or time, as is demonstrated for a specific case of a thermal discharge of a power plant in the North Sea Canal in the Netherlands. As the hazard is quantified using an approach that is also commonly used for toxicants and some other stressors, it can be used for an integrative risk assessment of multiple stressors. Problems with certainty and reliability of the data were not as problematic as they were for **Chapter 3**.

In addition to the certainty and reliability, the relevance of bioassay data is also important when applied in hazard quantification and thus risk assessment. Bioassay data is considered relevant when it is based on experiments in which relevant species have been tested with exposure concentrations, exposure routes and following conditions that are considered relevant. This relevance naturally depends on the context within and objective for which the data is used. Therefore, a real field case is studied in **Chapter 5** in order to evaluate relevance of bioassay

data in risk assessment.

In 2010 an oil terminal next to nature reservation Saliña Goto (Bonaire) caught fire. Firefighting resulted in elevated levels of specific toxicants in the salt lake. Within months flamingo abundance in Goto dropped to near complete absence. Toxicological effects on abundance of prey are likely the main cause for the flamingo absence. As flamingos were not present in the years after the fire, they were not exposed to food contaminated with the toxicants. Although the sediment is still polluted, flamingos returned, and started to feed on contaminated organisms. Based on estimated intake, the birds and other wildlife are considered to be at risk of being exposed to levels that can cause adverse effects. Monitoring bird populations is advised to assess potential toxic effects on birds and their offspring. This case study also revealed some uncertainties, among which the mismatch between the tropical and saline conditions in the study area versus the generally temperate and freshwater conditions in the lab on which hazard indicators were based.

This mismatch is partly due to the current stimulation of generating bioassay data following standardised protocols. These conditions, such as water temperature; salinity; acidity and dissolved organic carbons, can have a considerable effect on toxicant effect concentrations. Differences in effect concentrations caused by (natural) variance of these conditions can be high and in the same range as differences between the sensitivity of species for these toxicants. A mismatch between conditions used in laboratory experiments and the field can thus result in inaccurate hazard indicators and risk assessment, unless it is corrected for.

The same is true for species and life-stages selected for laboratory experiments. Although on average there is little difference found between species from different climatic regions and between freshwater and marine water species, there is variation among species. Hence hazard quantification can be biased when species are selected non-randomly. Life-stages also respond differently to toxicants. Typically, smaller and younger life-stages are more sensitive than the larger older life-stages.

Reliable bioassay data are generally considered data that are reproducible. This is currently evaluated by applying a scoring system. Many of the existing scoring systems focus on the level of standardisation and the documentation of an experiment. Selecting reliable bioassays may therefore not necessarily yield the most relevant data. In addition, the scoring systems are no guarantee for error-free data (**Chapter 2**) and they are difficult to compare quantitatively with other quality aspects, such as relevance and statistical certainty.

Current European guidelines for selecting bioassay data, for the use in hazard quantification and eventually risk assessment, focus on reliability and relevance of the bioassay data. The selection procedure itself will affect the outcome and certainty of the hazard quantification. It is therefore argued that the certainty of the bioassay data and that of the subsequent hazard assessment, should be

considered in the data selection stage. This in contrast to current practise when it is done in a later stage during risk assessment. In addition, guidelines are not explicit on why specific selection criteria should be applied. It is proposed to make goals of the data selection process explicit (how much should reliability, relevance and certainty improve and why?). Then apply these criteria in an iterative selection process, where in each iteration the consequences of the applied selection criteria are evaluated against the goals set. Criteria can be adjusted each iteration in order to optimise the balance between reliable, relevant and certain bioassay data.

Further research should focus on a framework for (quantitatively) comparing the data quality aspects of reliability, relevance and certainty, for optimising the data selection process. One path that can be taken is using all data with a weight of evidence approach is, instead of selecting subsets of data for hazard quantification. Although the work presented in this thesis focuses on existing *in vivo* bioassay data, it is encouraged to also include noval experimental and statistical techniques in hazard quantification. Nonetheless, existing bioassay data provide a valuable source of information for hazard quantification when selection criteria are well balanced. This is especially true when proper corrections are applied to translate data from the laboratory conditions under which they were generated to field conditions relevant to the context in which they are applied.

Samenvatting

Bij ecotoxicologische risico's gaat het om ongewenste effecten die door organismen in een ecosysteem worden ondervonden door de blootstelling aan één of meer gifstoffen. Voordat bepaald kan worden hoe ernstig het risico is, moet eerst de 'hazard' beoordeeld worden. 'Hazard' is een Engels woord zonder goede Nederlandse vertaling. In feite geeft 'hazard' aan hoe groot de potentie van een stof is om schade aan te richten aan het milieu. Dit wordt onder andere met behulp van levende testorganismen in het laboratorium bepaald. Deze testen worden in het Engels '*in vivo* bioassays' genoemd. Effecten, zoals verminderde overleving en voortplanting, veroorzaakt door gifstoffen worden in dergelijke experimenten bepaald bij verschillende blootstellingsconcentraties. Resultaten van bioassays worden vaak samengevat in een (geen) effect concentratie. Voorbeelden zijn de EC50, de concentratie waarbij 50% effect optreedt en de NOEC, de hoogst geteste concentratie waarbij geen effect is waargenomen.

Resultaten van bioassays worden vastgelegd in databases, waarvan bij ecotoxicologische risicobeoordeling dankbaar gebruik wordt gemaakt van deze bestaande gegevens. Bestaande data moeten natuurlijk wel geschikt zijn voor het doel waarvoor het wordt gebruikt. Er bestaan daarom richtlijnen voor het selecteren en beoordelen van de data op relevantie en betrouwbaarheid. In dit proefschrift is, aan de hand van verschillende casestudies, onderzocht hoe deze richtlijnen voor het selecteren en evalueren van bioassay data voor het gebruik bij 'hazard' inschattingen kunnen worden verbeterd.

Voor het inschatten van betrouwbaarheid van bioassay data bestaan op dit moment verschillende classificatiesystemen. Het proces van classificeren, zoals bijvoorbeeld het toekennen van zogenaamde Klimisch score, is een tijdrovend karwei dat beoordeelt of een test is uitgevoerd volgens gestandaardiseerde protocollen en of de test goed gedocumenteerd is. De aanwezigheid van onregelmatigheden en (on)bewuste fouten in de gegevens wordt daarmee echter niet getoetst. **Hoofdstuk 2** laat zien dat de wet van Benford kan worden toegepast voor het onthullen van dergelijke onregelmatigheden in bioassay gegevens. Deze wet neemt aan dat de aanwezigheid van het eerste cijfer in natuurlijke gegevens een specifieke logaritmische verdeling heeft en wordt al succesvol toegepast in

bijvoorbeeld accounting. De wet kan worden toegepast om afwijkingen (mogelijk als gevolg van manipulatie) op te sporen in grote datasets met (geïnterpoleerde) (geen) effect concentraties zoals de EC50, zonder eerst elk individueel gegeven te moeten beoordelen. Deze aanpak is daarom vooral geschikt en waardevol voor systemen waarin grote hoeveelheden gegevens worden geregistreerd en de belangen groot en mogelijk verstrengeld zijn. De EU verordening 1907/2006 ('*inzake de registratie en beoordeling van en de vergunningverlening en beperkingen ten aanzien van chemische stoffen (REACH)*') is een voorbeeld van een kader waarin toepassing de wet van Benford nut kan hebben.

In sommige gevallen zijn protocollen voor bioassays (nog) niet gestandaardiseerd, en zijn er weinig tot geen betrouwbare gegevens beschikbaar. **Hoofdstuk 3** beschrijft hoe zo'n gebrek aan betrouwbare bioassay gegevens de beoordeling van 'hazard' wordt beïnvloed, door te richten op bioassays waarin de effecten van verhoogde koolstofdioxide (CO_2) gehalten zijn bestudeerd. Deze studies kennen een groeiende belangstelling door de wereldwijde stijging van CO_2 gehalten in de atmosfeer, en verzachtende maatregelen zoals de opslag van CO_2 in uitgeputte offshore aardgasreservoirs. Voor dit doeleinde zijn bioassay effectstudies verzameld uit de literatuur om een kwantitatieve inschatting te maken van de 'hazard' van de blootstelling van mariene organismen aan verhoogde CO_2 gehalten. Hieruit werd duidelijk dat er onvoldoende geschikte gegevens waren voor een dergelijke aanpak. Veel van de verzamelde studies bestudeerden alleen de effecten van een mogelijk toekomstig gehalte aan CO_2 en testten daarbij naast het controle experiment maar één of twee verhoogde CO_2 gehalten. Daarnaast zijn een aantal andere zaken essentieel voor een betrouwbare inschatting van het CO2 'hazard'. Waaronder een beschrijving van de volledige kwantitatieve relatie tussen blootstelling en effect; een consistente maat voor de beschrijving van effect concentraties; en gestandaardiseerde protocollen. Voor het geval van CO2 ontbraken deze veelal of waren beperkt. Wanneer alleen de meest betrouwbare gegevens worden geselecteerd uit de verzamelde literatuur, wordt de omvang van de gegevens set sterk verkleind. Hiermee wordt ook de zekerheid van de 'hazard' kleiner. Het 'hazard' niveau verandert weinig door de strengere selectie, maar de onzekerheid neemt wel toe. Al bij de selectie van bioassay gegevens zou dus niet alleen rekening moeten worden gehouden met de betrouwbaarheid van de gegevens, maar ook met de gevolgen van die selectie voor de zekerheid in het 'hazard' niveau.

Er zijn ook andere niet-giftige drukfactoren, zoals thermische stress, welke ook met bioassays worden bestudeerd, maar dan met een hogere maat van standaardisatie. **Hoofdstuk 4** beschouwt de temperatuurgevoeligheid, gebaseerd op bioassays. Van 50 verschillende aquatische soorten, zijn deze gegevens gebruikt om het 'hazard' niveau van warmtelozingen te bepalen en wordt gedemonstreerd aan de hand van het voorbeeld van een koelwaterlozing van een energiecentrale in het Noordzeekanaal. Aangezien de gekozen aanpak breed gebruikt wordt

voor gifstoffen en andere drukfactoren en gebruik maakt van kanswerking, kan deze onder specifieke aannames gebruikt worden bij de risicobeoordeling van gecombineerde drukfactoren. Aspecten met betrekking tot betrouwbaarheid en zekerheid van 'hazard' beoordeling vormden in dit geval geen belemmering zoals dit voor CO_2 wel het geval was.

Niet alleen de zekerheid en betrouwbaarheid spelen een rol bij de selectie van bioassay gegevens bij 'hazard'- en dus risicobeoordeling. De gegevens moeten ook relevant zijn. Bioassays worden beschouwd als zijnde relevant wanneer de relevante organismen zijn getest; bij relevante blootstellingsconcentraties (gehalten die in het veld te verwachten zijn); bij relevante blootstellingsroutes (bijvoorbeeld via het voedsel of via de waterkolom); en onder relevante omstandigheden (bijvoorbeeld temperatuur en zuurgraad). De relevantie hangt daarom ook af van de context waarin en het doel waarvoor de gegevens worden gebruikt. De relevantie van bioassay gegevens is in **Hoofdstuk 5** bestudeerd aan de hand van een risicobeoordeling in een werkelijke veldsituatie.

Op Bonaire is naast het natuurreservaat Saliña Goto in 2010 een petrochemische brand ontstaan bij een olieoverslag als gevolg van blikseminslag. Bij de brand en de bestrijding ervan zijn verontreinigende stoffen in het zoute meer (Saliña Goto) terecht gekomen. Binnen enkele maanden na de brand zijn de aantallen flamingo's bij het meer gedaald tot het niveau van nagenoeg afwezigheid. Het ontbreken van voedsel door de verontreiniging wordt als belangrijke oorzaak hiervan beschouwd. Tijdens hun afwezigheid konden flamingo's niet worden blootgesteld aan de vervuiling in het meer. Hoewel het meer nog steeds is verontreinigd, keerden flamingo's terug en voeden zij zich nu met verontreinigde prooidieren uit het meer. Op basis van schattingen van voedselinname kunnen ongewenste effecten op flamingo's, als gevolg van de verontreiniging, niet worden uitgesloten. Het wordt geadviseerd om de vogelpopulatie te blijven volgen en daarbij bij voorkeur werkelijke (interne) blootstellingsgehalten te bepalen om potentiele toxische effecten nauwkeuriger te kunnen beoordelen. Deze case studie laat een aantal onzekerheden zien bij de risicobeoordeling, waaronder het ontbreken van bioassays die corresponderen met de tropische, zoute omstandigheden in het meer. De bioassays waarop de gebruikte waterkwaliteitsnormen zijn gebaseerd zijn vooral uitgevoerd met zoetwatersoorten uit een gematigd klimaat met daarbij horende testomstandigheden.

Deze 'mismatch' ontstaat gedeeltelijk doordat het genereren en gebruik van bioassays met gestandaardiseerde protocollen wordt gestimuleerd door huidige richtlijnen. Testomstandigheden in deze protocollen, zoals water temperatuur; zoutgehalte; zuurgraad; en opgelost organisch koolstof kunnen een aanzienlijk effect hebben op de effect concentratie van een stof. Verschillen in effect concentraties, veroorzaakt door (natuurlijke) variatie van deze omstandigheden, zijn groot. Ze kunnen in dezelfde orde grootte liggen als de variatie in gevoe-

ligheid van verschillende organismen voor dezelfde stof. Wanneer de testomstandigheden afwijken van de omstandigheden in het veld, kan dit leiden tot een onnauwkeurige inschatting van de 'hazard', tenzij voor de afwijkende omstandigheden gecorrigeerd wordt.

Hetzelfde geldt ook voor de gekozen testorganismen en het levensstadium dat getest wordt. Gemiddeld genomen is er weinig tot geen verschil in gevoeligheid van soorten uit verschillende klimatologische regio's of tussen zoetwater en zoutwater organismen. Desalniettemin is er variatie in gevoeligheid van soorten. Dit betekent dat wanneer testorganismen niet willekeurig gekozen zijn, er een bevooroordeeld beeld kan ontstaan van de 'hazard'. Daarbij komt dat verschillende levensstadia (bijvoorbeeld, larve, juveniel, adult) verschillend kunnen reageren op dezelfde dosis van een stof. Doorgaans zijn de jongere en kleinere levensstadia gevoeliger dan de oudere en grotere levensstadia. Bij de selectie van bioassay gegevens kunnen dus ook de selectie van organismen en levensstadia effect hebben op de 'hazard' bepaling.

Betrouwbare bioassay gegevens worden gevormd door gegevens die reproduceerbaar zijn. Omdat het herhalen van een test kostbaar en onpraktisch is, wordt betrouwbaarheid van geproduceerde gegevens momenteel beoordeeld met scoringssystemen. De meeste van die systemen richten zich op het niveau van standaardisatie en de documentatie van een experiment. Hierdoor worden bij het selecteren op betrouwbaarheid niet altijd de meest relevante gegevens gekozen. Daarnaast is een dergelijk scoringssysteem geen garantie voor gegevens die vrij van fouten zijn (**Hoofdstuk 2**) en is bovendien lastig kwantitatief te vergelijken met andere kwaliteitsaspecten van gegevens, zoals de relevantie en (statistische) zekerheid.

Europese richtlijnen voor het selecteren van bioassay gegevens, voor het gebruik in 'hazard'- en uiteindelijk risicobeoordeling, richten momenteel vooral op de betrouwbaarheid en relevantie van de data. Het selectieproces zelf heeft echter ook invloed op de statistische zekerheid van de 'hazard' beoordeling. Het zou daarom goed zijn om bij het selectieproces al rekening te houden met deze zekerheid. Nu gebeurt dit pas nadat het selectieproces is afgerond bij het beoordelen van het risico. Daarnaast zijn de huidige richtlijnen niet expliciet over waarom specifieke criteria worden gebruikt bij de selectie van data. Het wordt daarom voorgesteld om expliciet te maken hoeveel de betrouwbaarheid, relevantie en statistische zekerheid moet toenemen bij de selectie van data en waarom. In dat geval kan worden bepaald of de toegepaste selectie daadwerkelijk de verbetering heeft bereikt die men voor ogen had. De verbetering van een enkel aspect (bijvoorbeeld betrouwbaarheid) kan daarbij ten koste zijn gegaan van een ander aspect (bijvoorbeeld zekerheid). Criteria zouden daarom iteratief moeten worden bijgesteld tot er een geoptimaliseerde balans tussen betrouwbaarheid, relevantie en zekerheid is ontstaan.

Toekomstig onderzoek zou zich moeten richten op het ontwikkelen van een

raamwerk waarin data kwaliteitsaspecten (betrouwbaarheid, relevantie en zekerheid) onderling (en kwantitatief) met elkaar vergeleken kunnen worden. Een mogelijke richting daarbij is de 'weight of evidence' (bewijskracht) benadering. Bij die benadering wordt de kwaliteit van data meegerekend bij de bewijskracht daarvan, om zo tot een gebalanceerd beeld van het 'hazard' te komen. Daarnaast moeten we nieuwe ontwikkelingen in experimentele en statistische technieken bij het uitvoeren en het uitdrukken van bioassays niet uit het oog verliezen. (Ver)ouder(d)e bioassay resultaten blijven niettemin een waardevolle bron van informatie. Dit zolang de gegevens vanuit hun laboratoriumomstandigheden juist worden vertaald naar de veldomstandigheden, relevant voor de context waarin ze worden gebruikt.

Dankwoord

Het zal medio 2013 zijn geweest dat ik de knoop definitief door heb gehakt om mijn wetenschappelijke werk te vertalen naar een promotieonderzoek. Omdat ik het mij niet kon veroorloven om als AIO in dienst te treden, besloot ik het traject naast mijn reguliere baan als contractonderzoeker te bewandelen. Ik had eigenlijk geen slechter moment kunnen kiezen. Wageningen Marine Research (toen nog IMARES) was in zwaar weer terecht gekomen en voor mij en collega's was het alle hens aan dek. Voor 'indirecte' taken zoals promoveren was eigenlijk geen ruimte meer. Zonder de hulp van anderen had ik het resultaat dat nu voor u ligt niet voor elkaar kunnen krijgen. Mijn dank gaat dan ook uit naar iedereen die mij gesteund hebben, een aantal mensen wil ik hier in het bijzonder noemen zonder af te doen aan iedereen die ik niet expliciet bij naam noem.

In mijn dynamische werkomgeving heb ik verschillende leidinggevenden mee mogen maken tijdens mijn promotieonderzoek. Elk van hen wil ik graag danken voor hun rol in dit traject. John Schobben wist me uit te stimuleren om te gaan publiceren. Floris Groenendijk gaf me de kans en de ruimte om er formeel mee te starten en daar een flinke slag in te maken. Jeannette Riensema hielp me na onze reorganisatie te prioriteren en de staart van mijn promotie te combineren met het reguliere werk.

Tinka Murk wist al dat ik ging promoveren voordat ik het zelf wist. Je betrokkenheid bij mijn allereerste publicatie en je overredingskracht hebben mij over de streep getrokken om het promotietraject te formaliseren. Ik heb het als een voorrecht ervaren dat je mij daarbij als begeleider hebt ondersteund. Dank voor je inhoudelijke input, maar ook de nodige peptalks. Je wist me altijd net dat zetje te geven dat ik nodig had. Edwin Foekema heb ik natuurlijk niet voor niets als copromotor gevraagd. Ik heb altijd veel bewondering voor hoe Edwin 'outside the box' kan denken; volgens mij zit hij niet eens ìn de box. Je creatieve manier van denken heeft mij geholpen dit proefschrift tot één geheel te maken. Dank dat je zonder twijfel inging op mijn verzoek.

De overige leden van de commissie wil ik ook danken. Deelname in een commissie als deze wordt vaak als erezaak beschouwd, wat eigenlijk betekent dat je deze taak vooral in de avonduren en weekenden uitvoert. Met deze woorden

wil ik mijn waardering voor jullie inzit benadrukken.

Een paar collega's wil ik in het bijzonder bedanken. Diana Slijkerman, zonder jouw motivatie, inbreng, (morele) ondersteuning en het beschikbaar stellen van middelen (tijd) was het 'flamingo'-hoofdstuk er nooit geweest. Jacqueline Tamis, wij werken veel in projecten samen en hebt mij regelmatig uit de brand gered door mij werk uit handen te nemen, zodat ik onder andere aan mijn proefschrift kon werken. Daarnaast had je altijd tijd voor een ontspannend potje tafelvoetbal. Dit geldt overigens ook voor de overige collega's. Ruud Jongbloed, Robbert Jak en Jan Tjalling van der Wal om er een paar te noemen (ook al was niet iedereen van het tafelvoetballen).

De meeste coauteurs heb ik inmiddels al de revue laten passeren. Maar ook de overige ben ik natuurlijk dankbaar: Dick de Zwart, Chris Klok, Christiaan Kwadijk, Leo Posthuma, Mathijs Smit en Michiel Kotterman. Jullie waardevolle kritische inbreng hebben een belangrijke bijdrage geleverd aan de kwaliteit van de verschillende hoofdstukken.

Met Jantine Leeflang als studiegenoot is het mij wonderwel gelukt om mijn studie scheikunde tot een succesvol einde te brengen. Ik ben heel blij met onze vriendschap en dat je mij als paranimf wilt ondersteunen tijdens mijn verdediging.

Mijn schoonouders, Gert-Jan en Henriëtte, wil ik niet alleen danken voor hun ondersteuning aan zowel praktische als morele kant. Fijn ook dat Gert-Jan als ervaringsdeskundige mij als mijn paranimf wil bijstaan.

Mijn ouders groeiden op in een tijd dat waarin toegankelijkheid tot een goede opleiding beperkt was. Ik besef dan ook dat het een groot goed is, en niet vanzelfsprekend, dat ik dit proefschrift mag verdedigen. Dankbaar ben ik voor mijn ouders dat ze mij met liefde hebben grootgebracht en mij gestimuleerd hebben om mij te ontwikkelen tot wie ik ben en waar ik nu sta. Dat geldt ook voor mijn broers Jasper en Wibo die met broederlijke rivaliteit maar ook veel lol daaraan hebben bijgedragen. Wibo ook dank voor het ontwerp van de omslag van dit proefschrift, dat maakt het werk echt af.

Promoveren wordt ook wel eens gezien als een huwelijk met de wetenschap. En in die analogie voelde het schrijven van het proefschrift ook wel wat als overspel. Zeker is in ieder geval dat er heel wat vrije tijd in gaat zitten, tijd die ik niet aan mijn vrouw en kinderen heb kunnen besteden. Ik ben jullie dankbaar voor de liefdevolle ondersteuning en de ruimte die jullie mij hebben gegeven om dit proefschrift te schrijven. Geke, Myrthe en Eline, dit proefschrift is ook van en voor jullie. Ik hou van jullie.

Curriculum vitae

Pepijn de Vries was born in Hoorn in 1979 and was raised in Enkhuizen. In Enkhuizen he also enroled in vwo (pre-university secondary education) at the Regionale Scholen Gemeenschap (Regional School Community) from which he graduaded in 1997.

In that same year he started his chemistry education at the faculty of science of the Vrije Universiteit Amsterdam. His programming skills were put to the test in his minor in computer science. His focus was finaly drawn to pharmaceutical science, and he majored in molecular toxicology. His time in university had to be balanced with (primarily administrative) part time jobs.

After obtaining his Master degree in 2005, he accepted a new challenge and started working at TNO (the Dutch Institute for Applied Science) in 2006 in a field that was new to him: environmental risk assessment. He initially worked on hazard and risk assessment of chemicals.

In 2008 the department of TNO merged with the Nederlands Instituut voor Visserij Onderzoek (RIVO, Dutch Fisheries Research Institute) and a part of AL-TERRA to form the research institute that was initially called IMARES, but is currently known as Wageningen Marine Research (WMR). Pepijn joined this new organisation during the merger where his focus on chemicals widened to assessment of anthropogenic impacts in general on the marine environment.

At WMR he worked on various projects, but a reoccuring theme is his work with the offshore oil and gas industry. In the years 2010 to 2012 he was hired as an advisor to the Dutch delegates in OSPAR's Offshore Industry Committee (OIC). In this period he assisted with drafting of OSPAR Recommendation 2012/5 for a risk-based approach to the Management of Produced Water Discharges. He also advised on the Dutch implementation of this recommendation and compared different methodological approaches used in this framework for the Norwegian industry.

Currently Pepijn works on a diversity of projects, where data processing, analysing and presentation are key. Many of the projects involve cumulative effect assessment, as there is a growing demand for a consistent approach to assess effects of human activities in the complex marine environment.

List of publications

Peer reviewed articles

De Vries, **P**., Murk, A. J., 2013. Compliance of LC50 and NOEC data with Benford's Law: An indication of reliability? Ecotox Environ Safe 98, 171–178.

De Vries, **P**., Slijkerman, D. M. E., Kwadijk, C. J. A. F., Kotterman, M. J. J., Posthuma, L., De Zwart, D., Murk, A. J., Foekema, E. M., 2017. The toxic exposure of flamingos to *per*- and Polyfluoroalkyl substances (PFAS) from firefighting foam applications in Bonaire. Mar Pollut Bull 124 (1), 102–111.

De Vries, **P**., Smit, M. G. D., Van Dalfsen, J. A., De Laender, F., Karman, C. C., 2010. Consequences of stressor-induced changes in species assemblage for biodiversity indicators. Environ Toxicol Chem 29 (8), 1868–1876.

De Vries, **P**., Tamis, J. E., Foekema, E. M., Klok, C., Murk, A. J., 2013. Towards quantitative ecological risk assessment of elevated carbon dioxide levels in the marine environment. Mar Pollut Bull 73 (2), 516–523.

De Vries, **P**., Tamis, J. E., Murk, A. J., Smit, M. G. D., 2008. Development and application of a species sensitivity distribution for temperature-induced mortality in the aquatic environment. Environ Toxicol Chem 27 (12), 2591–2598.

Foekema, E. M., Fischer, A., Parron, M. L., Kwadijk, C., **De Vries**, **P**., Murk, A. J., 2012. Toxic concentrations in fish early life stages peak at a critical moment. Environ Toxicol Chem 31 (6), 1381–1390.

Foekema, E. M., Kotterman, M., **De Vries**, **P**., Murk, A., 2016. Maternally transferred dioxin-like compounds can affect the reproductive success of European eel. Environ Toxicol Chem 35 (1), 241–246.

Piet, G. J., Jongbloed, R. H., Knights, A. M., Tamis, J. E., Paijmans, A. J., Van der Sluis, M. T., **De Vries**, **P**., Robinson, L. A., 2015. Evaluation of ecosystem-based marine management strategies based on risk assessment. Biol Conserv 186, 158–166.

Piet, G. J., Knights, A. M., Jongbloed, R. H., Tamis, J. E., **De Vries**, **P**., Robinson,

L. A., 2017. Ecological risk assessments to guide decision-making: methodology matters. Environ Sci Policy 68, 1–9.

Roques, J. A. C., Schram, E., Spanings, T., Van Schaik, T., Abbink, W., Boerrigter, J., **De Vries**, **P**., Van de Vis, H., Flik, G., 2015. The impact of elevated water nitrite concentration on physiology, growth and feed intake of African catfish *Clarias gariepinus* (Burchell 1822). Aquac Res 46 (6), 1384–1395.

Schram, E., Roques, J. A., Van Kuijk, T., Abbink, W., Van de Heul, J., **De Vries**, **P**., Bierman, S., de Vis, H. V., Flik, G., 2014a. The impact of elevated water ammonia and nitrate concentrations on physiology, growth and feed intake of pikeperch (*Sander lucioperca*) aquaculture. Aquaculture 420-421, 95–104.

Schram, E., Roques, J. A. C., Abbink, W., Spanings, T., **De Vries**, **P**., Bierman, S., Van de Vis, H., Flik, G., 2010. The impact of elevated water ammonia concentration on physiology, growth and feed intake of African catfish (*Clarias gariepinus*). Aquaculture 306 (1-4), 108–115.

Schram, E., Roques, J. A. C., Abbink, W., Yokohama, Y., Spanings, T., **De Vries**, **P**., Bierman, S., Van de Vis, H., Flik, G., 2014b. The impact of elevated water nitrate concentration on physiology, growth and feed intake of African catfish *Clarias gariepinus* (Burchell 1822). Aquac Res 45, 1499–1511.

Slijkerman, D. M. E., De León, R., **De Vries**, **P**., 2014. A baseline water quality assessment of the coastal reefs of Bonaire, southern Caribbean. Mar Pollut Bull 86 (1-2), 523–529.

Tamis, J. E., **De Vries**, **P**., Jongbloed, R. H., Lagerveld, S., Jak, R. G., Karman, C. C., Van der Wal, J. T., Slijkerman, D. M. E., Klok, C., 2016. Towards a harmonised approach for environmental assessment of human activities in the marine environment. Integr Environ Assess Manag 12, 632–642.

Tulp, I., Bolle, L. J., Meesters, E., **De Vries**, **P**., 2012. Brown shrimp abundance in northwest European coastal waters from 1970 to 2010 and potential causes for contrasting trends. Mar Ecol Prog Ser 458, 141–154.

Poster and platform presentations

De Vries, **P**., 2010a. Environmental aspects of offshore exploration and production of oil and gas. Research School SENSE, Special Topics in Ecotoxicology, 't Horntje.

De Vries, **P**., 2010b. Probabilistic ecological risk assessment for thermal discharges. Workshop Sustainable Use of Cooling Water from the Wadden Sea, Nieuweschans.

De Vries, **P**., Karman, C. C., 2009. Environmental risk assessment of produced water. Twinfilter, Seminar Produced Water, Zaandam.

De Vries, **P**., Klok, C., Karman, C. C., 2010. A risk based approach for the management of produced water – for better or worse? TUV-NEL, 8th Produced Water Workshop, Aberdeen.

De Vries, **P**., Tamis, J. E., Murk, A. J., Smit, M. G. D., 2008. Development and application of a species sensitivity distribution for temperature-induced mortality in the aquatic environment. SETAC Europe 18th Annual Meeting, Warsaw.

Jak, R., **De Vries**, **P**., Tamis, J., Hjorth, M., Falk-Petersen, S., Hemerik, L., 2016a. Matrix models for Arctic *Calanus* species: using experimental toxicity data for oil components to assess population level effects. 7th SETAC World Congress, Orlando, FL, USA.

Jak, R., Tamis, J. E., **De Vries**, **P**., Hemerik, L., Hjorth, M., 2016b. Matrix models to assess acute and chronic effects of oil spills on Arctic *Calanus* populations. Arctic Frontiers, Tromsø.

Kater, B., Van den Hevel-Greve, M., Thijsse, P., Beegle-Krause, C., Bos, O., Grasmeijer, B., Griffin, L., Jak, R. Van Onselen, E., Van Overzee, H., Piet, G., , Tamis, J., Tuijnder, A., **De Vries**, **P**., Van der Wal, J., Steins, N., 2016. An overview of the adequacy of Arctic sea basin data. ICES Annual Science Conference, Riga, Latvia.

Klok, C., Ravagnan, E., Grøsvik, B. E., Hjort, M., Hansen, B. H., Farmen, E., Harlang Olsen, G., De Laender, F., Hendriks, J., Karman, C., **De Vries**, **P**., Nordtug, T., Janssen, C., De Hoop, L., Sanni, S., 2012. Applying dynamic energy budget theory to estimate the impact of oil components in an Arctic food chain. SETAC Europe 22nd Annual Meeting, Berlin.

Schram, E., Roques, J., Abbink, W., Yokohama, Y., Spanings, T., **De Vries**, **P**., Bierman, S., Van de Vis, H., Flik, G., 2012. Exposure of fish to high nitrate in recirculating aquaculture systems affects growth, feed intake and physiology. World Aquaculture Society - AQUA 2012, Prague.

Tulp, I., Bolle, L., Meesters, H. W. G., **De Vries**, **P**., 2011. Brown shrimp *Crangon crangon* abundance in northwest European coastal waters and estuaries in the period 1970-2010 and potential causes for contrasting trends. International Symposium on the Ecology of the Wadden Sea, 't Horntje.

Van Erp, R., Mathôt, J., Kortekaas, P., Verlinden, J., Van Koperen, R., **De Vries**, **P**., 2010. Update of the E3 tug development - economically viable, environmentally friendly, efficient in operation. The 21st International Tug and Salvage Convention and Exhibition, Vancouver.

Selection of additional publications

Bierman, S. M., **De Vries**, **P**., Kaag, N. H. B. M., 2012. The development of a full standard methodology for testing ballast water discharges for gross non-compliance of the IMO's ballast water management convention (EMSA/NEG/12/2012). Tech. Rep. C124a/12, IMARES, Den Helder.

Deerenberg, C. M., Brinkman, A. G., **De Vries**, **P**., 2015. Inventarisatie van modellen van IMARES geschikt voor beantwoording beheervragen Rijkswaterstaat. Tech. Rep. C006/15, IMARES Wageningen UR, Den Helder.

De Vries, **P**., 2012. Optimising environmental performance – the E3 tug project. SWZ Maritime 133, 24–27.

De Vries, **P**., Hintzen, N. T., Slijkerman, D. M. E., 2015. Fisheries displacement effects of managed areas: a case study of De Voordelta. Tech. Rep. C136/15, IMARES, Den Helder.

De Vries, **P**., Karman, C. C., 2009. Environmental risk assessment of produced water discharges on the Dutch Continental Shelf. Tech. Rep. C012/09, Wageningen IMARES, Den Helder.

De Vries, **P**., Karman, C. C., 2010. Whole effluent toxicity and substance-based hazard assessment of produced water. Tech. Rep. C162/11, IMARES, Den Helder.

De Vries, **P**., Karman, C. C., 2012a. OSPAR risk based approach: Intercomparison of phase II calculations on produced water. Tech. Rep. C021/12, IMARES, Den Helder.

De Vries, **P**., Karman, C. C., 2012b. Whole effluent toxicity and substance-based risk assessment of produced water: Phase II calculations on Dutch and Danish platforms. Tech. Rep. C020/12, IMARES, Den Helder.

De Vries, **P**., Smit, M. G. D., Jongbloed, R. H., Karman, C. C., 2007. Random sampling of HOCNFs and underlying test reports. Tech. Rep. C82/07, Wageningen IMARES, Den Helder.

De Vries, **P**., Sneekes, A. C., Velzeboer, I., Karman, C. C., 2009. Alkylphenols in produced water: A first tier risk assessment. Tech. Rep. C007/09, Wageningen IMARES, Den Helder.

De Vries, **P**., Tamis, J. E., 2014. Manual to proposed Dutch implementation of OSPAR's risk-based approach to the management of produced water discharges. Tech. Rep. C057/14, IMARES, Den Helder.

De Vries, **P**., Tamis, J. E., Bolman, B. C., Lahr, J., Rau, M.-L., Rietra, R. P. J. J., 2014. Pilot project mapping of literature on discharge of mud and cuttings - project and synthesis report. Tech. Rep. C007/15, IMARES, Den Helder.

De Vries, **P**., Tamis, J. E., Van den Heuvel-Greve, M. J., Thijsse, P., Kater, B., 2016. Collecting literature for identifying data sets and data sources. Tech. Rep. C072/16, IMARES Wageningen UR, Den Helder.

De Vries, **P**., Tamis, J. E., Van der Wal, J. T., Jak, R. G., Slijkerman, D. M. E., Schobben, J. H. M., 2011. Scaling human–induced pressures to population level impacts in the marine environment – implementation of the prototype CUMULEO-RAM model. Tech. Rep. Werkdocument 285, Wageningen UR, Wettelijke Onderzoekstaken Natuur en Milieu, Wageningen.

Elschot, K., De Groot, A., Dijkema, K., Sonneveld, C., Van der Wal, J. T., **De Vries**, **P**., Brinkman, A. G., Van Duin, W., Molenaar, W., Krol, J., Kuiters, A. T., De Vries, D., Wegman, R. M. A., Slim, P. A., Koppenaal, E. C., De Vlas, J., 2017. Ontwikkeling kwelder Ameland-Oost : Evaluatie bodemdalingsonderzoek 1986-2016. In: De Vlas, J. (Ed.), Monitoring effecten van bodemdaling op Oost-Ameland. De Waddenacademie, Leeuwarden, Ch. 4, pp. 185–328.

Griffioen, A. B., **De Vries**, **P**., Twijnstra, R. H., De Graaf, M., 2017. Glass eel monitoring in the netherlands. Tech. Rep. C010/17, Wageningen Marine Research, IJmuiden.

Hulskotte, J. H. J., Jongbloed, R. H., **De Vries**, **P**., Appelman, W. A. J., Heslinga, D. C., 2007. Afvalwaterketenonderzoek (AKON): Emissiebronnen en maatregelen in het verzorgingsgebied van twee RWZI's in het beheersgebied van Aa en Maas. Tech. Rep. 2007-A-R0326/B, TNO Bouw en Ondergrond, Apeldoorn.

Jansen, J., Kersting, K., **De Vries**, **P**., Davaasuren, N., Sneekes, A., Smaal, A., 2013. Effecten van mosselvisserij op de bodemstructuur, een analyse van sidescan sonar opnames. Tech. Rep. C164/12, IMARES, Den Helder.

Klok, C., **De Vries**, **P**., Jongbloed, R., Tamis, J., 2012. Literature review on the sensitivity and exposure of marine and estuarine organisms to pesticides in comparison to corresponding fresh water species. Tech. Rep. 2012:EN-357, IMARES, Den Helder.

Klok, C., Hille Ris Lambers, R., **De Vries**, **P**., Tamis, J. E., Wijsman, J. W. M., 2009. Model instruments for marine biodiversity policy. Tech. Rep. Werkdocument 148, Wageningen UR, Wettelijke Onderzoekstaken Natuur en Milieu, Wageningen.

Kotterman, M. J. J., **De Vries**, **P**., 2015. Dioxines en PCB's in Chinese wolhandkrab; invloed van grootte en variatie door het seizoen. Tech. Rep. C057/15, IMARES Wageningen UR, IJmuiden.

Lagerveld, S., Jonge Poerink, B., **De Vries**, **P**., 2015. Monitoring bat activity at the Dutch EEZ in 2014. Tech. Rep. C094/15, IMARES Wageningen UR, Den Helder.

Lagerveld, S., Jonge Poerink, B., **De Vries**, **P**., Scholl, M. M., 2016. Monitoring bat activity at the Dutch EEZ in 2015. Tech. Rep. C001/16, IMARES Wageningen UR, Den Helder.

Pijnenburg, J., Pietersz, T., **De Vries**, **P**., 2007. Risk assessment of benzene emissions from offshore installations in the North Sea. Tech. Rep. 2007.011, RWS Rijksinstituut voor Kust en Zee, Den Haag.

Slijkerman, D. M. E., De León, R., **De Vries**, **P**., Koelemij, E., 2012. Water quality monitoring Bonaire: Results monitoring November 2011 and recommendations for future research. Tech. Rep. C028/12, IMARES, Den Helder.

Slijkerman, D. M. E., De León, R., **De Vries**, **P**., Koelemij, E., 2013a. Water quality of the coastal zone of Bonaire: Results field monitoring 2011–2013. Tech. Rep. C158/13, IMARES, Den Helder.

Slijkerman, D. M. E., **De Vries**, **P**., Kotterman, M. J. J., Cuperus, J., Kwadijk, C. J. A. F., Van Wijngaarden, R., 2013b. Saliña Goto and reduced flamingo abundance since 2010: Ecological and ecotoxicological research. Tech. Rep. C211/13, IMARES, Den Helder.

Tamis, J. E., Karman, C. C., **De Vries**, **P**., Jak, R. G., Klok, C., 2011. Offshore olie- en gasactiviteiten en Natura 2000: Inventarisatie van mogelijke gevolgen voor de instandhoudingsdoelen van de Noordzee. Tech. Rep. C144/10, IMARES, Den Helder.

Tulp, I., Bolle, L. J., Dänhardt, **De Vries**, **P**., Haslob, H., Jepsen, N., Scholle, J., Van der Veer, H. W., 2017. Fish. Wadden Sea quality status report 2017. Tech. rep., Common Wadden Sea Secretariat, Willemshaven.

Van der Wal, J. T., **De Vries**, **P**., Tamis, J. E., 2016. Growth and innovation in the ocean economy: North Sea checkpoint; data adequacy report – oil platform leak challenge. Tech. Rep. C095/16, IMARES Wageningen UR, Den Helder.

Netherlands Research School for the
Socio Economic and Natural Sciences of the Environment

D I P L O M A

For specialised PhD training

The Netherlands Research School for the
Socio Economic and Natural Sciences of the Environment
(SENSE) declares that

Pepijn de Vries

born on 31 May 1979 in Hoorn, the Netherlands

has successfully fulfilled all requirements of the
Educational Programme of SENSE.

Wageningen, 16 April 2018

the Chairman of the SENSE board	the SENSE Director of Education
Prof. dr. Huub Rijnaarts	Dr. Ad van Dommelen

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

The SENSE Research School declares that **Pepijn de Vries** has successfully fulfilled all requirements of the Educational PhD Programme of SENSE with a work load of 35 EC, including the following activities:

SENSE PhD Courses

- Research in context activity: 'Contributing to the accessible presentation and dissemination of marine research outcomes: To explore the potential of coastal and marine nature to improve the quality of life (2017)'

General skills Courses

- Dose Related Risk and Effects Assessment Model (DREAM) Training, SINTEF, Trondheim, Norway (2006)
- Klantgericht handelen, Boertien Training (2007)
- Persoonlijke effectiviteit, Nijssen Partners (2007)
- Onzekerheids en gevoeligheidsanalyse voor modelbouwers, Wageningen Business School, Wageningen University & Research (2008)
- Use of the Web ICE modelling application to estimate species sensitivity to contaminants, SETAC Europe (2008)

External training at a research institute

- Managing and executing research projects, TNO, The Netherlands (2006 2009)

Management and Didactic Skills Training

- Organising a half day workshop for Dutch oil & gas producers on the implementation of OSPAR's risk based approach for the management of produced water discharges (2014)

Selection of Oral Presentations

- *Development and application of a species sensitivity distribution for temperature induced mortality in the aquatic environment*. SETAC Europe 18th Annual Meeting, 26 29 May 2008, Warsaw, Poland
- *Environmental risk assessment of produced water.* Seminar Produced Water, 15 October 2009, Zaandam, The Netherlands
- *A risk based approach for the management of produced water – for better or worse?".* TUV NEL, 8th Produced Water Workshop, 23 24 June 2010, Aberdeen, United Kingdom

SENSE Coordinator PhD Education

Dr. Peter Vermeulen

Colofon

Cover design:
Wibo de Vries

Photography:
Chapter 2, page 14; Pepijn de Vries
Chapter 3, page 38; Oscar Bos
Chapter 4, page 70; Pepijn de Vries
Chapter 5, page 104; Steve Geelhoed

Printed by:
Digiforce ‖ `www.proefschriftmaken.nl`